Analysis of design strategies for RF ESD problems in CMOS circuits

Pugsley, William

Analysis of design strategies for RF ESD problems in CMOS circuits

William Pugsley

A Thesis presented for the degree of

Doctor of Philosophy

Centre for Electronic Systems

School of Engineering

University of Durham

England

2007

Declaration

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

Acknowledgements

I would like to thank my project supervisor, Simon Johnson, for his time, help and encouragement, my colleagues at Durham, Craig Robinson, Richard McWilliam and Thomas Grocutt who gave me no end of help in setting up and running computers to do my finite element simulations. I also received considerable help from the technical staff in the Engineering department, in particular Stuart Watson whose considerable machining skill was crucial to the construction of a low cost TLP tester. I am very much indebted to the staff at CSR in particular Ian Sabberton, James Collier, Simon Chang, Justin Penfold, Paul Egan and Jens Zolnhofer all of whom provided me with considerable help on the theoretical side of things as well as practical help in implementing test structures on a $0.18\mu m$ commercial chip. I would like to thank Robert Manning and my mother for their patience whilst proof reading this thesis. I would also like to thank the open source programming team that created LaTeX and thus saved me from Word.

Abstract

This thesis analyses the design strategies used to protect RF circuits that are implemented in CMOS technologies. It investigates, in detail, the physical mechanisms involved when a ggNMOS structure is exposed to an ESD event and undergoes snapback. The understanding gained is used to understand why the performance of the current RF ESD clamp is poor and suggestions are made as to how the performance of ggNMOS clamps can be improved beyond the current body of knowledge. The ultimate aim is to be able to design effective ESD protection clamps whilst minimising the effect the circuit has on RF I/O signals.

A current ggNMOS based RF ESD I/O protection circuit is analysed in detail using a Transmission Line Pulse (TLP) tester. This is shown to be a very effective diagnostic tool by showing many characteristics of the ggNMOS during the triggering and conducting phase of the ESD event and demonstrate deficiencies in the clamp design. The use of a FIB enhances the analysis by allowing the isolation of individual components in the circuit and therefore their analysis using the TLP tester. SPICE simulations are used to provide further commentary on the debate surrounding the specification required of a TLP tester for there to be a good correlation between a TLP test and the industry standard Human Body Model (HBM) ESD test.

Finite element simulations are used to probe deeper in to the mechanisms involved when a ggNMOS undergoes snapback especially with regard to the contribution parasitic components within the ggNMOS make to the snapback process. New

ggNMOS clamps are proposed which after some modification are shown to work. Some of the finite element experiments are repeated in a $0.18\mu m$ process CMOS test chip and a comparison is made between the two sets of results.

In the concluding chapter understanding that has been gained from previous chapters is combined with the published body of knowledge to suggest and explain improvements in the design of a ggNMOS for RF and standard applications. These improvements will improve homogeneity of ggNMOS operation thus allowing the device size to be reduced and parasitic loading for a given ESD performance. These techniques can also be used to ensure that the ESD current does not take an unintended path through the chip.

Contents

List of Figures

List of Tables

Glossary

μ_n	Electron mobility
μ_p	Hole mobility
β	Current gain of a bipolar transistor.
Back biasing	Technique to alter the operating mode of a MOS transistor by changing the bulk voltage
BC01b	CSR's first commercial product
BigFET	ESD protection methodology that uses large standard nMOS transistors
CDM	Charged Device Model
CMOS	Complimentary Metal Oxide Semiconductor
C_S	Socket capacitance
C_P	Parasitic capacitance
CSR	Cambridge Silicon Radio
D_n	Electron diffusivity
D_p	Hole diffusivity
DUT	Device Under Test
Electromigration	Movement of material due to the flow of current in the material
EMI	Electro Magnetic Interference
EMMI	PhotoEmission Microscopy
EOS	Electrical Over Stress

EOS	Electrical Over Stress
EPI	Epitaxial [layer] (Thin crystalline growth on a single crystal substrate).
ESD	Electrostatic Discharge
ESDA	Electrostatic Discharge Association
Eye Diagram	A graphical method for determining whether a signal level is sufficiently high/low over a number of digital bits to ensure reliable data transfer
FIB	Focused Ion Beam
f_{MAX}	Frequency when power gain reaches unity
f_T	Frequency when current gain reaches unity
ggNMOS	grounded gate n-type MOS
Ga^+	Gallium Ion
G_M	Transconductance of a MOSFET (dI/dV)
G_n	Rate of electron generation
G_p	Rate of hole generation
GPS	Global Positioning System
GUI	Graphical User Interface
HBM	Human Body Model
HNO_3	Nitric Acid
Hotplugging	The action of removing or adding a component whilst it is operational
I_{CE}	Collector to emitter current in a bipolar transistor
I/O	Input Output
Impact ionisation	Generation of carriers in high electric fields

IV	Current Voltage
$I_{DS@2V}$	I_{DS} when V_{DS}=2V
I_{DS}	Current flow from the drain to source
I_{DUT}	Current flow through the Device Under Test (DUT)
I_{hold}	The minimum current required to keep a ggNMOS in the snapback mode
I_{leak}	The leakage current
I_{max}	Maximum current
I_{peak}	Peak current
I_{t1}	Current required to trigger a ggNMOS in to the snapback mode
I_{t2}	Current at which a ggNMOS fails whilst in the snapback mode
I_{test}	Test current
I_{TLP}	The current drawn from the TLP tester
JEDEC	Joint Electron Device Engineering Council
J_n	Electron current density
J_p	Hole current density
LDD	Lightly Doped Drain
L_P	Parasitic inductance
LNA	Now Noise Amplifier
LRC	Resonant circuit consisting of an Inductor, Capacitor and (parasitic) Resistor.
L_{drain}	Length of the silicide blocked drain extension
L_{gate}	Gate length
MFT	Multi Finger Turn on (Mechanisms)

MM	Machine Model ESD test
M	Avalanche multiplication gain
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
n	electron carrier density
Pocket implants	Additional dopants in the channel near the source and drain junctions to improve punch through control.
p	hole carrier density
RF	Radio Frequency
$R_{contact}$	Contact resistance
R_P	Parasitic resistance
R_{DUT}	Resistance of the Device Under Test
R_{error}	Error in the resistance calculation
$R_{ESD-clamp}$	Resistance of the ESD clamp
$R_{ESD-source}$	Resistance of the ESD source
R_{HBM}	Output resistance of the HBM tester
R_{on}	The differential resistance of the ggNMOS once triggered in the snapback mode, i.e. $\frac{dV}{dI}$ and not $\frac{V}{I}$
SCR	Silicon Controlled Rectifier
SD	Source Drain (with reference to MOSFET doping)
Secondary breakdown	Thermal failure of a device leading to permanent damage
Snapback	Mode of operation for the ggNMOS ESD clamp whilst it is diverting ESD current
SPICE	Simulation Program with Integrated Circuit

Substrate pump	Emphasis Circuit specifically designed to inject carriers in to the substrate
Sequoia Device Designer	Semiconductor finite element electrothermal simulation software supplied by Sequoia
STI	Shallow trench isolation
Taurus	Semiconductor finite element electrothermal simulation software supplied by Avant!
TCL	Tool Command (software) Language
Teradyne Tester	Automatic test equipment for use in an IC manufacturing environment
TL	Transmission Line
TLP	Transmission Line Pulse (tester)
TX_A/B	Transmitter A/B
t_{6V-HBM}	Time taken for the HBM pulse to reach 6V from 0V
t_{6V-TLP}	Time taken for the TLP pulse to reach 6V from 0V
t_F	Fall time
t_L	Pulse length
T_{MAX}	Maximum temperature
T_{OX}	Gate oxide thickness
t_R	Rise time
t_{RC}	Time constant (Resistance*Capacitance)
t_{R-HBM}	10% to 90% rise time of an HBM pulse
t_{R-TLP}	10% to 90% rise time of an TLP pulse
U_n	Rate of electron recombination
U_p	Rate of hole recombination
UWB	Ultra Wide Band

V_{ava-sb}	Voltage difference between the avalanching voltage and triggering voltage (V_{t1})
V_{BE}	Base emitter voltage of a bipolar transistor
V_{BS}	Bulk source voltage of a MOSFET
V_{CC}	Circuit power supply voltage
V_D	Drain voltage
V_{DD}	Circuit power supply voltage
V_{DS}	Drain to source voltage across a MOSFET
V_{DUT}	Voltage across a Device Under Test (DUT)
V_{GS}	Gate to source voltage across a MOSFET
V_{HBM}	HBM test voltage (the voltage across the capacitor that is discharged in to the DUT)
V_{hold}	The minimum voltage required to keep a ggNMOS in the snapback mode
V_{ox-max}	Voltage threshold at which permanent damage occurs in the gate oxide.
V_{source}	Source voltage
V_{SS}	Circuit ground voltage
V_T	Threshold voltage for a MOSFET during normal operation
V_{t1}	Voltage required to trigger a ggNMOS in to the snapback mode
V_{t2}	Voltage at which a ggNMOS fails whilst in the snapback mode
$VF-TLP$	Very Fast TLP

Introduction

This PhD was initiated by James Collier, Ian Sabberton and Justin Penfold, all of whom are co-founders of CSR. They recognised that ESD protection of RF pins on a CMOS process was going to be an issue. All non RF pins on the $0.35\mu m$ generation of BlueCore products, current at the time this PhD was started, achieved 2KV HBM performance (a standard customer requirement), yet the RF pins struggled to achieve 500V. Given that CSR's mission is to produce radio transceivers in silicon (at the time of writing CSR is developing or has plans to develop Bluetooth, GPS, FM and UWB solutions) RF ESD is clearly an important issue.

Protecting the RF pins from ESD events is not technically challenging in itself but the problem is that, at the time this PhD was started, all the ESD solutions available to CSR that passed 2KV HBM, subjected the RF pins to excessive capacitive loading. The balancing of ESD requirements and RF performance is a common problem in the semiconductor industry, so much so that in some cases ESD protection is absent to preserve RF performance [7][8].

In general ESD has also become an increasingly important issue for standard digital and analogue IOs. This is due to the increasing pin count per chip, decreasing gate oxide thicknesses, power supply voltages and the competitive pressures to reduce die size. So any results that can improve ESD protection structures will also help other IOs.

Chapter 1 covers the current state of ESD knowledge.

In Chapter 2 the Transmission Line Pulse tester (TLP) is introduced. This is a diagnostic tool that is especially useful for understanding and troubleshooting ESD protection circuits. In the following Chapter the TLP tester is used to diagnose the cause of ESD weakness on CSR's BC01b chip RF TX pins. The TLP tester is able to give detailed insight into the operation of the various circuit components under ESD conditions.

In Chapter 3 the TLP tester is used to analyse the RF outputs of CSR's first Bluetooth chip. The TLP tester is able to demonstrate why, despite there being a large ESD clamp, the RF inputs only achieve 500V HBM ESD rating. This case study is a textbook example of why the TLP tester is useful.

The simulation results for $0.18\mu m$ CMOS technology ggNMOS ESD protection structures are reported in Chapters 4 and 5 which also deal with the relationship between various physical parameters and device ESD performance. In the final section of Chapter 5 a new structure is proposed and shown not to work. The simulation demonstrates why it does not work and a modified solution is demonstrated.

Chapter 6 presents test structures manufactured in $0.18\mu m$ CMOS process; they are all connected to a single IO pad therefore to test them individually they have to be isolated. This was achieved by using a FIB to isolate the desired test structure so that it can be characterised. Some measurements agree with the results from Chapter 5; for the ones that did not a possible explanation was simulated and shown to be plausible, using finite element simulation.

Chapter 1

Literature Review

1.1 Introduction

1.1.1 What is ESD?

Electrostatic Discharge (ESD) is part of a broader category of Electrical Over Stress (EOS) [9][10]. EOS is: "the exposure of an object to a current or voltage beyond its maximum ratings" [11]. At one end of the spectrum EOS includes lightning, with extremely high voltage and extremely high power while, at the low power end, *specific EOS* has low voltage (1-100V), long duration (1-10ms) and low power. ESD is generally characterised as having high-voltage (1-15KV), short duration ($\sim 100ns$) and a fast rise time. ESD has also been described as "transient discharge of charge" [12].

For completeness, we shall also mention electromagnetic interference (EMI) [9] which occurs when an electromagnetic field induces a transient in an unshielded conductor. These transients are likely to be too weak to cause permanent damage but frequently are able to cause a system to malfunction temporarily.

1.1.2 How is ESD affecting the Electronics Industry?

The ESD Association [13] was founded in 1982 in response to the increasing problems that the semiconductor industry was having with ESD and related failures. Studies have shown that $10\% - 30\%$ of semiconductor failures can be attributed to ESD [9] [14][15] and in some cases suspected *specific EOS* failures can be attributed to ESD [12][16]. The continuing drive for smaller size and increased complexity in the semiconductor industry [17] coupled with the changes in semiconductor technology (e.g. LDD junctions and silicides) [12] has only exacerbated this problem.

1.1.3 How is the ESD Problem being solved?

The problem of ESD related failures has been tackled in many ways. The first and most obvious solution is to prevent the ESD events from happening at all. The aim is to prevent charge build up or, failing that, to allow charge to be conducted away safely. This has been tackled by introducing antistatic technologies like wrist straps, antistatic worktops, air ionisation, humidity control and operator education [9].

The alternative approach, in part covered by this thesis, is to develop IO blocks for integrated circuits that are more tolerant of ESD events. Section 1.2 covers the common ESD protection circuit elements and their operation. Section 1.4 and Chapter 2 cover all the common testing techniques as the various ESD solutions need to be benchmarked against recognised standards. The aim is not only to compare circuit performance and verify compliance to ESD standards but also to understand how a component or circuit operates under ESD conditions. It is particularly important to find out the mode of operation and the current path(s) during an ESD event so that the circuit can be improved.

Section 1.2.3 covers ESD protection strategies that have been suggested in the liter-

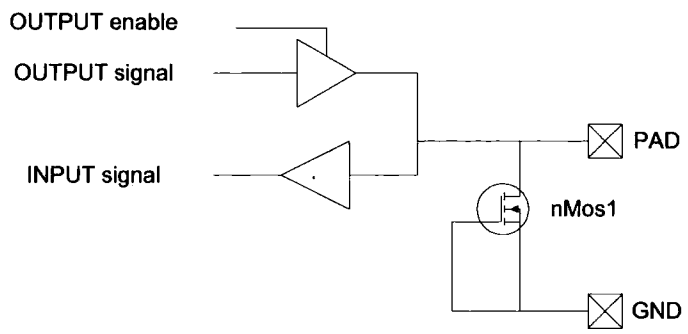

Figure 1.1: An IO circuit with ESD protection

ature as opposed to just concentrating on the individual ESD protection structures. These protection methods are particularly interesting as they are designed to protect the chip as a whole and not just the individual pin.

1.2 Protection Structures

All ESD protection circuits are a compromise as they must prevent ESD damage to the chip yet, at the same time, allow the chip to operate to specification at all other times. This is illustrated in Figure 1.1 which shows a standard bidirectional IO circuit. nMos1 protects the input and output buffers from ESD pulses between the output pin and ground whilst allowing the circuit to operate normally at all other times. The protection circuit could provide excessive capacitive or resistive shunt loading on the IO; it could also accidentally trigger, thus shorting the IO pin to ground. The ESD structure and the output driver need to be designed in such a way that the ESD protection structure will not conduct during normal operation (or at least, if it does conduct it will turn off sufficiently quickly that it will not disrupt normal circuit operation).

Figure 1.2 shows the safe operating area, in green, for an ESD protection device. During an ESD event the voltage between the PAD and GND (Figure 1.1) must

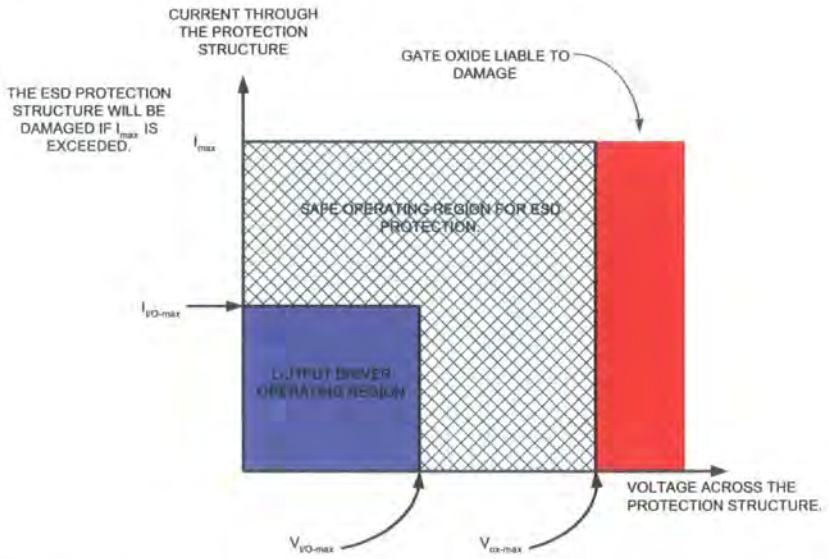

Figure 1.2: The safe operating area for an IO ESD protection circuit

not exceed V_{ox-max} otherwise the gate oxide of the input driver will be damaged. The blue area shows the current-voltage profile that the output driver is capable of supplying. The IV characteristics of the protection structure must also not cross into this area otherwise the ESD protection structure might interfere with the operation of the output driver. I_{max} is the maximum current that the protection structure can sustain before permanent damage occurs. This current should be as high as possible as it is directly related to the energy in the ESD event thus it is proportional to the level of ESD protection provided to the circuit.

The protection structures will be functioning well outside their traditional operating regime of $0.18\mu m$ process transistors. The rising slew rates are considerably faster, the current densities higher, but for a relatively short period of time, and the switching speed requirements are much more demanding.

These requirements determine the performance required of the ESD protection circuit under ESD conditions. The protection circuit has to allow the circuit to operate to its specification under normal operating conditions and must not present excessive series resistance or shunt loading to any IO signal. It also must not cause the cir-

cuit to consume excessive current from the power supply (including during standby conditions). The shunt loading and the series resistance is a particular problem at radio frequencies. n-well or poly silicon resistors are often used in series with the inputs to limit the ESD current though this will also have the effect of decreasing the maximum power transfer into and out of the IO pin. Shunt loading is also a problem at RF frequencies as many ESD protection structures are large, due to the fact that they have to dissipate power, and thus present a significant parasitic capacitive loading to ground.

1.2.1 ggNMOS Protection Structures with Emphasis on RF Requirements

The bulk of this thesis investigates the operation of the grounded gate NMOS (ggNMOS) as an ESD protection structure. It is a commonly used device as it can easily be made using the CMOS process. The standard nMOS transistor design is not a particularly effective ESD protection structure; therefore its mode of operation under ESD conditions needs to be understood so that the design can be altered and improved so as to minimise area, capacitance and leakage.

1.2.1.1 MOSFET Construction

Wolf [18] provides detailed information on the physical makeup of a typical $0.18\mu m$ CMOS process MOSFET including the manufacturing methods. This information is derived from the interpretation of published data and conversations with scientists and engineers from the semiconductor industry. Figure 1.3 shows the basic doping profile of an nMOS.

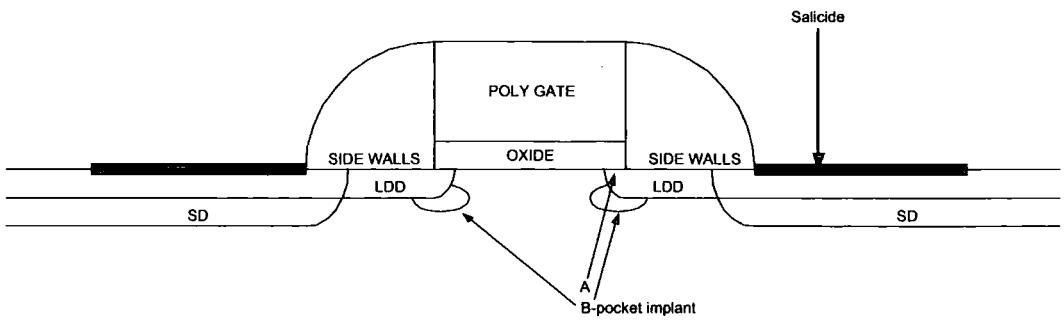

Figure 1.3: Doping profiles in a typical deep sub-micron MOSFET

1.2.1.2 *n*-type doping in an *n*-type MOSFET

The Source/Drain (SD) doping is high to ensure an ohmic contact with the metal contacts, the Lightly Doped Drain (LDD) doping is lower to reduce the maximum electric field [19][20] in the drain-substrate depletion region just under the gate oxide as indicated by (A) in Figure 1.3.

1.2.1.3 *p*-type doping in an *n*-type MOSFET

The substrate is doped to a uniform low background level ($\sim 10^{12} cm^{-3}$). Other implants are added to alter the threshold voltage of the transistor and the dopants are implanted in such a way that the peak implant density is not at but below the silicon surface. This type of doping is called a retrograde implant and has the advantage of allowing the dopant to affect the threshold voltage without the drift and diffusion velocity of the carriers being affected by high implant densities. Deeper in the silicon, high concentration pocket implants are added to increase the punch through voltage. These are indicated by (B) in Figure 1.3.

The gate oxide is around 32Å to 80Å thick and is covered with an *n-doped* polysilicon layer.

1.2.1.4 Understanding the Snapback Mechanism

Under ESD conditions the MOSFET enters a snapback mode. To understand how this occurs it is necessary to be aware of the parasitic components in the MOSFET (Figure 1.4).

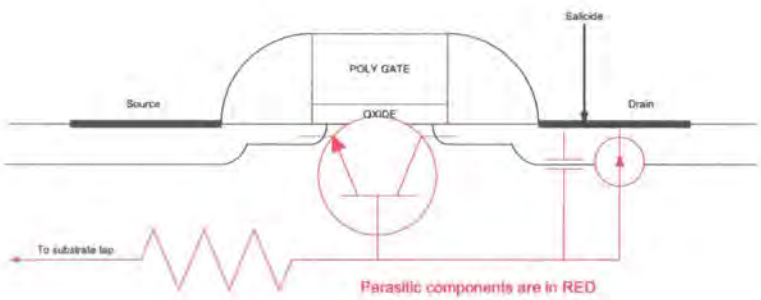

Figure 1.4: Location of the parasitic components in a typical deep sub-micron MOSFET

Duvvury [12], Ochoa et al. [21] and Gaitonde [22] cover the snapback mechanism in detail in a MOSFET when it is subjected to an ESD event. In this explanation it is assumed that the gate, source and substrate are grounded and that the ESD current is passed into the drain, through the device, to ground. Initially, as V_D is increased, the voltage is dropped across the reverse biased drain-substrate pn junction which will eventually start conducting. This conduction mode is called avalanche multiplication (characterised by an avalanche multiplication factor, M) and is due to carrier multiplication in the high electric field in the reverse biased pn junction. The avalanche multiplication factor is an exponential function of the applied voltage across the drain-substrate junction[5] thus as the drain-substrate voltage increases M and $\frac{dM}{dV}$ increases. Any carriers that cross the pn junction are accelerated in the high electric field found in the depletion region and when these carriers collide with electrons in the valance band, the electrons are promoted to the conduction band. This effect creates two carriers, a hole and an electron, which are also accelerated by the electric field and can therefore create further carrier pairs. Once this process has started, it rapidly produces a large number of

carriers. More details on the mathematics and the mechanisms of the breakdown are available in the literature [5][12][22][23]. The avalanche current flows from the drain into the substrate and out of the substrate tap. The substrate has a finite resistance therefore; given enough current, the potential of the substrate region just below the oxide will increase to $\sim 0.7V$. This is enough to forward bias the source-substrate junction which allows electrons to diffuse from the source to the substrate region. As the minimum distance between the source-substrate and drain-substrate pn junctions is short (less than a diffusion length), some of the injected carriers from the source will diffuse into the reverse biased drain-substrate pn junction.

From this, we can see that the effect of the ESD potential across the substrate is to turn on the parasitic bipolar transistor shown in Figure 1.4. The bipolar transistor will then, in turn, inject electrons into the avalanching drain-substrate junction thus decreasing the avalanche multiplication required to produce the base current of the bipolar transistor. As the multiplication factor of a reverse biased diode is voltage dependent, the increased electron current provided by the bipolar transistor gain results in a lower voltage across the diode and therefore across the ggNMOS. This effect is called snapback. Before the ggNMOS is triggered, the avalanching junction needs to multiply the leakage current (a few nA) in the MOSFET sufficiently to turn the bipolar transistor on (a few mA) requiring a multiplication factor of around 10^6. Once the device is triggered the junction needs to multiply I_{CE} of the bipolar transistor ($\sim 100mA$) to produce enough carriers for the base of the bipolar transistor ($\sim 1mA$) requiring a multiplication factor of around 10^{-2}. Therefore once triggered a multiplication factor M of only $\frac{1}{\beta}$ is needed to sustain the current, therefore the voltage across the drain-substrate junction decreases sharply. The decrease in source-drain voltage is further exacerbated by the increase in β caused by the increase in I_{CE} in the bipolar transistor [5]. In summary, the avalanching junction has a greater influence on the breakdown voltage whereas once the device is triggered the parasitic bipolar transistor gain has a greater influence. This has been

illustrated by Ker et al. [24] where they showed that there could be two distinct snapback modes when a highly n^+ doped region underneath the drain doping was added. The initial snapback mode is during triggering, and in this case the current flows deep in to the silicon to the highly doped n^+ region where the drain-substrate junction depletion width is thinner and therefore the avalanche breakdown voltage is lower. Once a significant current starts to flow ($\sim 1mA/\mu m$) and the β of the bipolar transistor increases the current flow moves to the drain-substrate junction where the n-type doping is lower (and therefore the depletion region is wider) as the carrier generation by avalanche multiplication is now less important. Similar conclusions can be drawn from Esmark et al. [25] and Litzenberger et al. [26][27] papers where they found that during triggering the current flowed where the electric field was highest (at the edge of the drain each side of the transistor) and once the transistors had been supplied with just enough current to trigger the device the current would flow where the bipolar gain was highest (in the center of the device).

Figure 1.5 shows the representation of a typical IV curve of a ggNMOS. The voltage required to trigger the transistor is around 6V-12V (known as V_{t1}) and once triggered the drain voltage drops to 4V-6V (known as V_{hold}). The on resistance of the device once triggered is low as an increase in the drain current will result in a marginal increase in V_{DS}. The failure voltage and current are known as V_{t2} and I_{t2} respectively.

1.2.2 Other ESD Protection Structures

Although this thesis does not specifically investigate other types of protection structures, it is useful to know that they exist and how they operate. For example the diode is commonly found as a parasitic device in a ggNMOS and the triggering operation of an SCR can aid understanding of the triggering operation of a ggNMOS.

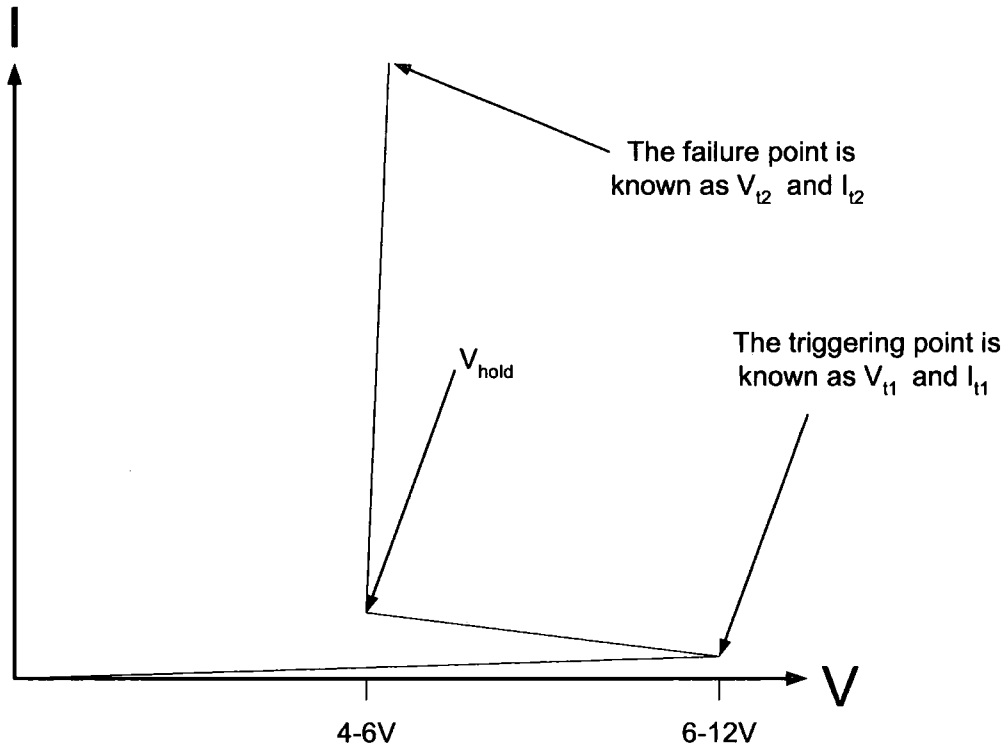

Figure 1.5: IV profile of a ggNMOS under ESD conditions

1.2.2.1 Diodes

Diodes are commonly used in ESD protection as they have a high failure current, low capacitance and low area usage [28] although there are applications where they are inappropriate. Often diodes are found in ESD protection circuits in conjunction with other protection devices; a common example is shown in Figure 1.6 [29].

Diodes are commonly used where they will be reverse biased under normal operation (A in Figure 1.6). Their low forward voltage drop inhibits their use where they will be forward biased under normal operation unless a number of diodes are put in series (B in Figure 1.6). If this is not suitable, then another technology can be used such as a ggNMOS or SCR (C in Figure 1.6).

The diode can be implemented in the substrate [29] or in polysilicon [30] (Figure 1.7). Their HBM failure currents are around 20-50mA/μm (the failure current per unit

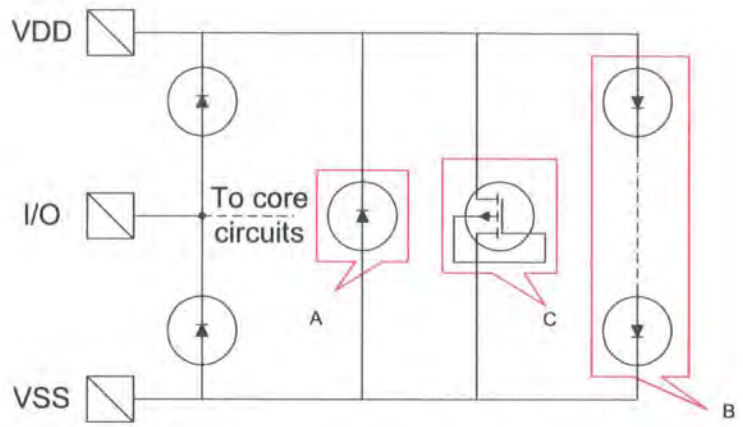

Figure 1.6: Typical ESD protection circuit. (A)Reverse current ESD protection (B)ggNMOS for forward current ESD protection (C)diode string for forward current ESD protection

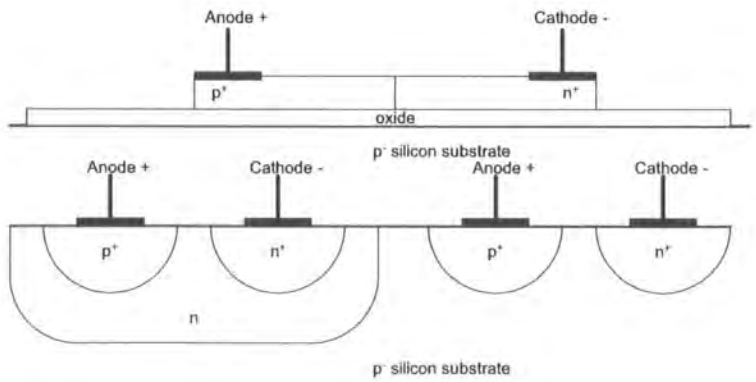

Figure 1.7: Top: Diode made in polysilicon.
Bottom: Diodes made using wells in a p^- substrate

area is around $10mA/\mu m^2$) and the typical on resistance is in the order of 1Ω [31]. The failure current and noise figure for the diode scales linearly [32] but the ESD performance of such devices in reverse bias is very poor [1].

Velghe et al. [33], Leroux et al. [34] and Hyvonen et al. [35] have studied an alternative approach to assessing the impact the diode parasitics have on a circuit by measuring the Q-factor of a diode. They concluded that the Q-factor of a diode was much higher than that of a ggNMOS and did not change with area. They also showed that the capacitive loading effect of a diode can be cancelled out by adding an inductor in series. In this particular instance, high gain and low noise was needed. Vassilev et al. [36] took a third approach. They looked at the S-parameters of the IO protection structures. This information makes it easier to calculate the effect that the protection will have on the overall circuit. Hyvonen et al. [35] also compared different diode types that can be constructed in a deep N-well CMOS process and showed that there was a significant difference in ESD performance relative to parasitic capacitance for different types of diodes.

Richier et al. [37] pointed out that the capacitance of a diode changes with applied voltage but also demonstrated how careful tuning of diode dimensions can minimize the variation in capacitance.

Many authors [1][10][33] show that the ESD performance of a reverse biased diode is very poor, as expected. For a start, the reverse breakdown voltage is much higher than the forward biased diode voltage. Secondly, Pierce [1] also stated that heat generated in a diode is much more concentrated if it is reverse biased, as all the voltage will be dropped across the depletion region. When the diode is forward biased, some of the heat will be generated in the n and p bulk regions due to the high current densities.

Design guidance to reduce parasitic resistances and capacitance is given by Worley

et al. [38]. They suggested that the metal tracks should be tapered such that there is a near constant current density and the spacing between metal tracks should be maximised to reduce capacitance (Figure 1.8).

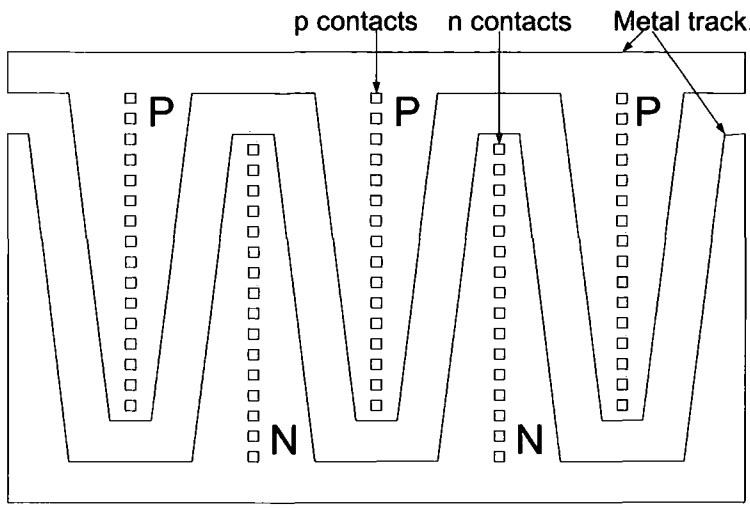

Figure 1.8: Suggested layout of metal tracks to achieve a constant current density and minimise capacitance in a pn junction

1.2.2.1.1 Diode failure current. (I_{t2}) According to Pierce [1] and Worley et al. [31] the failure current (I_{t2}) of a diode is dependent on its perimeter, as the current flows though the sidewalls. Worley et al. [31] added that the parasitic capacitance is a function of area. So overall, the highest perimeter and the lowest area diode would be best. The current has a tendency to crowd at the corners of the diode reducing overall efficiency though Worley et al. [31] have shown that this can be alleviated by decreasing the number of contact vias at the corner of the diode. This advice is apparently contradicted by Worley et al. in the same paper where it states that failures tend to occur in the contacts or the metal tracks, so one should ensure that there are as many contacts as possible and that the metal tracks are as wide as possible.

A forward biased diode is particularly effective in conducting ESD current as the voltage drop across the pn junction is small ($\sim 0.7V$) thus the temperature rise of the junction will be small. The bulk p and n regions will also contribute to the

overall voltage drop but they will not heat up the pn junction. Any heat generated in the diode will be spread out. Pierce's ESD-EOS tutorial [1] adds that, in reverse bias, the voltage drop across the pn junction is higher ($> 10V$) and therefore the junction will reach its melting point at much lower currents creating a short circuit. The reverse biased ESD performance of a diode is very poor and often of no value at all [29].

Smedes et al. [39] do point out that IC processes are designed to maximize device performance, latchup immunity, substrate noise isolation and RF performance, but little regard is taken for ESD. In this particular case, Smedes studied the effect of EPI resistance on the failure current of a diode. The EPI layer is deposited on undoped silicon to provide a thin substrate which has the advantage of reducing cross talk in the circuit. The doping density in the EPI layer determines the substrate resistivity and Smedes et al. [39] found that the failure current is around 10% higher for a low substrate resistivity. In this paper it was found that the IV curve of the diode was unaffected by the substrate resistivity.

1.2.2.1.2 Using a diode to reduce the effective parasitic capacitance of other ESD protection structures Verhaege's tutorial [40] demonstrated that diodes can also be used to reduce the parasitic loading of an ESD protection structure (Figure 1.9). He explained that, under normal operation, the resistor charges up the parasitic capacitance of the SCR so that the diode is reverse biased. As the two structures are in series and the capacitance of the diode is far less than that of the SCR, the capacitive loading is determined by the diode.

1.2.2.1.3 Using multiple diodes in series to increase forward voltage and to decrease effective parasitic capacitance Multiple diodes can been put in series which will have the effect of reducing capacitance and increasing turn on voltage of the protection circuit. B in Figure 1.6 shows a chain of diodes protecting

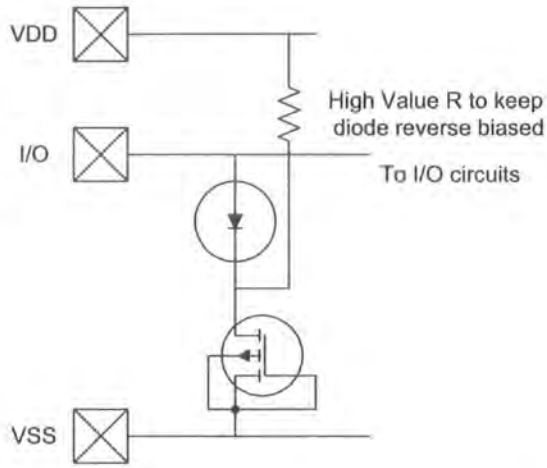

Figure 1.9: The use of the diode to reduce the capacitive load on an IO due to an ESD protection structure

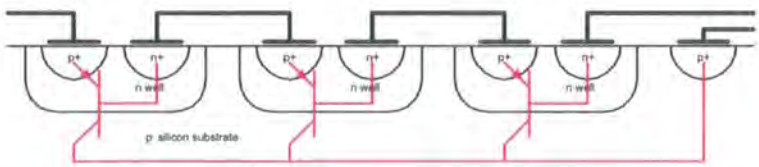

Figure 1.10: The location of the parasitic bipolar transistors found in a series of well diodes. These parasitic structures amplify any leakage current produced by the diodes.

the power supply bus from positive ESD pulses ($V_{DD} > V_{SS}$). This protection technique is likely to be very area efficient and have low parasitic capacitance, but is prone to high leakage during circuit operation at high temperatures. Maloney et al. [41] explain that this is due to the decrease in forward voltage of a diode as the temperature increases. They go on to explain that the leakage current is further amplified by parasitic bipolar transistors which occur when a chain of diodes are placed in close proximity of each other (Figure 1.10). Any leakage from the diode on the right will be multiplied by the transistors on the middle and on the left. The gain of the transistors will also increase exponentially for a given voltage (across the base-emitter junction in the parasitic bipolar transistor) as the temperature rises. Many solutions [41][42][43][44] have been proposed that reduce the 'leakage multiplication' problem.

The leakage current can be reduced by increasing the number of diodes in the chain

though this may result in an unacceptably high voltage across the protection structure during an ESD event.

Wang et al. [30] and Ker et al. [45] also reported 'leakage multiplication' in substrate diodes and investigated polysilicon diodes as an alternative which has the advantage that they do not suffer from 'leakage multiplication'. They found that the ESD performance of poly diodes was poor and a silicide block had to be applied to prevent the silicide shorting the diode. Ker et al. [45] solved the poor ESD performance of the poly diodes by modifying their doping profile.

An obvious alternative to a diode string is a zener diode, though van Dalen et al. [46] found that low voltage zener diodes have a high capacitance and that the on resistance is high. One must also remember that it is very difficult to implement a zener diode in a standard CMOS process without altering the manufacturing parameters. Ker et al. [45] have demonstrated that a chain of diodes can be used to trigger other protection devices. In this paper, they used the diodes to substrate pump a ggNMOS thereby lowering the triggering voltage of the ggNMOS (Figure 1.11). This technique is used in a similar circuit described in section 1.3.3.

Diodes have been shown to be useful where low capacitance ESD clamps are required, though their high leakage current at high temperatures and moderate forward voltages makes them unsuitable for IOs that have a DC voltage greater than $\sim 200mV$. This makes diodes suitable for low DC voltage RF IOs but the leakage current rules out their use for power supply clamps. The leakage current could be reduced by using small diodes to trigger other larger ESD protection clamps.

1.2.2.2 SCRs

SCRs have been shown [1][29][47][48][49][50] to have a higher ESD current handling capability per unit area than other types of protection. Caillard et al. [51] specifically

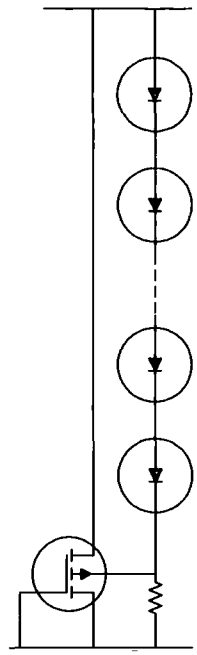

Figure 1.11: The use of diodes to pump the substrate of a ggNMOS to assist triggering

mention a current handling capability of over $76mA/\mu m$. The structure is also known for its low parasitic capacitance [50], low hold voltage (~1-3V) [1][48][49], good width scaling [48] and low on resistance [48][49].

Although this component is robust, there are serious problems with its ability to protect MOSFET IO circuits. The high trigger voltage is of particular concern [1] [29][47][50][52] and the slow triggering speed further exacerbates this problem though Kunz et al. [52] showed that the trigger voltage would fall for a high $\frac{dV}{dT}$ which would alleviate this problem. The low holding voltage of an SCR is an advantage when it comes to reducing the power dissipation in the component for a given ESD current. The disadvantage is that if, as is usually the case for SCRs, the power supply voltage is greater than the holding voltage, the output driver (Figure 1.1) could, if driving a high output voltage, keep an accidentally triggered SCR in a triggered state. Accidental triggering can be caused by noise or a by spike on the power supply bus or on the IO pin.

The literature is full of proposed solutions to these shortcomings. Kunz et al. [52]

have shown that the dimensions of the SCR are critical to the trigger voltage and, if the dimensions are minimised, the triggering voltage will be minimised. Small dimensions also help speed up the triggering time [47].

Hybrid solutions have been proposed by Morishita [50] where a ggNMOS or a diode string is used to trigger the SCR and the SCR conducts the majority of the ESD current (Figure 1.12).

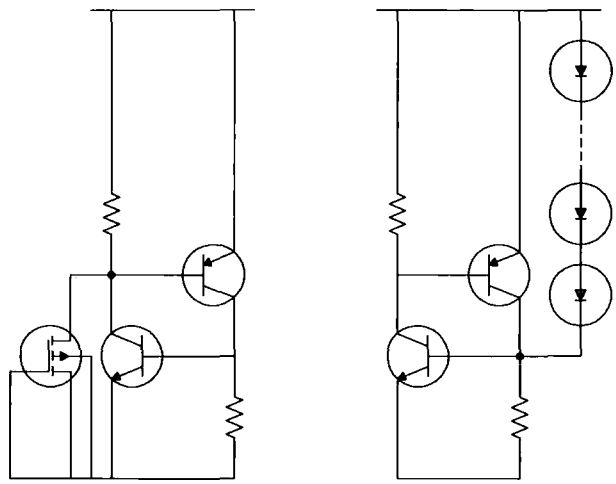

Figure 1.12: Left: This SCR is triggered by a small ggNMOS. Right: This SCR is triggered by the diode string

Caillard et al. [51] have also proposed a hybrid solution with the added feature of latchup immunity. It uses an RC circuit coupled with a NOT gate to trigger the SCR (Figure 1.13). The triggering voltage is variable and rise time dependent. This circuit also has the advantage of being able to trigger multiple SCR fingers.

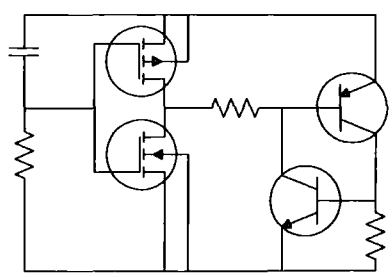

Figure 1.13: In this solution an RC trigger is buffered to turn on an SCR

Mergens et al. [48] have provided a solution to the problem of latchup. It does this by increasing the minimum current required to keep the SCR in a latchup state. This

means that, if it is accidentally triggered, the latchup state will stop as the chip's power supply is unable to provide sufficient current to sustain the latchup state. This is achieved by decreasing the resistance between the base and the emitter of the npn transistor, thus more current will flow directly to ground, (A) in Figure 1.14, and less current will flow into the npn transistor (B). A higher total current (C) will be needed to keep the SCR in a triggered state. It is known that at low current levels, the gain of a bipolar transistor falls, so a higher minimum current will be needed to keep the SCR in a triggered state [5].

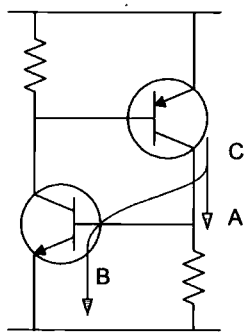

Figure 1.14: The operation of an SCR. If the resistance from point (A) to ground is decreased, less current will flow in to npn transistor base (B). Therefore the collector current from the pnp (C) will have to be increased to keep the SCR triggered.

Knowledge of how SCRs operate is important: Duvvury et al. [53] have demonstrated that some ESD failures are due to parasitic SCRs triggering and then failing during an ESD event. An alternative structure based on the SCR has also been designed and tested by Feng et al. [54]. It is designed to provide ESD protection between any three terminals, for example an IO, V_{DD} and V_{SS} terminals, in any polarity. Its operation is based around an SCR and only one device is needed for each IO. It therefore provides a low capacitance and lower area consumption solution than the traditional solution that uses a separate protection structure for each pair of terminals.

1.2.2.3 Bipolar Transistors

Bipolar transistors can be used as ESD protection devices either with the base shorted to ground (Figure 1.15:A), or with the base connected to ground via a resistor (Figure 1.15:B) or, in some cases, with a triggering mechanism (Figure 1.15:C&D).

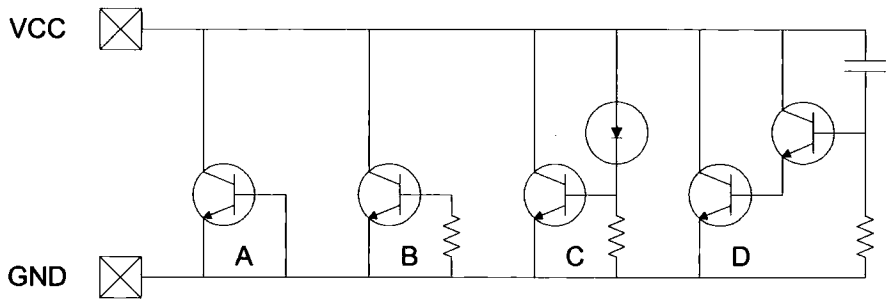

Figure 1.15: Three ESD protection solutions using bipolar transistors

The mode of operation of the bipolar transistor ESD protection is the same as that in a ggNMOS as explained in section 1.2.1. Voldman et al. [55] have used the bipolar transistor to provide a high Q, low capacitance and low noise ESD protection for RF circuits. SCRs and MOSFETs were not good enough; diodes were used to trigger a darlington transistor and this was found to be seven times more area efficient than a MOSFET based protection. Ma et al. [56] have also used bipolar GaAs transistors for RF applications as SCRs are too noisy, capacitive and the Q factor is too low. They did consider using diodes but the diodes series resistance, per unit area, is too high to keep the voltage low during ESD discharge. The problem was solved by triggering a BJT or a darlington transistor using diodes (Figure 1.15:C). In a further paper, Ma at al. [57] reported that the diode solution worked well but had the drawback of high leakage and a non linear impendence causing distortion, they therefore investigated a darlington, of both 2 and 3 transistor construction (Figure 1.15:D), which were shown to have good ESD and RF characteristics. The 2 transistor darling solution was found to have a better ESD performance and the 3 transistor darlington solution a better RF characteristics.

1.2.2.4 MOSFETs Operating in the Linear Region

Section 1.3 explains that reducing the triggering voltage (V_{t1}) improves the current spreading and increases the maximum current handling capability (I_{t2}). One of the methods used to decrease V_{t1} was to increase V_{GS} during an ESD event by allowing the gate to float [58] or to drive it high [59][60][61]. Merrill et al. [16] proposed that this should be taken to an extreme, i.e. drive $V_{GS} >> V_T$ so that the MOSFET conducts as an ordinary switch thus preventing any avalanche multiplication or snapback.

There are three parts to the circuit shown in Figure 1.16:

A triggering circuit. This increases V_{GS} when an ESD event is detected. The most common trigger circuit is an RC ($t_{RC} \sim 1\mu s$) transient trigger circuit. The capacitor is made using the gate of a MOSFET and the resistor can be made using an n-well or a small MOSFET with the gate tied to the drain. One can also use diodes or MOSFET clamps to trigger the circuit.

A buffer: This is needed to drive the large gate in the BigFET.

A BigFET: This is a large MOSFET $\sim 8000\mu m$ wide [16] which is used to conduct the ESD current.

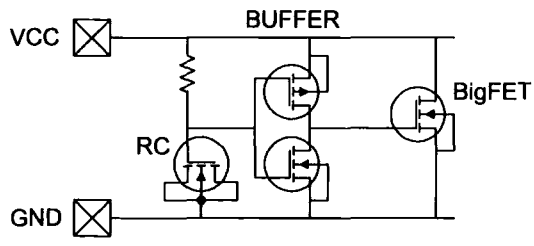

Figure 1.16: The basic concept of a BigFET ESD protection

Researchers at Texas Instruments have improved the design [62] by converting the buffer into a latch which is achieved by adding feedback to the buffer (Figure 1.17). The advantage of this solution is that the t_{RC} of the RC circuit can be reduced.

The feedback loop will ensure that the BigFET will stay on so long as there are a few volts on the power supply. This will reduce the area consumption and increase the immunity to false triggering during power up (due to the reduced t_{RC}). Li et al.[63][64] addressed a downside of the feedback circuit where false triggering could cause the power supply to be shorted resulting in high current consumption and possibly catastrophic failure during normal operation. They proposed two modified circuits (Figures 1.18-1.19) and showed that they had good ESD performance and did not suffer from false triggering during normal operation. The first circuit uses feedback to accelerate the turn on of the BigFET once it starts to be triggered. The second solution provides two paths for the RC signal to reach the BigFET; a fast one to turn it on when an ESD event is detected and a second, slower, path to turn it off once the ESD event is complete.

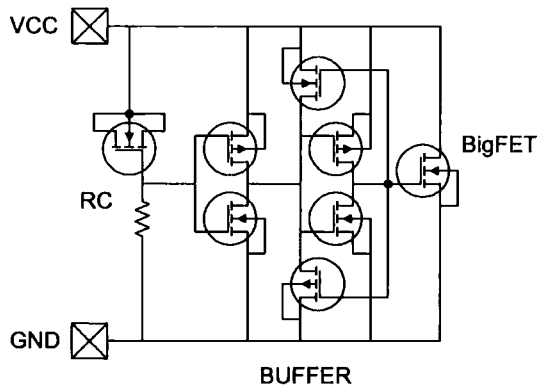

Figure 1.17: BigFET ESD solution with positive feedback to assist triggering

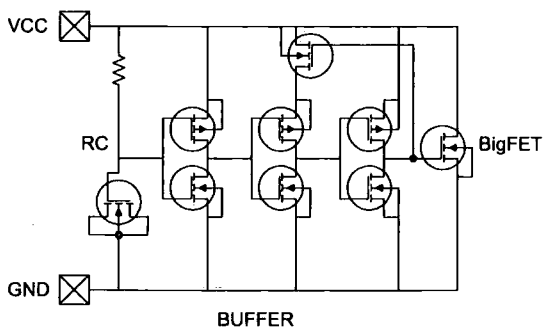

Figure 1.18: BigFET ESD solution with modified positive feedback to assist triggering and prevent power supply latchup

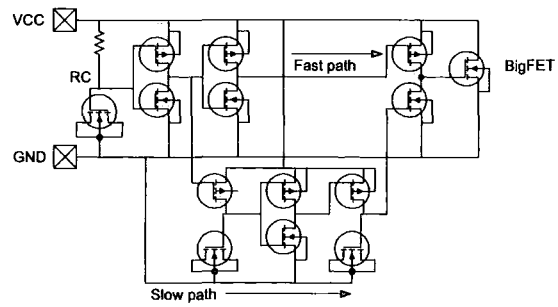

Figure 1.19: BigFET ESD solution with modified RC triggering circuit to minimise area consumption of the MOSCaps and eliminate the requirement for positive feedback

Intel Corp. [28][65] are using pMOS for the BigFETs which is counter intuitive as it is well known that nMOS can carry more current than pMOS transistors due to the difference between electron and hole mobility [5]. This can be explained by other research that has shown that pMOS devices are much weaker snapback devices than nMOS devices and therefore less likely to go into snapback mode. If a BigFET is turned on weakly, e.g. when V_{GS} is low, it is susceptible to snapback and to all the damage that snapback causes to standard MOSFETs as explained in section 1.2.1.

Maloney et al. [65] also looked at ways to reduce the leakage current in the BigFETs for battery powered applications where a low standby current is required. The trend for smaller gate lengths and lower threshold voltages means that leakage currents are high. This is exacerbated by the large gate widths of BigFETs. Recently Intel has used *back biasing* to reduce the leakage current in the pMOS devices whilst the chip is in a low power mode. This involves biasing the substrate of the pMOS devices above VCC which in turn increases V_T of the devices and reduces I_{leak} as shown in Figure 1.20. The biasing reduces the sub threshold conduction but does not address the band to band pn junction tunnelling and gate oxide tunnelling. The *back biasing* is undesirable for ESD protection BigFETs during an ESD event as this reduces the current drive capability.

Intel have proposed a circuit shown in Figure 1.21. The NOT gates 1 and 3 drive the pMOS BigFET and NOT gate 2 pulls the substrate tap to V_{CC} during an ESD event

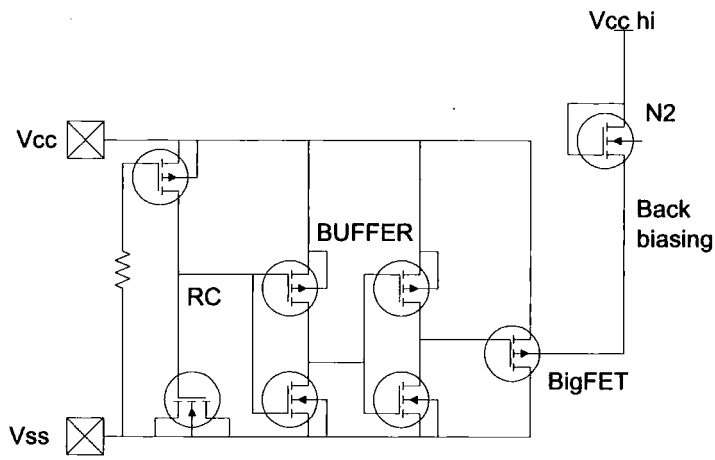

Figure 1.20: The use of *back biasing* to reduce leakage current in a BigFET

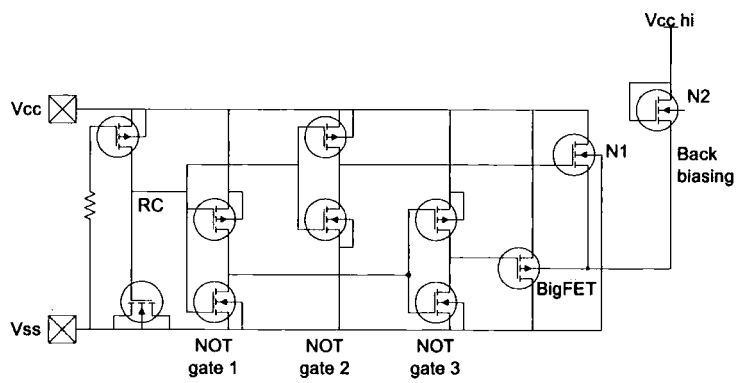

Figure 1.21: *Back biasing* can be disabled during an ESD event to improve ESD performance

using N1. N2 is used to connect the substrate tap to a higher VCC for low power operation. This technique has been proven to be as good as the unbiased BigFET but the leakage has been reduced by 70% – 80%. Further designs are proposed where cascoded devices are used to reduce leakage current where *back biasing* is not available.

Poon et al. [66] have also provided solutions to the excessive leakage found in MOSFET based capacitors in RC circuits. The leakage prevents the capacitor from charging up fully resulting in the input of the NOT gates not being at the power rails. This increases the leakage current of the NOT gate. The solution proposed is a feedback loop which, once the RC capacitor is charged, turns on a second pull up pMOS so that the capacitor is fully charged (Figure 1.22). This results in a decrease in the I_{leak} of the NOT gate.

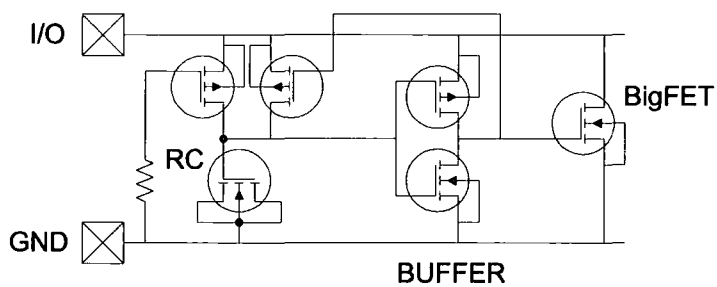

Figure 1.22: Feedback can be used to reduce leakage whilst the protection structure is inactive

Researchers at Motorola [67][68] implemented BigFETs on a large scale spanning multiple IOs using a single protection structure. They found that it was more area efficient and easier to predict the performance as the circuit could be simulated using SPICE. Detractors of this approach might point out that the ggNMOS transistor, during snapback, has been measured as four times more effective than the BigFET transistor under ESD conditions [69]. To get around this Motorola used a single circuit based on the BigFET principle to protect a whole bus of IO pins so that the complete design is more area efficient.

Originally they tried to use one large clamp to protect all the IOs. The fundamental mode of operation is to divert the ESD current on to the power supply bus via diodes and then put a single ESD protection circuit on the bus to conduct the current to ground (Figure 1.23).

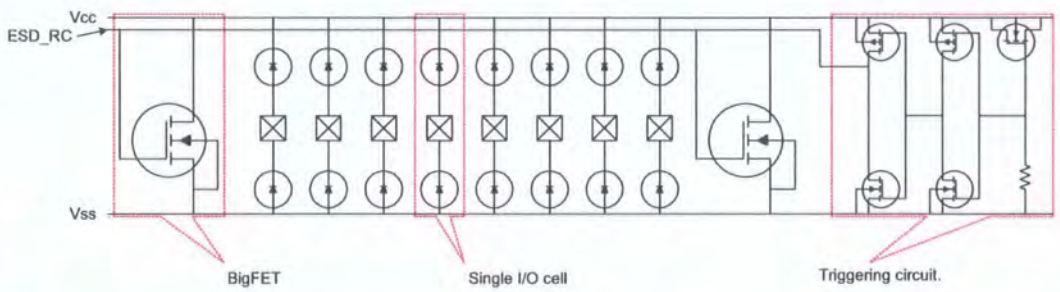

Figure 1.23: The use of a few BigFETs to protect a large number of IOs

The aim is to save silicon area by having only one protection structure for many IO pins. Motorola used SPICE simulation to show that the bus resistance was such that pads positioned far away from the protection structure went over voltage during ESD events.

The solution proposed by Torres et al. [67] was to put a small BigFET protection structure at each IO and to add a large BigFET protection structure at the V_{DD} and V_{SS} pads where there are no output drivers (Figure 1.24).

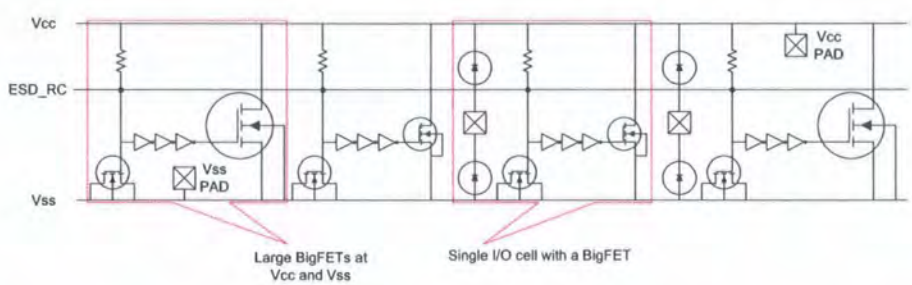

Figure 1.24: The distribution of BigFETs along a chain of IOs to mitigate the effects of bus resistance

They also showed that a more area efficient variant would work where the ESD detection is carried out by a single RC circuit located in the V_{DD} or V_{SS} power pads. The output signal from the RC circuit is used to turn on all the BigFETs

simultaneously. A trigger line added to the pad design carries this trigger signal to each individual BigFET (Figure 1.25).

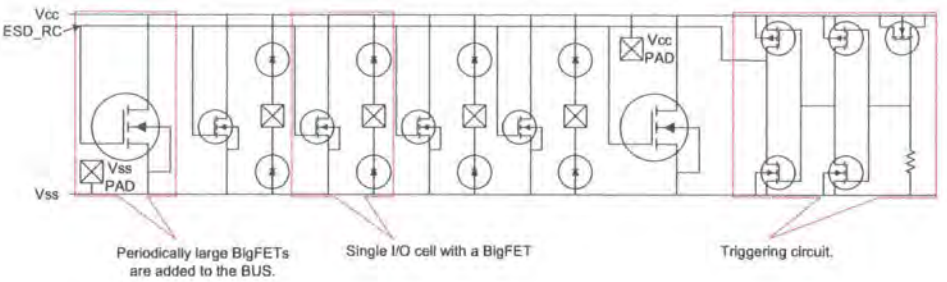

Figure 1.25: The triggering of distributed BigFETs by a single ESD detecting circuit

SPICE simulation showed that the maximum voltage of a given pad was not dependent on its location as the distributed BigFETs negated the effect of the V_{DD} and V_{SS} bus resistance. This increased the ESD performance of the chip as a whole.

Further refinements to this solution has been carried out by researchers at Motorola. In the original design [67] the power rail carries the ESD current; and because of the power bus resistance the voltage available to drive the BigFET's gate is reduced which decreases the current carrying capability of the BigFET. A separate power rail is added to power the triggering circuit [68]: as the current levels are low the voltage drop across this bus is low. Overall this means that the BigFET is driven harder during an ESD event (Figure 1.26).

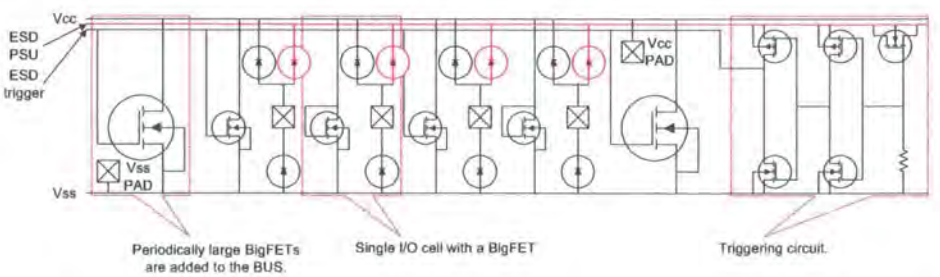

Figure 1.26: The use of a separate ESD detection circuit power supply bus to mitigate the effect of bus resistance

The researchers have also shown that size of the distributed BigFET can be optimised [67] using SPICE to maximize ESD performance and minimize area consumption for all discharge combinations. This design technique has the advantage

of filling vacant spaces in the IC with extra distributed BigFETs thus making more efficient use of the die area.

In some designs there will be some free space, for example between pads, this area can be taken advantage of thus allowing the size of BigFETs elsewhere in the circuit.

Solutions have also been proposed to prevent the inadvertent triggering of protection circuitry due to hotplugging (this is where two circuit boards are connected together whilst one of the boards is powered). Hotplugging causes fast rise times to occur in the power supply [67] which can activate the RC triggered circuits. A solution would be to have a shorter RC time constant for the RC circuit so that only the fastest rise times would be detected. The problem with this approach is that the RC time would be so short that it would not keep the BigFETs fully on for the duration of the ESD event. The solution proposed is to trigger a monostable using the RC circuit so that the BigFET is kept on for the duration of the ESD event.

Stockinger et al. [68] argue that, given these new techniques, the BigFET ESD solution is more area efficient than the ggNMOS solution. The protection structure is easier to simulate as SPICE simulation is sufficient and finite element simulation is not required. It is process portable and independent as the circuits are designed based on the SPICE models provided by the foundries. ESD solutions based on avalanche breakdown (ggNMOS and SCRs) are very process sensitive as the trigger voltage is dependent on avalanche voltage and the β of the parasitic transistors in the device. Therefore process optimisation is required to improve ggNMOS/SCR operation. Miller [29] supports these arguments and adds that the BigFET designs may add leakage and capacitance to the circuit. The protection structure may be susceptible to false triggering and the concept is more complicated than ggNMOS based solutions to implement. On the positive side he adds that the circuit should be easily converted to new processes.

1.2.2.5 n-well Resistors

n-well resistors can be used to improve the characteristics of other ESD protection devices such as the SCR. The SCR has a high I_{t2} value but its triggering voltage is high, as a consequence any MOSFET that it is protecting will be damaged before the SCR triggers. A ggNMOS can be used to protect a MOSFET but its I_{t2} value is lower; therefore the ggNMOS needs to be larger to provide the same level of protection. A hybrid solution [12] which combines the I_{t2} of an SCR with the V_{t1} of a small ggNMOS using a resistor is shown in Figure 1.27.

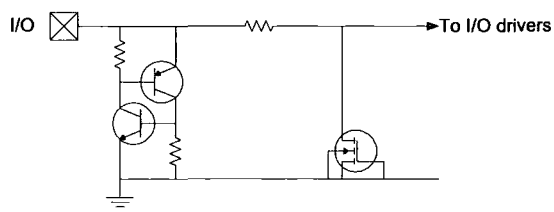

Figure 1.27: The use of an n-well resistor, SCR and a ggNMOS to create a hybrid ESD protection solution

The SCR diverts the majority of the ESD current and the n-well resistor limits the current into the ggNMOS which as a consequence can be small. The ggNMOS limits the maximum voltage seen by the IO transistors.

In this case, the resistor needs to be small to minimize area consumption, parasitic capacitance to ground and attenuation of the IO signal but it also needs to be sufficiently large to prevent permanent damage to the ggNMOS. Worley et al. [31] have studied the IV characteristics of n-well resistors and found that at low voltages the resistor is ohmic. At high voltages the current remained the constant irrespective of the applied voltage or resistor length within practical limits. This is due to the saturation velocity of the carriers in the resistor. If the applied voltage is increased further still, the number of carriers will increase due to the high electric field causing impact ionisation [5]. The maximum voltage is limited by secondary breakdown and the breakdown voltage of the n-well p substrate diode. At this point, filamentation

occurs where Joule heating increases the carrier density which increases local current density creating localised hot spots and ultimately causes the silicon to melt [12].

1.2.2.6 LRC Methods

The LRC protection method is a neat concept that takes advantage of parasitic capacitances in an IO circuit. The basic principle is that in all IO circuits there will be some parasitic capacitance to ground. An inductor is added to the circuit in parallel to the parasitic capacitance (Figure 1.28) and is tuned to the operating frequency of the IO. The IO signals at the resonant frequency of the LC circuit will pass through while all other signals (including ESD currents) will be shorted to ground. HBM ESD signals have frequency content that extends into the low MHz [1][70] and it will all be shorted to ground. This technique has two effects: the first is that it counteracts any parasitic capacitance and secondly it shunts HBM ESD signals to ground.

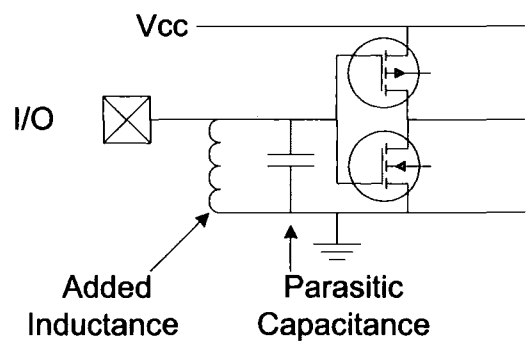

Figure 1.28: The use of an inductor as an ESD protection device

This principle has been demonstrated by Hyvonen et al. [70] in order to protect narrow band RF circuits that operate at 1GHz. They showed that this is an effective solution for HBM ESD but it did not work for CDM. They also found that if a ggNMOS was added, in parallel with the capacitor, it would damp the LC resonator during CDM ESD events improving the performance.

Vassilev et al. [71] also investigated LRC methods and showed that it was effective in protecting the input of an LNA as well as reducing the effect of the parasitics. Thijs et al. [72] used inductors though they were able to demonstrate that an inductor could be added without altering the LNA circuit for matching whilst at the same time achieving good ESD performance though they also warned that the series DC resistance of an inductor needed to be taken in to account as this, in some cases, caused significant voltages to be applied to gate oxides and therefore result in transistor failure.

1.2.2.7 Vias, Contacts and Metal Tracks

Vias, contacts and metal tracks connect all the components of a circuit together, they are as important as the individual components. Like components they can suffer latent and permanent damage due to ESD events [10].

It is well known [18] that the high current density found in metal tracks of ICs can cause electromigration issues. Electrical failure can happen in one of two ways: 1. A break on the track causing an open circuit; 2. A short circuit to an adjacent line due to the accumulation of material. Wolf [18] also adds that Cu is around 10 times more resistent to electromigration than Al.

Researchers at Chartered Semiconductor [73] found that the crystal structure of Cu was changed when a TLP pulse was put through it. They showed that this reduced the lifetime of the metal tracks in accelerated lifetime tests. The failure was due to electromigration.

Pierce [1] carried out mathematical modelling of electro-thermomigration of aluminium contacts that shorted out the pn junction in the silicon. In some situations, ESD pulses caused the metal in aluminium contact to migrate into the silicon and through the pn junction causing a short across the junction. He concluded that

electromigration caused the Al atoms to move into the silicon and that it was not a thermal effect.

1.2.2.8 Metal track Failure due to IR Heating

Smedes et al. [74] and Tan et al. [75] found that the failure current of metal tracks scaled linearly with width. Smedes et al. added that there was an offset which could be attributed to heat loss at the edges of the track. Wunsch Bell models have been used by Pierce and Durgin to correlate current pulse length to thermal failure [1][10][76]. They argued that if the pulse was very short, the heat dissipation would be adiabatic and therefore, all the heat generated would be absorbed within the thermal heat capacity of the metal. In the adiabatic region, the failure energy is constant regardless of pulse width or amplitude. At the other extreme, for long pulses, the heat is dissipated away as fast as it is generated; it is a steady state thermal dissipation. In this case, the failure power is constant regardless of pulse width or amplitude. In the middle there is a smooth transition region which is partially adiabatic. See Figure 1.29 for the results of the practical experiment and Figure 1.30 for the Wunsch Bell models.

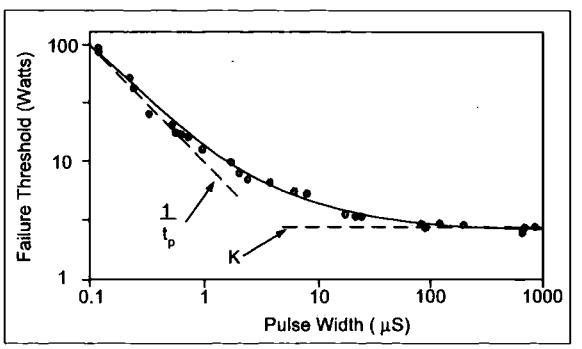

Figure 1.29: Variation of the failure power threshold in relation to the pulse width for a metal track [1]

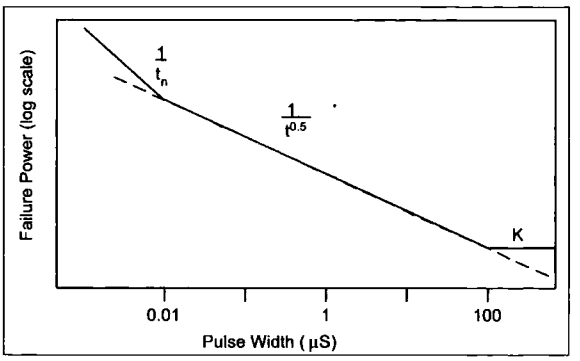

Figure 1.30: The Wunsch-Bell curve for failure due to IR heating [1]

1.2.3 Overall Protection Strategies

Sections 1.2.1 and 1.2.2 discuss individual components that are able to protect an individual IO circuit though, as illustrated in *MOSFETs operating in the linear region* (section 1.2.2.4, Page 29), a more holistic approach can reap dividends. For example the BigFET section explained how a series of IO could be protected by a single BigFET circuit giving rise to considerable reduction in area consumption.

Ito et al. [77] and Ker et al. [78] have demonstrated a variant of the LC protection for RF circuits. Standard protection structures (such as SCRs) would ordinarily severely attenuate any RF signals due to their parasitic capacitance. In this solution, the protection structures are split up and separated by inductors creating a transmission line (Figure 1.31). Ideally, the parasitic capacitances should be split up into an infinite number of small parts separated by an infinite number of inductors; in practice a four inductor solution was demonstrated to be effective. Ito at al. [77] also showed that impedance matching, using quarter-wave transformers, could be achieved using the same technique.

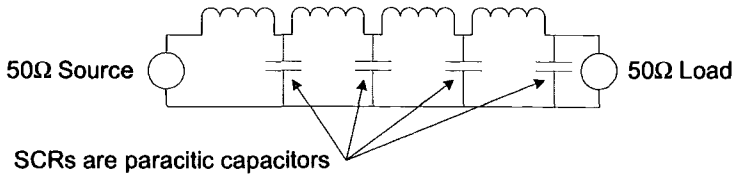

Figure 1.31: The use of inductors to mitigate the capacitive loading effect of a SCR to create a transmission line

1.3 Improving the Snapback Mechanism

The MOSFET mechanism described in section 1.2.1.4 which is taken advantage of in a ESD protection structure called the 'grounded gate nMOS' (ggNMOS) will operate as a weak ESD protection device when added in parallel to any IO structure. Many problems have been reported and solved in the literature and tweaks to the basic design illustrated in Figure 1.32 have been tried and proven over the years. Salling et al. [69] have shown that ggNMOS can, during snapback, conduct four times the current of a standard nMOS switch without suffering damage.

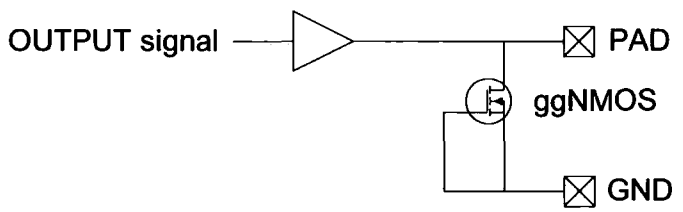

Figure 1.32: Basic ggNMOS design using a standard nMOS

1.3.1 Width Scaling

The most common problem discussed in the literature is width scaling. This occurs when the width of the ggNMOS, or two ggNMOSs in parallel, is not proportional to the maximum ESD voltage. This suggests that the ESD current does not flow uniformly across the width of the device and it may cause localised hot spots and ultimately premature failure. This problem manifests itself both within a transistor as

well as between separate fingers of the same transistor [27][58][59][61][79][80][81][82][83][84].

Polgreen et al. [85] confirmed that silicided ggNMOS transistors do not scale with width, have low minimum failure currents (I_{t2}) and the variance in the I_{t2} values is high. However, for device widths of less than $2\mu m$, the I_{t2} values did scale with width which would suggest that the width of the conducting channel in a standard nMOS (wired up as a ggNMOS) is around $2\mu m$. It must be pointed out that this problem has been solved in the past [86]. However it is important to ensure width scaling occurs as efficiently as possible (minimum area consumption, minimum parasitics and leakage) and to explore new techniques in light of the constantly evolving semiconductor technologies. Esmark et al. [25] and Litzenberger et al. [26][27] studied the current flows though a single finger ggNMOS using back side laser interferometry and 3D finite element simulation and found that during triggering the current was concentrated at the interface between the drain and gate at the edge of the transistor where the electric field is highest. Once triggered, the current tends to flow at the center of the device for low currents and as the current is increased the current flow becomes uniform.

It is widely accepted [2][59][61][83][84][85][87][88] that if $V_{t2} > V_{t1}$ then the ESD current will spread across the width of the protection device. The reason for this can be understood by looking at the operation of a number of identical ggNMOSs in parallel. All the devices will have slightly different geometries and doping profiles due to the natural variation in the wafer manufacturing process. Each protection structure will therefore have a slightly different V_{t1} causing one device to trigger before all the others. As the current is increased, all the current will be conducted by the single triggered transistor and the voltage across the protection structures will increase. If V_{t2} is reached before V_{t1} (because in this case $V_{t2} < V_{t1}$), the single transistor will fail. Conversely if $V_{t2} > V_{t1}$, the other protection structures in parallel will be triggered, and will conduct ESD current, before the individual

transistor fails. Therefore once V_{t2} is reached, all the ggNMOSs will be conducting and the protection structures will be conducting the maximum amount of current. A single wide ggNMOS may be treated as a number of thin ggNMOSs operating in parallel.

Koizumi et al. [61] showed that further improvements in I_{t2} could be achieved by increasing the margin of V_{t2} has over V_{t1}. Researchers at Texas Instruments [20][89] also argued for a low V_{t1} by showing that the HBM failure voltage is strongly inversely correlated to the snapback voltage, V_{t1}. Gupta et al. [89] added that decreasing the gate length increases β which in turn decreases the required M to sustain snapback, resulting in a lower V_{hold} and therefore a lower power dissipation. Researchers from Texas Instruments and Infineon Technologies [25][90] have also used electrothermal simulation to show that on triggering, by far the highest current density is found in the center of the device. They went on to state that the Negative Temperature Coefficient (NTC) of the avalanche multiplication phenomena causes the current to spread out, though the effect of the NTC is limited as the speed of the avalanche multiplication spreading is $\sim 1um/ns$ which is not sufficiently fast to prevent silicon melting. Tan et al. [75] and Bychikhin et al. [91] have used backside laser interferometry to show that current tends to crowd into the center of the device. Tan et al. go on to explain that as the conducting channel heats up, the level of avalanche multiplication in the conducting channel decreases due to the negative temperature coefficient of impact ionisation. This would cause the current to spread out where it not for the stronger positive temperature coefficient of β causing the current to crowd. This is further exacerbated by the fact that once the silicon gets hot to the point where the intrinsic carrier density exceeds doped carrier density the carrier density increases sharply ultimately resulting in failure [25].

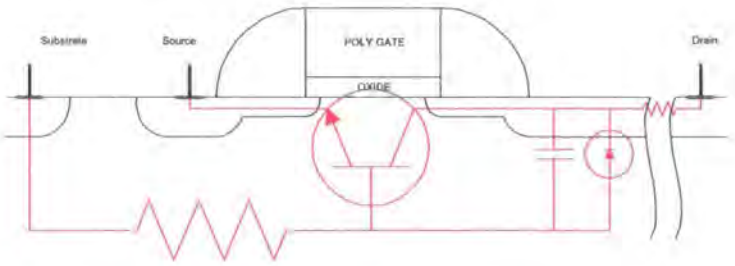

Figure 1.33: The location of parasitics and drain extension design modification

1.3.2 Methods for Increasing V_{t2}

The established solution to the problem of width scaling is to increase the drain length by $1-4\mu m$ (Figure 1.33) and to prevent silicide deposition on the drain (silicide is deposited on the MOSFETs to reduce the resistance between the source/drain contacts and the gate to improve the switching speed of the transistor). The added drain length is designed to act as an n-well resistor in series with the ggNMOS. The IV curves of the n-well, the unmodified ggNMOS and the modified ggNMOS are shown in Figure 1.34. Esmark et al. [25] have shown that this is not as straight forward as it sounds since the slope of the ggNMOS IV curve once it is triggered (caused by the drain resistance) is lower than the resistance measured, at low currents, in an n-well resistor. This is caused by conduction modulation which occurs when a large number of carriers, generated by avalanche multiplication, overwhelm carriers provided by the dopants thus reducing the semiconductor's resistivity. Gauthier et al. [88] showed that silicide blocked ggNMOS transistors scale well with width, whereas silicided transistors do not except with fast rise time TLP tests where the performance of the silicided transistors is improved. This result also demonstrates that a fast rise time makes the current spread out across the width of the device. The paper showed that ggNMOS protection circuits can give consistent protection even if the process changes.

The solution works well though the research community has tried to reduce the

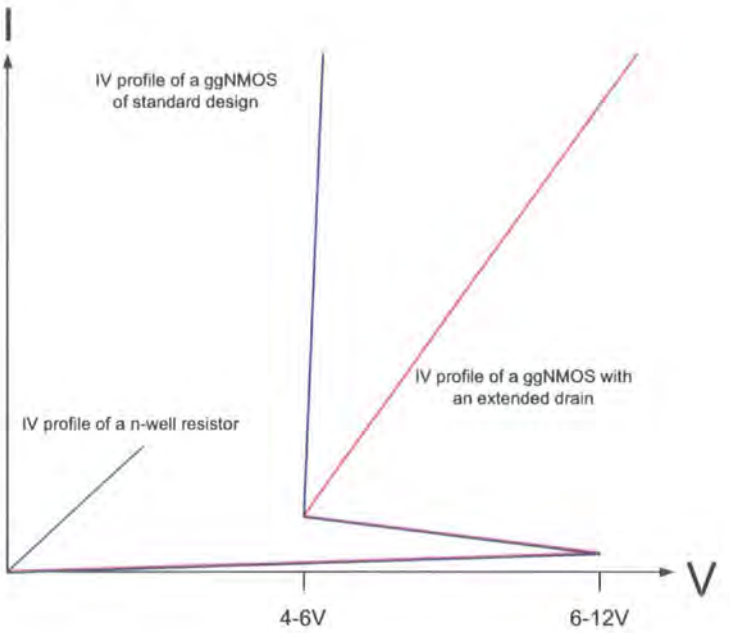

Figure 1.34: The IV profile of an unmodified and modified (drain extended) ggNMOS

silicon area consumed by the ggNMOS as well as the parasitic capacitance and alter the triggering voltage of the device whilst maintaining a high I_{t2}.

Sarnoff Corp. have proposed a number of alternative solutions to adding resistance to the drain of the ggNMOS. They demonstrated that if the resistor was on the source side it would be equally effective. The first device of this kind was published in 2000 [2] and it uses vias, metal tracks and poly resistors to provide resistance to the source and drain of the nFET. The use of these parts means that the source and drain resistors are not planar (Figure 1.35) and therefore the structures have a higher resistance per unit area compared to the extended n-well drain resistor. The advantage of this circuit is that the extended drain resistor in the silicon (Figure 1.33) is eliminated, which means that less silicon area is used and the drain-substrate parasitic capacitance is reduced.

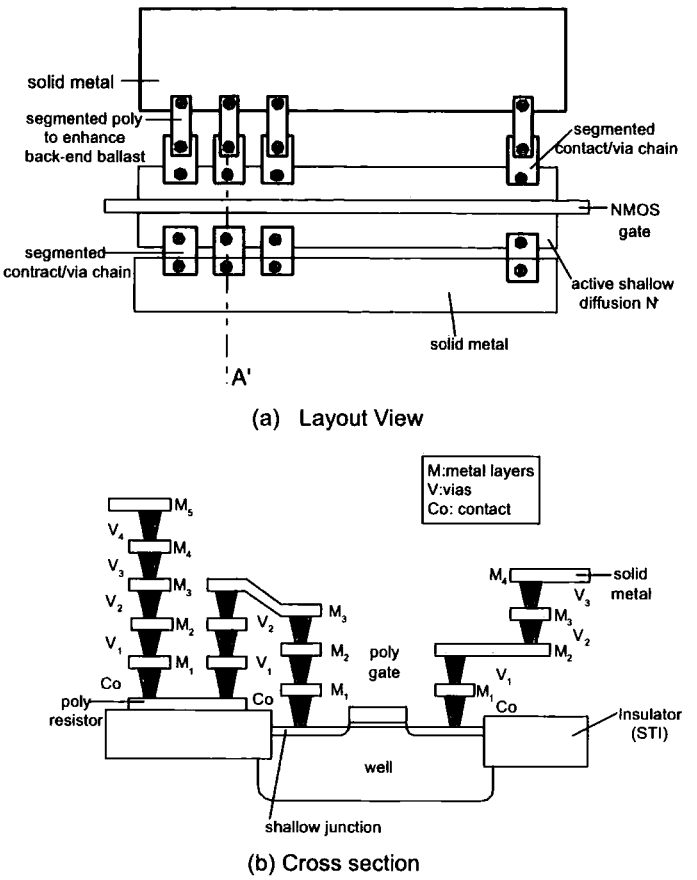

Figure 1.35: Use of vias, metal tracks and poly resistors to replace the drain resistor in a ggNMOS [2])

A second solution was published in 2003 [3] proposing the separation of the drain resistor into strips so that the effective resistance per unit area was increased (Figure 1.36). The authors also demonstrated poly resistor fingers (Figure 1.37). See Figures 1.35 and 1.36 for these hybrid solutions.

Alternative solutions have been demonstrated by Richier et al. [37] showing that the drain parasitic capacitance can be reduced by adding an n-well doping under the SD doping. Kim et al. [92] block the silicide with a gate and use an n-well to produce a drain resistor.

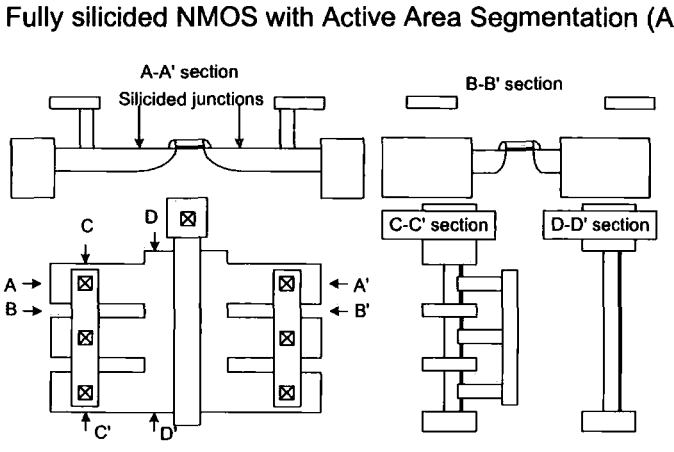

Figure 1.36: Modified drain extension resistor in a ggNMOS used to increase the resistance per unit area [3]

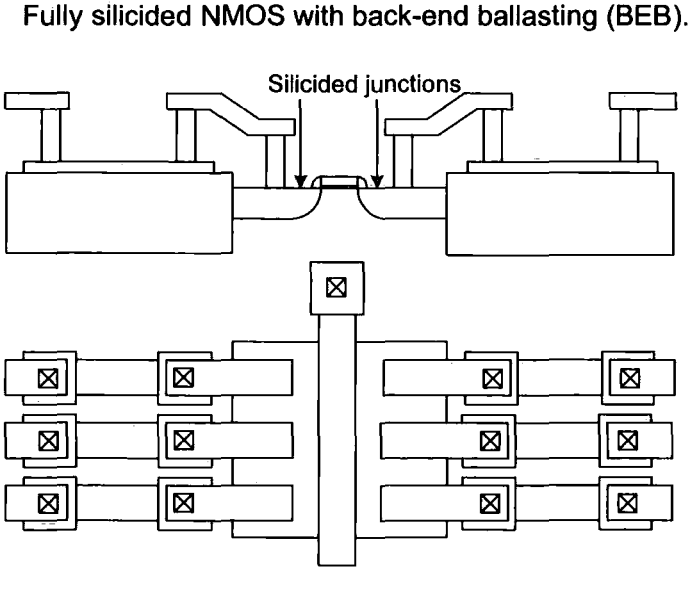

Figure 1.37: A hybrid solution based on the ideas illustrated in Figures 1.35 and 1.36 [3]

1.3.3 Methods for Improving the Failure Current by Modifying the Triggering and Decreasing V_{t1}

Wolf et al. [93] and Polgreen at al. [85] have shown that increasing the gate length increases V_{t1} (as β is dependent on the bipolar transistor base length) and therefore a shorter gate length will improve current spreading. They also added that the long gate length will increase the triggering time, as it will take longer for the base current to diffuse across the base. This means that the ESD pulse will reach a higher voltage before the device is triggered increasing the likelihood of permanent damage to the protected device.

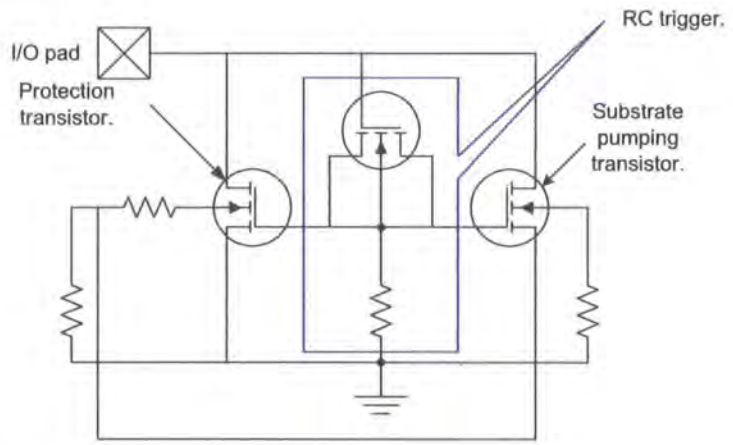

Figure 1.38: RC triggered nMOS switch which pumps the substrate of the protection transistor to assist triggering

Researchers from Texas Instruments [83] have demonstrated an RC triggered substrate pump ggNMOS trigger as shown in Figure 1.38 (also see [85][94]). In this example, the ggNMOS is triggered and kept on by a substrate pump which removes the need for impact ionisation. The RC circuit detects the ESD event and turns an nFET on which, in turn, pumps the substrate of a second nFET. The action of substrate pumping triggers the second nFET into snapback mode by injecting current into the base of the parasitic transistor. As the whole transistor is pumped at the same time, the ESD current flow will tend to be uniform along the whole width. In a standard ggNMOS, the device will inevitably trigger at a single point

and it is up to the designer to force the current to spread out across the width of the transistor. In this pumped circuit, the base of the parasitic transistor is supplied with current across the whole width; therefore I_{DS} will flow uniformly across the width of the transistor.

Koizumi et al. [61] showed a correlation between V_T and V_{t1}. This would suggest that if V_{GS} is increased then the V_{t1} decreases. The pumped device (Figure 1.38) takes advantage of this by increasing V_{GS} to reduce V_{t1} and then supplying the parasitic bipolar base with current to trigger the device.

Researchers at Sarnoff Corp. [59] also stated that for current to spread before the transistor fails, V_{t2} must be greater than V_{t1}; one needs therefore to increase V_{t2} or decrease V_{t1}. The authors took the approach of decreasing V_{t1} and demonstrated this approach using multi finger turn on mechanisms (MFT). Ker [60] showed that if the gate voltage is increased, then V_{t1} will be reduced. The Sarnoff Corp researchers took advantage of this technique to create three new MFT designs.

(i): Use V_{source} (I_{DS} is passed through a resistor) to increase V_{GS} on the next device (Figure 1.39 and [85]). By increasing V_{GS} on the next device, V_{t1} is reduced. As V_{t1} of all the devices is similar, when the first device triggers (the device with the lowest V_{t1}), it will decrease the V_{t1} of the next device. This chain effect will ensure that all the devices will be triggered.

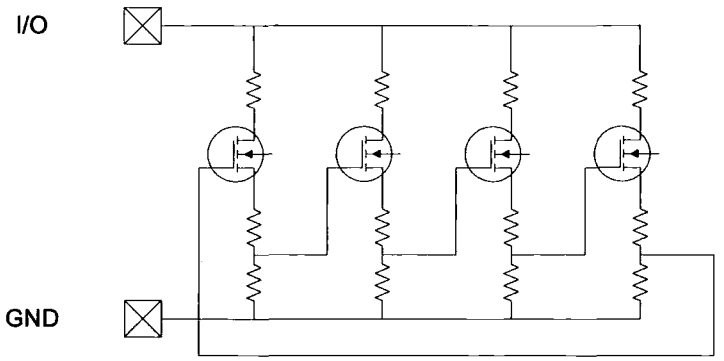

Figure 1.39: The first MFT design where the increased V_{source} is used to increase V_{GS} on the next device and hence propagate triggering

A similar concept is shown to work by Chen et al. [95] where an RC circuit detects the ESD event which in turn increases V_{GS} of the nMOS to lower the V_{t1} (Figure 1.40). Although increasing V_{GS} decreases V_{t1}, it is debatable whether this is still snapback. Once $V_{GS} > V_T$ it should, arguably, be considered as a BigFET protection structure as described in section 1.2.2.4.

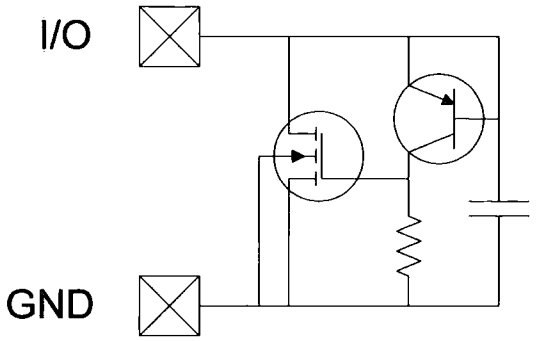

Figure 1.40: The use of an RC based circuit to decrease the triggering voltage of a ggNMOS during an ESD event

(ii): Use V_{source} (I_{DS} is passed through a resistor) to turn on small nMOS which raises V_{GS} of all devices (Figure 1.41). As soon as one device has triggered all the devices will trigger.

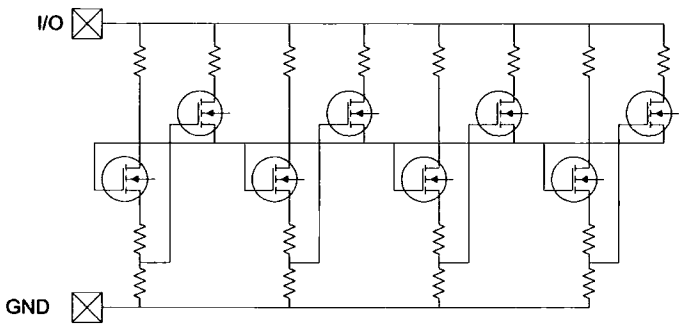

Figure 1.41: The second MFT design. The increase V_{source} in a triggered device is used to reduce the triggering voltage of subsequent devices and therefore propagate triggering.

(iii): Use V_{source} (I_{DS} is passed through a resistor) to raise V_{GS} of all devices via a diode. When one device has triggered, V_{GS} of all devices will be increased which will in turn decrease the V_{t1} of all the devices and all the devices will trigger (Figure 1.42).

Sarnoff Corp [59] also used the substrate pumping principles demonstrated by Texas

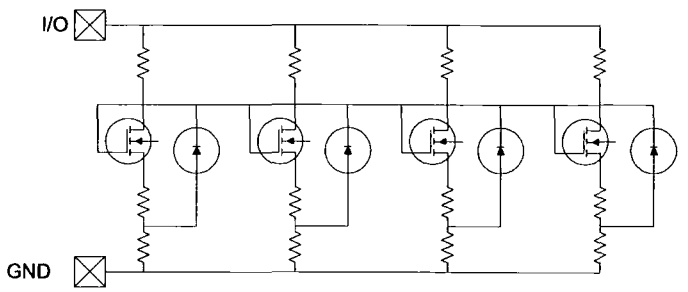

Figure 1.42: The third MFT design. The increase V_{source} in a triggered device is used to reduce the triggering voltage of subsequent devices and therefore propagate triggering.

Instruments [83] and Polgreen at al. [85] to produce a variant of the third design where all the substrate taps are disconnected from the ground and are connected together (Figure 1.43). As one of the devices triggers its substrate voltage will increase to $\sim 0.7V$; it will increase the substrate voltage of all the other devices reducing their V_T and V_{t1}. This will force the other devices to trigger as their V_{t1} will be close to the V_{t1} of the first device to trigger.

Another way of looking at this circuit is that the first device to trigger pumps the substrates of the other devices as demonstrated by Texas Instruments [83].

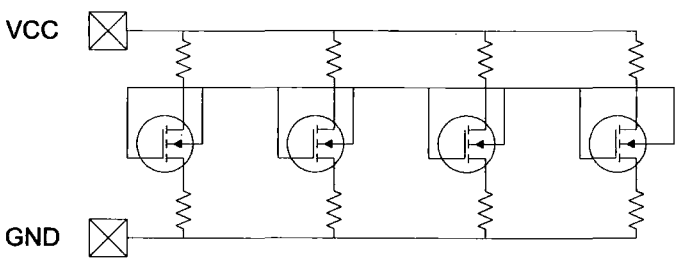

Figure 1.43: The use of substrate pumping principles to propagate triggering between devices

Researchers at IMEC [96] agreed that as you lengthen L_{gate}, I_{t2} decreases due to the increased power dissipation, though they found that if L_{gate} was increased significantly from $0.25\mu m$ to $0.5\mu m$ (in a $0.25\mu m$ process) the I_{t2} value would actually increase. Using a TLP they found that if $L_{gate} = 0.25\mu m$, I_{leak} increased abruptly for an escalating ESD current, whereas if $L_{gate} = 0.5\mu m$, I_{leak} increased gradually. They explained that if $L_{gate} = 0.5\mu m$, any silicon filamentation would not be able

to reach across from the drain to the source in a single ESD event, but after multiple ESD events there would be enough filaments to reach across from the source to the drain. In a $L_{gate} = 0.25\mu m$ transistor the filament is able to reach across the source-drain in a single ESD event.

Chatty et al.[84] showed that special ESD implants in the drain could decrease the triggering voltage of ggNMOS devices in $45nm$ and $65nm$ CMOS technology. The implant also causes the ESD current to flow deeper in to the transistor resulting in a slightly lower R_{on}, lower peak Joule heating and a higher I_{t2} though it does have the effect of increasing drain-substrate capacitance and the leakage current. The higher I_{t2} coupled with the lower V_{t1} allows the reduction in silicide blocked drain extension, thus saving area and reducing capacitance.

1.3.4 Methods for Increasing the Speed of Snapback

Research carried out by Krieger simulated the operation of a ggNMOS and found that V_{DS} can exceed V_{t1} for very fast rise times (100ps-1ns). This is due to the finite time that it takes for snapback to occur which is a function of the transit time of the electrons from the source to the drain [97]. In most cases, this overshoot would be suppressed by the overlap capacitance between the gate and the drain until such time as snapback is triggered. In extreme cases, the overshoot could be significant but very brief ($< 1ns$). This ties in well with research done by Fong et al. [98] where they found that the electric field that could be sustained by an oxide increased as the pulse width decreased and as the maximum electric field was approached shifts in V_T occurred. Thus speed limitations of ggNMOS can cause V_T shifts in a MOSFET though any measurements of a ggNMOS performance using a TLP tester are inconclusive as the rise time of the tester is too slow. TLP testers are covered in more detail in Chapter 2.

1.3.5 The Future of Snapback

Snapback has been shown to work well in $> 0.18\mu m$ process but as we move on to smaller geometries the structure of the transistor changes to address the issues associated with deep sub micron processes. This will have implications for the snapback mechanism.

Boselli et al. [99] looked at the future of snapback in sub $0.13\mu m$ processes and found that ggNMOSs became weaker protection structures due to the decreased substrate resistance which inhibits the snapback operation. The performance can be improved by substrate pumping or other forced triggering methods. It was also found that increased gate length improved the I_{t2} current. Conversely, it was found that the performance of ggPMOS devices improved when implemented in a sub $0.13\mu m$ process.

Thin EPI layers were developed to improve the device performance, latchup immunity, substrate noise isolation and RF performance, though Smedes et al. [39] queried its effect on ESD performance. It was found that the substrate resistivity would have little effect on the trigger voltage though it did affect the performance of the device under snapback. Simulations showed that in high resistivity substrates, the current tended to flow deeper into the device thus mitigating the effect of the increased resistivity. It was also found that high resistivity substrates suffered increased leakage at lower drain currents.

Gupta et al. [89] also found that the increasingly abrupt drain-substrate junctions of current devices are making avalanche multiplication easier causing a reduction in V_{t1} and V_{hold} and thus improving I_{t2}. They also found that pocket implants were having the opposite effect. They decreased β (and increased carrier multiplication, though the change in β dominates) thus increasing V_{t1} and V_{hold} and ultimately degrading I_{t2}. Gupta et al. [89] also found little correlation between I_{t2} and T_{ox} (range 40Å

to 120Å). This is a surprising result, given that T_{ox} is related to V_T and Koizumi et al. [61] showed a correlation between V_T and V_{t1} suggesting a correlation between T_{ox} and I_{t2}.

Another example of contradictory results can be found in Miller's ESD-EOS tutorial [29] on page 6 where it is reported that on shortening L_{gate}, V_{t1} decreases but at the same time I_{t2} degrades. This contradicts other results cited in this literature review.

1.4 ESD Test and Verification

A quick overview of ESD testing is required to help the reader appreciate the performance parameters of a protection structure. The JEDEC ESD test standard [6][100] requires an ESD pulse to be applied between all 2 pin combinations in both polarities. The number of tests required for an n pin IC is given by Equation 1.1.

$$2\sum_{i=1}^{n-1} i \tag{1.1}$$

These tests can be subdivided into the following types:

1. ESD tests across pins that are all on the same power bus as shown by the red lines in Figure 1.44. ESD compliance is not difficult so long as the power buses are sufficiently wide to accommodate the current transients caused by the ESD pulse.

2. ESD tests across the power supply as shown by the green line in Figure 1.44.

3. ESD tests from an IO pin to a power supply pin as shown by the blue lines in Figure 1.44.

4. ESD tests between the two IO pins as shown by the black line in Figure 1.44. ESD compliance is not difficult so long as items 2 and 3 are satisfied. The ESD pulse will travel from the IO pin through the power supplies to the other IO pin.

ESD engineers are primarily interested in items 2 and 3 of the list [29].

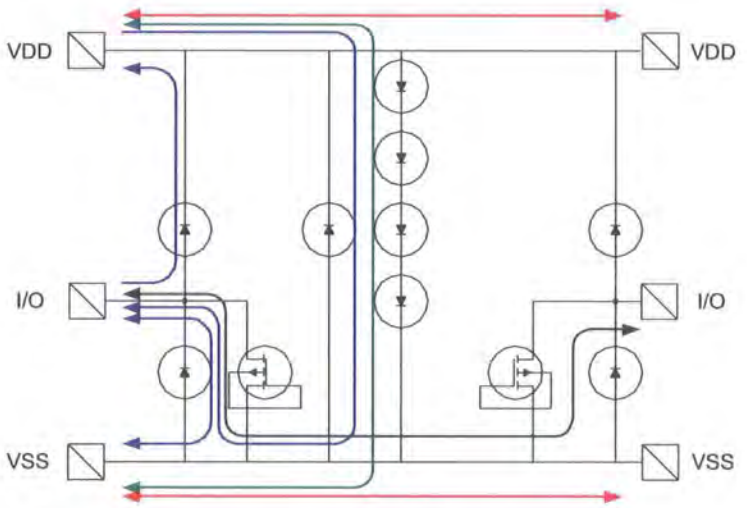

Figure 1.44: Current paths through a chip for all pin discharge combinations

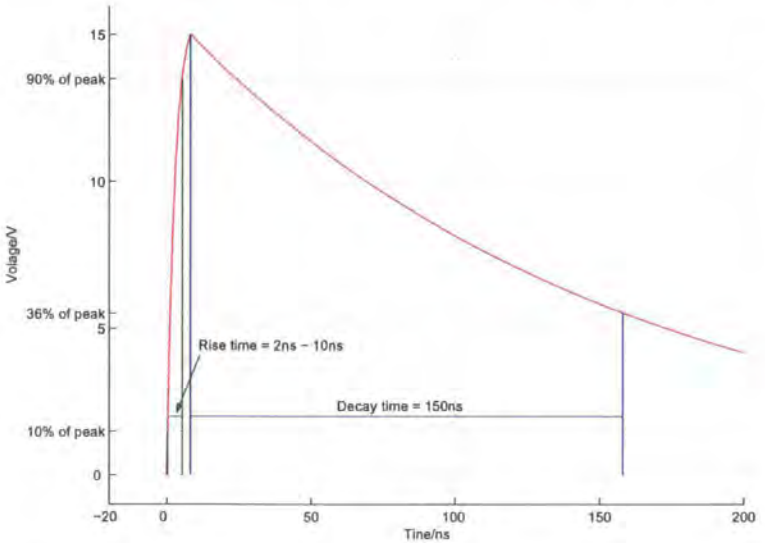

Figure 1.45: Current vs time for an HBM ESD event

The protection structures must be able to conduct a current pulse as described in Figure 1.45 without sustaining irreversible damage. For this graph the current

source is derived from a $100pF$ capacitor being discharged through a 1500Ω resistor (see section 1.4.3 for more details). The peak current is around $1-2A$, the rise time is in the order of 2-$10ns$ [6] and the pulse duration is approximately $150ns$. The protection structure does not necessarily have to conduct instantaneously as gate oxides can withstand high voltages for brief periods of time (Pierce [1]).

Every real life ESD event has a different current/voltage profile given that the electrostatic voltage is a function of environmental factors, the amount of charge to be discharged and the rate of discharge is determined by the parasitic circuit found in the environment. Therefore a common ESD discharge profile, which is a reasonable approximation of real life ESD, needs to be used to compare different ESD protection circuits. Once the standard ESD current has been applied to the test circuit, the circuit needs to be characterised to see if it has sustained any damage. This can be a time consuming process, so again a standardised methodology is required to determine if the circuit has failed to dissipate the ESD energy without sustaining damage. The current through and voltage across an ESD protection circuit can also be used to understand how the circuit is operating during the ESD event. The standard ESD test can be modified to reveal more information on the internal workings of the circuit. The modified test may not accurately simulate an ESD event but it will give more information as to how the protection circuit works under ESD conditions.

1.4.1 ESD Failure Definition

The ultimate aim in ESD research is to make ICs immune to ESD events. There are a number of bodies that provide specifications of standardised tests including the Department of Defense in the USA [101] and JEDEC [6][100] which propose a failure definition for an ESD test. The JEDEC HBM ESD specification states: '*A part will be defined as a failure if, after exposure to ESD pulses, it no longer meets*

the part drawing requirements using parametric and functional testing.'

In practice, testing for compliance with the published specifications every time a part of a chip's ESD performance is being investigated in the laboratory would take far too long. It has been common practice to measure the leakage current of an IO pin (set as an input) at the standard IO voltage +10% and if I_{leak} rises above an arbitrary level, or is significantly altered by the ESD tests, the pin will have been deemed to fail the ESD test. This is a quick and easy way of testing for ESD damage though is not suitable for formal ESD characterization.

Research has been carried out into the link between I_{leak} evolution and degradation of the device when compared to the part specification. If this link is not proven then I_{leak} is not an appropriate measure of damage due to ESD.

Voldman et al. [102] have investigated the correlation between I_{leak} evolution and the degradation in RF performance of a bipolar transistor. They found that f_{MAX} actually improves after low energy ESD pulses and they showed that there was a good correlation between the increase in leakage and the degradation of f_T and f_{MAX}. However when measuring the transistor performance using 'eye diagrams' they found that the eye diagrams did degrade before any increase in I_{leak} became significant. Hyvonen et al. [103] came to the conclusion that the RF performance metrics (Noise Figure, Linearity and S-parameters) degrade before any change in I_{leak} is detected. They also pointed out that in some cases, as their protection structure had an inductor to ground (see section 1.2.2.6), that it was not possible to measure the leakage current.

It has been shown that ESD can damage oxides, causing I_{leak} to increase and that subsequent ESD pulses can repair the damage. The leakage thus falls again to an acceptable level. It is possible that an ESD event could cause a short across an oxide and self-repair in the same ESD pulse though this is very difficult to verify.

Nevertheless the oxide has still suffered undetected damage.

On a different level researchers at Chartered Semiconductor Manufacturing [73] found that ESD pulses could change the crystal structure of copper tracks causing a reduction in the time to failure by electromigration. It is unlikely that a leakage measurement will detect a change in the crystal structure of Cu.

In conclusion, measuring I_{leak} is a common method for damage verification though the relationship between I_{leak} and damage is difficult to pin down.

1.4.2 ESD Testing

There are two types of ESD testing. The first type is a test that purely measures the tolerance of a circuit to ESD events by emulating the current and voltage characteristics of an ESD event and subsequently testing its performance against the circuit specification. The second type of test subjects the circuit to current and voltages that typically occur during ESD events and measures the real time current through and voltage across the circuit to aid understanding of the circuit's mode of operation during ESD events. The first type is used to compare the ESD tolerance of two circuits whereas the second one aims to understand and therefore improve circuit performance during ESD events.

There are a number of industry standard ESD tests. The most commonly mentioned are Human Body Model (HBM) test, Machine Model (MM) testing and Charged Device Model (CDM) testing. These tests are standard ways of measuring a circuits tolerance to ESD events. HBM testing was developed in the 60s and was, as the name suggests, designed to replicate ESD events due to human handling of parts specifically involving discharge from a human through the IC to ground. The MM test is very similar, the only difference is that it replicates a charged machine's ESD through an IC to ground. CDM was developed in the late 70s to replicate the ESD

events where the charge is stored in the IC and it is discharged to ground. An example of this is a chip that is triboelectrically charged (e.g. by sliding down a plastic tube) and then discharged when one of the pins hits a conductive surface.

1.4.3 Human Body Model

Human Body Model (HBM) testing is an industry standard benchmark test for a circuit's ESD tolerance. The HBM test has been defined by a number of industry bodies: JEDEC (US government) [6] US military [101], ESDA [104], Hyatt [105] and EIAJ [106]. Quoting from the JEDEC, HBM standard it is designed to:

'establish a standard procedure for testing and classifying microcircuits according to their susceptibility to damage or degradation by exposure to a defined electrostatic Human Body Model (HBM) discharge (ESD). The objective is to provide reliable, repeatable HBM ESD test results so that accurate classifications can be performed.' [6]

The test emulates the discharge that occurs when an electrostatically charged person touches a pin of an IC whilst another pin on the IC is grounded. A circuit model has been defined by the JEDEC standard and is shown in Figure 1.46. As $R_{HBM} >> R_{DUT}$ ($R_{HBM} = 1500\Omega, R_{DUT} =< 10\Omega$) the circuit provides a current to the Device Under Test(DUT) irrespective of the circuit operation of the DUT, so long as $V_{HBM} >> V_{DUT}$.

The standard states that the circuit must be calibrated by replacing the DUT with a short circuit. The measured t_R must be 2ns-10ns and I_{peak} must be within 10% of $\frac{V_{Capacitor}}{1500\Omega}$. Overall the current decay time is $\sim 150ns$. This calibration is necessary otherwise results differ widely and the effect of parasitics in this circuit are significant (Peirce [1]). To achieve the JEDEC calibration specifications, the parasitics of the HBM tester need to be understood; failing that, it can be difficult to achieve the specification (Chanie [107]). The parasitics are shown in red in Figure 1.47.

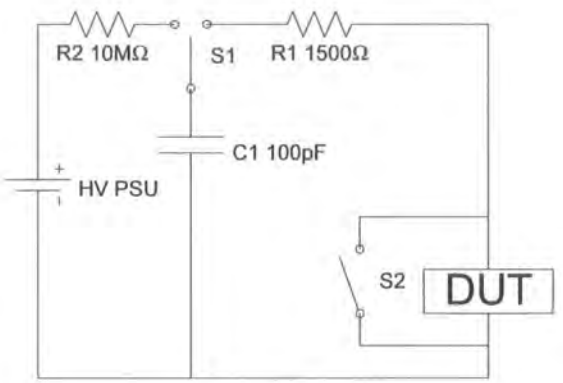

Figure 1.46: HBM ESD test circuit

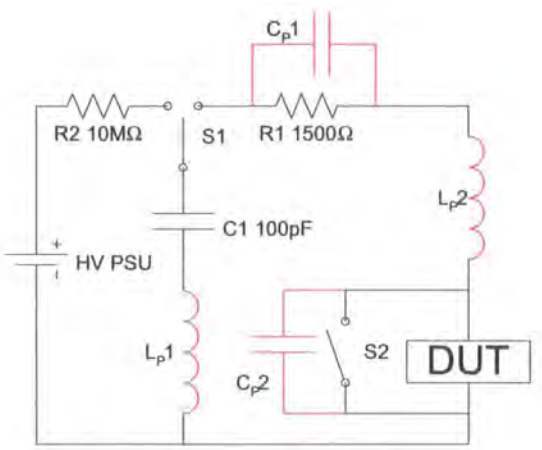

Figure 1.47: HBM ESD test circuit with parasitics

The most important parasitic component is C2 which is in parallel with R1. It reduces the rise time, increases I_{peak} and resonates with parasitic inductors causing ringing. The socket capacitance (C_S) can increase the stressing current as it will charge up until the trigger voltage of the DUT is achieved. After the device is triggered, it will dump all its charge, with a zero resistance, into the DUT. The socket capacitance also increases the rise time of the tester. These parasitic values have been shown to vary from tester to tester even if they were both the same model from the same manufacturer. The JEDEC standard has calibration procedures to control the effects of these parasitics though there is some debate as to whether the HBM pulse represents a real life HBM event. Research has been carried out to verify the validity of the various standards [1][108][109][110][111][112][113][114]. Barth et al. [115] and Hyatt [116] reported measurements made of actual human body discharges and found that the majority were faster than the 2ns-10ns range quoted in the JEDEC specification. Many recorded pulses were hitting the measurement limit of the equipment of $180ps$.

1.4.4 Machine Model

The Machine Model (MM) is more severe version of the HBM test developed by Electronic Industries Association of Japan [106] to replicate the ESD from a machine (e.g a wire bonder) through an IC to ground. This test has also been defined by JEDEC [100] and other bodies. The MM tester circuit is shown in Figure 1.48.

1.4.5 Charged Device Model

Charged Device Model (CDM) was originally described by Bossard et al. [117]. It is an ESD event where charge stored in the IC is rapidly removed through one pin. An example of this is a IC being charged whilst sliding down a plastic tube then falling

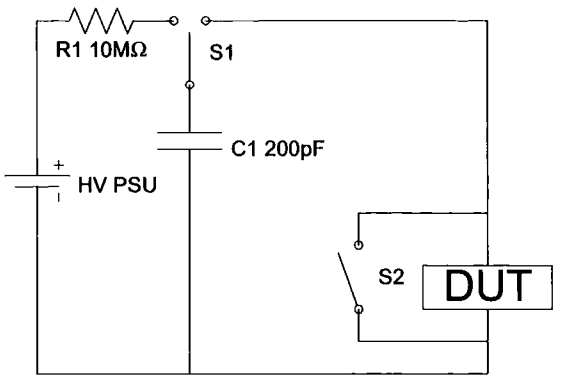

Figure 1.48: MM ESD test circuit

out on to a conductive surface. When one of the pins come into contact with the surface the chip discharges. The nature of the ESD event means that the discharge path inside the IC is hard to find. In an HBM ESD event the entry and exit points of the current are known and therefore it is often not too hard to deduce the current path based on the knowledge of the circuit and failure analysis. With CDM we know the exit point but, as the whole chip is charged up, all components in the chip supply part of the current. As the chip will be off during the ESD event, parasitic elements will play a significant part in conducting the current to the grounded pin. All this makes analysis and solving the problem particularly difficult.

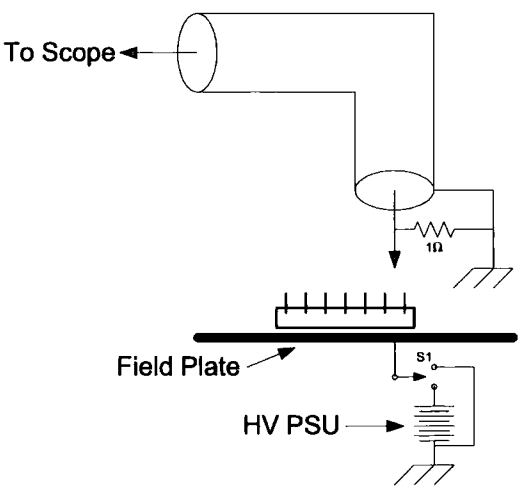

Figure 1.49: CDM ESD test circuit

Reproducing the CDM ESD test has also proved difficult. A common method is shown in Figure 1.49 [12][107][118]. The test is initialised by grounding the field

plate and pins of the DUT to ensure that there is no charge stored in the device. The device is then charged by disconnecting the scope probe and connecting the field plate to the high voltage PSU and the chip is charged by electrostatic induction. The test is then carried out by discharging the chip though one pin using the ground probe. The parasitics of this test are shown in Figure 1.50.

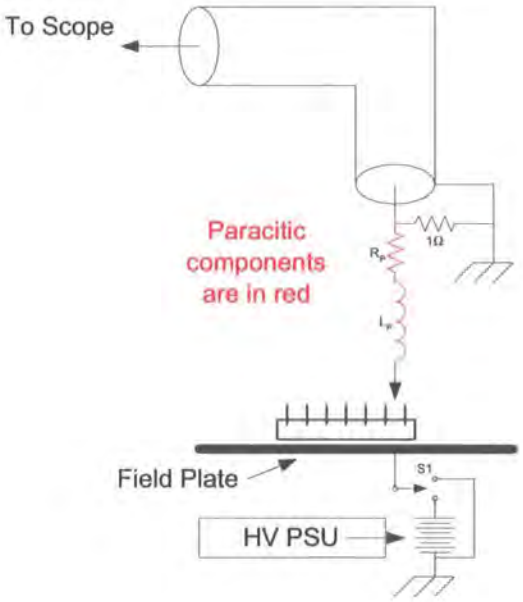

Figure 1.50: CDM ESD test circuit with parasitics

L_P and R_D need to be as small as possible to minimize the rise time and maximize the peak current. Typically the peak current is in the order of 10A and the rise time 100ps to 1ns [12][107]; this also means that the probes and oscilloscopes have to have a bandwidth of around 2-5GHz. Further details on achieving the correct waveform for CDM testing have been discussed by Henry et al. [118].

1.4.6 Transmission Line Pulse (TLP) Tester

Unlike the HBM, MM and CDM tests, the TLP test is not designed to replicate ESD events for the purpose of determining tolerance to ESD events. It is a test that is designed to provide data on how the circuit operates during the high current conditions found in ESD events ($\sim 1.3A$). This could be done using a curve tracer

except that the chip would fail at relatively modest currents. The TLP overcomes this problem by using a transmission line to apply a short ($\sim 100ns$) pulse to the device under test so that high current performance of the part can be measured without damaging the device. The TLP's short pulses allow high power and high current pulses to be applied to the device under test (DUT), whilst at the same time limiting the total energy supplied to the DUT and therefore preventing device from being damaged. If high currents are applied using a curve tracer, the DUT will be damaged, even at modest currents, since the pulses are at least a few ms long whereas the TLP pulses are around $100ns$ long. The TLP tester is used extensively in this thesis, Chapter 2 describes the construction and use of a TLP tester.

1.5 Conclusion of the literature review

1.5.1 Fidelity of the HBM test method

The research community use the HBM standard to replicate real world ESD events in the laboratory in order to provide a common test by which results can be replicated in different laboratories. The acceptance of this standard has become entrenched and there appears to be little questioning of the validity of this model developed over 30 years ago. The accuracy and speed of generic electronic test equipment such as oscilloscopes has increased considerably in the last 30 years, yet only two papers have been identified that actually measure an electro static discharge from a human [115][116]. They found that the rise time of a real electrostatic discharge from a human was significantly faster than the $2ns - 10ns$ range mentioned in the JEDEC ESD specification [6]. Barth at al. [119] also studied the initial rise time (i.e. the initial $\frac{dV}{dt}$) as this is the operating region where the triggering voltage is reached. They found that this portion of the rise time was considerably faster the expected figure implied by the $2ns - 10ns$ rise time stated in the specification.

It could be that the approximations in the JEDEC specification were appropriate in the 1960's and 1970's, but the rapid pace of evolution of silicon circuits and components means that these assumptions should be revisited. Clock speeds and analogue frequencies on silicon chips have increased by many orders of magnitude, meaning that the parasitic capacitance in a typical high speed IO has reduced. The parasitic capacitance of an IO may be able to attenuate fast rising edge signals and therefore give the circuit more time to react to the ESD event. The decreasing size of transistors may also result in an increase in the maximum temperature for a given ESD event as the Joule heating would be more concentrated.

The knowledge of which aspects of a real life ESD event is important is critical to ensuring good correlation between real life ESD events and ESD test equipment. It both ensures that we understand how to correlate ESD test results from different vendors of ESD test equipment and different types of ESD test equipment, such as HBM, MM and TLP testers. Although the basic structure of MOSFET components has remained unchanged in the last 30 years there are significant differences. The gate length and gate oxide thickness has decreased significantly, interconnect metallisation has changed from aluminium to copper, junction depth has decreased and the doping densities have increased resulting in thinner depletion regions. This means that for a given drive strength, the parasitic capacitance of the MOSFET has decreased reducing the ability of the MOSFET to bypass the initial portion of the ESD current before the transistor triggers. Resulting in a faster rise time to high voltages. The decreased oxide thickness has reduced the transistor's tolerance to over voltage. This applies to both the absolute voltage as well as the time integral of the voltage since there is less volume to dissipate the ESD energy before critical temperatures are reached. The decreased junction depth and the thinner depletion regions may reduce the volume that conducts the ESD charge and therefore increase the maximum temperature reached during an ESD event. Conversely the distances that the charge has to travel to initiate conduction has decreased significantly; this

may decrease the switching time and therefore increase a transistor's tolerance to fast rising edges.

In summary the rapid evolution of CMOS design within the semiconductor industry means that the specification of ESD test waveforms needs to be reviewed to ensure that they are still relevant.

1.5.2 Alternative ESD investigation techniques

Significant progress has been made in the last 20 years on ESD test techniques that allow further understanding of the operation of ESD clamps under ESD conditions. This was started by Maloney's with the TLP tester that was first published in 1986 [120] which showed that the IV characteristics of circuits under HBM and MM conditions. Later Geiser's VF-TLP [121] attempted to replicate and measure the IV characteristics of components under CDM conditions. Other creative examples include triggering external measurement equipment such as oscilloscopes or photo multipliers with the TLP pulse in order to allow other parameters to be measured during a controlled ESD event. There is certainly scope to further the collaboration between different types of measurement equipment, for example using a TLP triggered laser to induce non uniform triggering, and then ascertain, through photo emission microscopy, whether the current does flow uniformly through the transistor. This could be used to determine whether a transistor needs to be triggered in to a uniformly conducting state or whether it can be non-uniformly triggered and later recover from the non-uniform current flow. This would help designers understand how the circuits and components operate under ESD conditions and how each component of the mechanism works.

Although taking direct measurements of circuits in silicon under ESD conditions present significant challenges, it has a considerable advantage over simulations. Fi-

nite element semiconductor simulations can suffer from non convergence, inaccuracy, long simulations times and high demand on computing resources. SPICE simulations may not be suited to ESD conditions as the current densities are far higher than those found in conventional circuits and the models do not take into account the effects that occur outside the MOSFET's normal operating range.

1.5.3 Current state of ESD protection knowledge

In spite of significant restrictions (minimum/maximum dimensions, fixed doping profiles and oxide thicknesses, minimum/maximum poly/active/metal densities etc...) engineers, when designing in the CMOS process, have still been able to create a wide range of ESD clamps each with their own advantages and disadvantages.

The amount of literature available for the ggNMOS ESD clamp is limited, especially in recent years. This is surprising given the constant evolution in process technology which is likely to have an impact on ggNMOS operation. There are plenty of papers that cover how the transistor behaves when the ggNMOS is conducting an ESD event but there is a marked absence in detailed explanations as to the mechanisms involved. The literature limits itself to a simplistic explanation which in short states that the drain-substrate diode breaks down, this increases the substrate voltage and therefore turns on the parasitic bipolar transistor. We do not know why the hold voltage is significantly less than the triggering voltage. How does the evolving CMOS technology miniaturisation affect snapback performance? Can other clamping techniques, such as the BigFET, be combined with the ggNMOS to produce a clamp that triggers uniformly, but also allows the current to flow deeper in to the silicon (ggNMOS) thus reducing the peak temperature? BigFETs have probably been popular because they are an inherent low risk solution where it should be possible to get the design right the first time. This is because the width scaling issues are minimal, the technique scales well with technology (gain of the

transistors is constantly increasing) and performance can be predicted using SPICE simulations. As the transistor gain is constantly monitored by the foundry the clamp performance should remain relatively consistent for all process corners. Conversely, simulating ggNMOS clamps involves using finite element simulators and achieving good correlation with silicon results has repeatedly proved difficult. ggNMOS's also have the added complication that width scaling cannot be taken for granted therefore this technique probably requires more than one silicon test chip and is generally a higher risk solution. It is far more difficult to predict how the variations in process will affect ggNMOS performance as it is possible that manufacturing parameters are not controlled tightly enough to ensure a reliable operation. A better understanding of ggNMOS operation would help predict how process variations would affect performance under ESD conditions.

Chapter 2

Design and Build of the Transmission Line Pulse Tester

2.1 Introduction

This Chapter introduces the Transmission Line Pulse (TLP) tester; it is used as a diagnostic tool to help determine how circuits operate under ESD conditions. This is the fundamental difference between the TLP tester and all other types of standard ESD related tests which merely tell you whether a circuit is ESD tolerant or not, but fail to reveal any information as to why a circuit is ESD strong or weak. This Chapter covers the design and construction and explores advantages/disadvantages of various different configurations. It provides guidance on TLP operation and test data interpretation and shows how the latter can be correlated to other standard ESD tests. Other uses to TLP technology such as using EMMI at the same time as TLP to determine the location of the current flow are explored.

2.2 Theoretical Background

The TLP tester was developed by Maloney et al. [120] to meet a shortcoming of the HBM test. The HBM test (section 1.4.3) is designed to replicate real life ESD events as accurately as possible and provide a common standard so that different protection circuits can be compared. Initially, HBM tests are normally carried out at low energy levels and the test energy is increased until the device no longer meets the published chip specification, at which point the chip is deemed to have failed the test. The ESD rating of the chip is based on the highest HBM test voltage that the chip can withstand and still meet the published specifications. Therefore the HBM test tells you how immune a circuit is to ESD events but it does not provide any information on the internal operation of the circuit under ESD conditions.

The fundamental difference between the TLP test and the HBM test is that the TLP applies a constant voltage via a constant impedance for a short period of time to the Device Under Test (DUT) (Figure 2.1), whereas the HBM test is an RC circuit discharge (Figure 2.2) resulting in a variable voltage across the device. Figure 2.3 shows the waveforms for both test methods. The TLP tester's constant voltage/impedance pulse allows the current and voltage to be determined for any given DC operating point. The pulse length is kept short ($\sim 100ns$) so that the high current/high voltage IV characteristics can be measured without causing permanent damage to the IC; the pulse however has to be sufficiently long so that the parasitic capacitances do not affect measurement and any noise can be eliminated by averaging. The TLP test voltage is systematically increased. A high current DC IV curve can therefore be constructed ultimately allowing the high current mode of operation of the circuit to be understood.

This measurement cannot be performed using a curve tracer as the DUT would sustain damage before the high currents of a TLP test could be achieved. The

length of the TLP pulses can range from $3ns$ to $1\mu s$ though research has shown that, when comparing the peak current at which the DUT fails, a $100ns$ pulse can be equivalent to an HBM pulse [122][123][124][125][126][127][128].

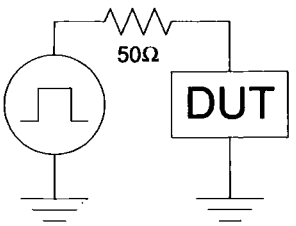

Figure 2.1: The basic concept of a TLP tester

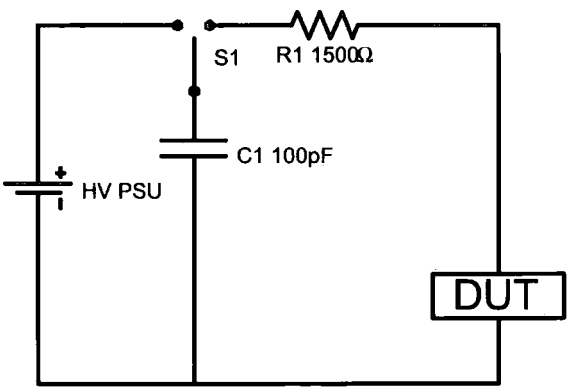

Figure 2.2: HBM ESD tester schematic

2.3 The Basic TLP Circuit

The basic set up of a TLP tester is shown in Figure 2.4. The Transmission Line (TL) is charged via the $10M\Omega$ resistor and SW1. It is discharged through the DUT to ground using SW2. The length of the pulse is dependent on the length of the TL and the speed of an electrical pulse travelling through it. The pulse length is around 10ns for every meter of TL. An oscilloscope is used to measure the current through and/or voltage across the DUT during the transmission line discharge.

In practice the impedance mismatch between the DUT and the TL in this circuit (Figure 2.4) will cause reflections to be propagated back along the TL. The open

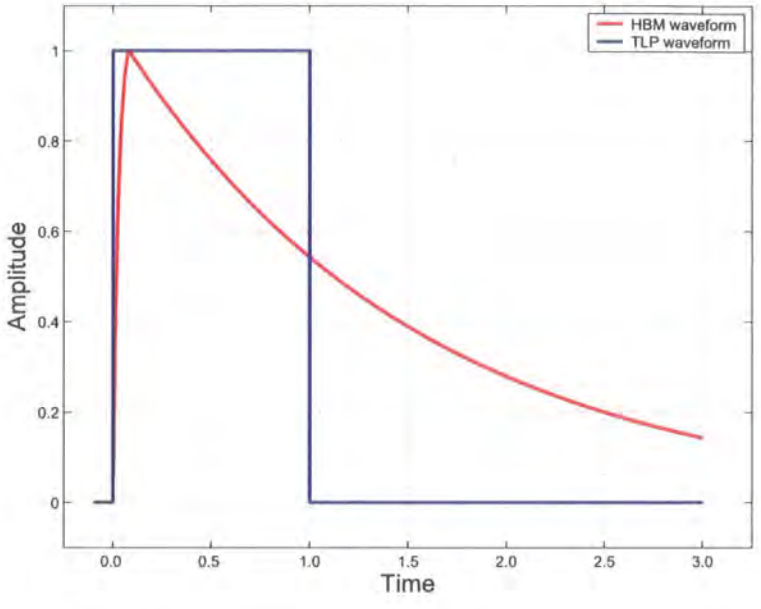

Figure 2.3: Normalised TLP and HBM waveforms

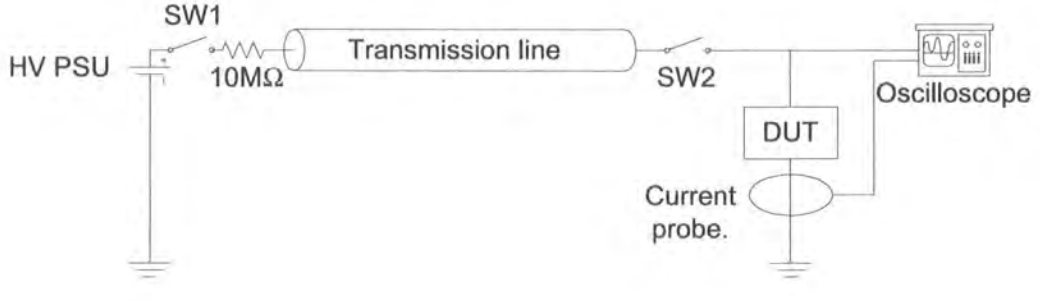

Figure 2.4: The basic set up of a TLP tester

circuit at the other end of the TL will invert and reflect the signal back to the DUT. This means that the DUT will be subjected to many pulses of unpredictable amplitude and polarity. Modifications are needed to ensure that the DUT is subjected to a single pulse of known current and duration. A number of solutions have been proposed in the past [129].

2.3.1 Constant Impedance TLP

In this circuit, shown in Figure 2.5, D1 and R1 have been added so as to present a 50Ω load to any negative polarity pulses that are reflected from the DUT thus preventing these pulses from being reflected back to the DUT. This solution requires that any reflections from the DUT are negative in polarity. This is almost always the case given that $R_{DUT} << 50\Omega$. (For an ESD clamp to be effective the voltage across it, depending on the process node, must be less than $\sim 12V$ for currents measuring up to $1.3A$. As a consequence its loading is almost always less than 50Ω.)

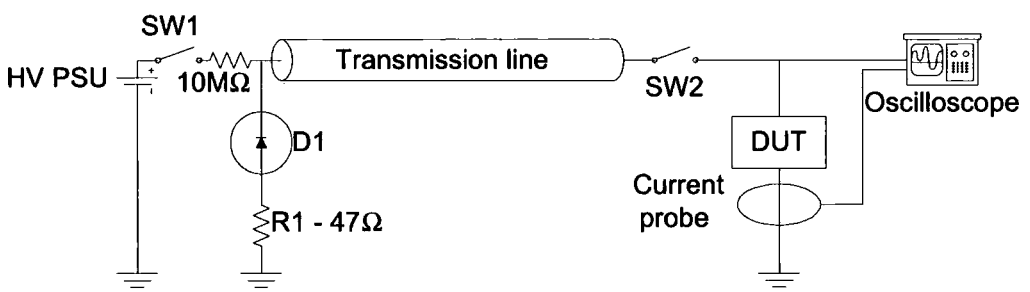

Figure 2.5: Constant impedance TLP tester

2.3.2 Constant Impedance TLP with 20dB Attenuator

An attenuator can be added to attenuate the reflections as shown in Figure 2.6. The attenuator will also reduce the TLP amplitude by 20dB so the TL has to be charged to 10 times the required test voltage. Any reflection generated by the mismatch

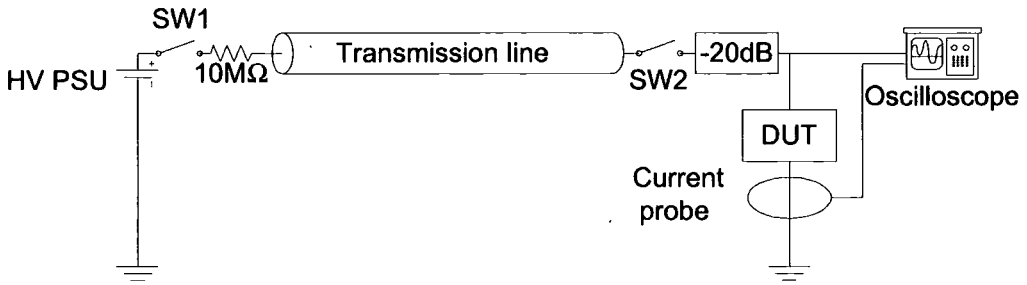

Figure 2.6: Constant impedance TLP tester with a 20db attenuator

in impedance will have to pass through the attenuator where its amplitude will be reduced by 20dB.

The open circuit at the end of the TL will reflect the attenuated reflection back towards the DUT but before it reaches the DUT, it will be attenuated by a further 20dB. At this point the pulse amplitude should be insignificant, if this is not the case then a higher value attenuator can be used.

The advantage of this circuit is that it virtually eliminates all reflections both positive and negative; but the disadvantage is that the TLP voltage needs to be increased by an order of magnitude to compensate for the attenuator.

2.3.3 Constant Current TLP

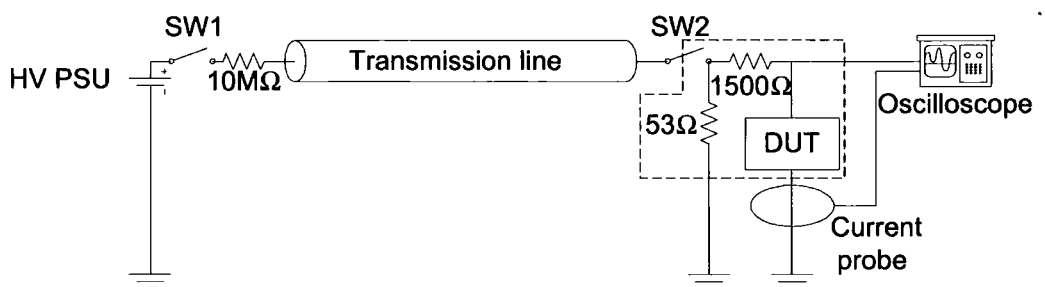

Figure 2.7: Constant current TLP tester

In this solution the TLP is set up such that the circuit in the red box has an impedance of almost exactly 50Ω (Figure 2.7). This means that there will not be any reflections. The majority of the power of the pulse is dissipated in the 53Ω resistor and the 1500Ω resistor provides a constant current for the DUT. The impedance of the DUT has little effect on the reflections generated in the circuit. Given that this solution has the same impedance and voltage as an HBM test, the rise time needs to be the same as an HBM test. This solution also has similar advantages and disadvantages as the *Constant impedance TLP with 20dB attenuator* solution.

2.3.4 Multilevel and Multi-impedance TLP testers

This type of TLP tester has recently been proposed [130][131] where two 50Ω cables are joined by a resistor to form a TL with variable impedance (as seen by the DUT). This means that when the switch (SW2) is closed the TLP tester operates as a constant impedance TLP tester until the pulse reaches the resistor that joins the two TLs. At this junction the amplitude of the pulse is reduced which allows the user to measure the performance of the DUT, once triggered, at low currents whilst still supplying enough initial voltage to ensure that the device is triggered.

2.4 TLP Tester Result Analysis

These quasi DC tests can be used to build up an IV trace for the DUT. An example of an IV trace, using a 50Ω TLP, of a ggNMOS is shown in Figure 2.8. This ggNMOS was TLP tested in 1V increments from 5V. Between 5V and 11V the DUT does not draw any current (black line). When the TLP tester is charged to 12V the DUT draws about 90mA and clamps the TLP voltage to around 7.5V. The current can be calculated as the impedance and the initial voltage of the TL are known

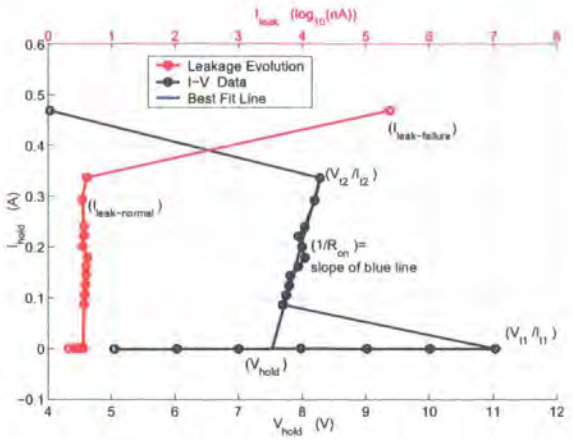

Figure 2.8: Example IV trace of a ggNMOS using a TLP tester

($\frac{12V-7.5V}{50\Omega} = 90mA$). As the initial voltage of the TL is increased, V_{DUT} and I_{DUT} increases. Stadler et al [126] have shown that TLP testing can provide the triggering voltage (V_{t1}), triggering current (I_{t1}), holding voltage (V_{hold}), on resistance (R_{on}), secondary breakdown voltage and current (V_{t2} and I_{t2}) as illustrated in Figure 2.8. V_{t2} and I_{t2} are the maximum currents before secondary breakdown occurs. This is when thermal generation of carriers exceeds the carriers produced by doping which results in high current densities and ultimately failure. This can be detected by observing that there is a sharp change in the IV trace as shown in Figure 2.8. In the graph the DUT undergoes failure when the current in the IV curve trace increases from $0.35A$ to $0.45A$ and the voltage decreases from $8.25V$ to $4V$. Barth et al. [132] argue that this kind of failure can be detected far more reliably by measuring the leakage current (I_{leak}) between the TLP tests, as shown by the red trace in Figure 2.8 where I_{leak} rises by 5 orders of magnitude when V_{t2} and I_{t2} are reached. They also argued that smaller damage that could not be detected in a TLP IV curve would show up in leakage measurements.

A similar IV curve can be created using a constant current TLP (Figure 2.7). This type of TLP test can give more information on the snapback operation at low currents which the 50Ω TLP is unable to do. Figure 2.9 illustrates the IV curves for 50Ω TLP and a 1500Ω TLP for the same DUT. The graph shows that up to $11V$

the DUT does not draw any current. The next TLP test undertaken was at $12V$; the DUT triggered and reduced the voltage to $\sim 7.5V$. To do this the DUT has to draw $90mA$ ($\frac{12V-7.5V}{50\Omega} = 90mA$) whereas with the 1500Ω TLP the DUT only has to draw 3mA ($\frac{12V-7.5V}{1500\Omega} = 3mA$). Therefore a constant current TLP is better able to measure low current performance. The TLP tester can also show how much current is required to keep the DUT in a triggered state (Figure 2.9).

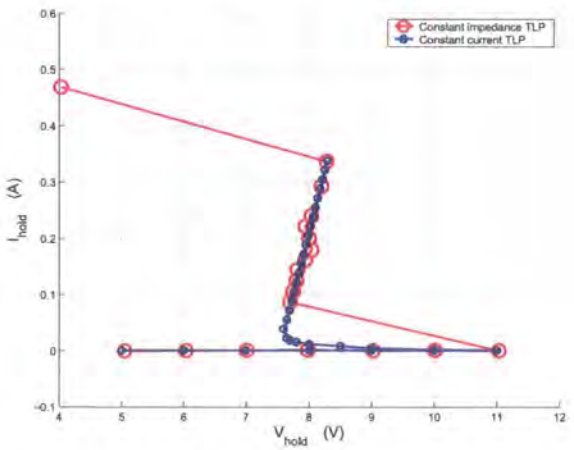

Figure 2.9: Comparison of an IV trace of a ggNMOS using a 50Ω TLP tester and a constant current TLP tester

2.5 Correlation

Ideally it would be useful to be able to correlate TLP measurements to HBM measurements so that the IV curves that are created are matched to the performance during an HBM event. The aim is to know for any given TLP result what the equivalent HBM pulse would be. Ideally we would have the same peak current at failure for both HBM and TLP tests (I_{t2}). A number of papers have been published on this topic addressing various factors in the design of the TLP. For a start it must be noted that there are some papers that question the ability of the HBM test to replicate real life ESD events as well as whether the HBM specification [6] is sufficiently tight to provide a consistent standard for comparison [110][116][123][119][133]. Moreover opinions are mixed as to whether TLP and HBM tests do correlate.

The length of the TLP pulse is important if correlation in I_{t2} is to be achieved between TLP and HBM tests. Lee et al. [134] carried out some 2D finite element simulations of TLP and HBM tests and calculated that if a TLP pulse was $95ns$ long, it would correlate well to an HBM pulse on the basis that, given the same peak current, the two pulses dissipated the same amount of energy. They added that a purely analytical calculation would suggest that a pulse length of $75ns$ would be more accurate. If the comparison, using finite element simulation, were done on the basis that the two pulses produced the same peak temperature, the TLP pulse would have to be $25ns - 30ns$ long. The authors of the papers also found good correlation between real HBM and TLP tests for TLP pulse lengths of $80ns - 85ns$.

In practice a pulse length of about $100ns$ has been shown on many occasions to correlate [122][123][124][125][126][127][128] and not to correlate [123][126][128][135]. The reader may note that some papers [123][126][128] provide evidence for correlation and miscorrelation in the same report. The reasons cited in these two cases were that the HBM specification is relatively loose and this alone can affect the ESD tolerance of a device and also that correlation and miscorrelation were process dependent.

Investigations were carried out by Barth et al. [123] where good correlation was demonstrated for TLP 50Ω systems with a rise time of $10ns$ though, if the rise time decreased to $0.2ns$, the correlation was very poor. Another paper by Barth et al. [119] adds that it is the initial $\frac{dV}{dT}$, of the ESD pulse, that is important. The time taken for an HBM pulse to reach the trigger voltage of a ggNMOS ($\sim 15V$) was found to be between $0.2ns$ and $1ns$ and it is suggested that the TLP system should have the same initial rise time if the two are to correlate. This means that the rise time of the TLP system has to be much faster than the rise time of an HBM tester. This is because, for a given test current, the HBM tester will reach much higher voltages than the $50\Omega - TLP$ tester. Musshoff et al. [128] also investigated the effect that

the rise time of the TLP pulse would have on the performance of a ggNMOS. They showed that a $2ns$ rise time TLP pulses tended to produce consistent results with reasonably good correlation to two standard HBM testers though the measured current profile of the TLP tester was significantly faster than that measured in the HBM tester. In an attempt to improve correlation they added two different types of passive filter (first and second order) to improve the correlation between the current profiles of the TLP and HBM testers. The results were mixed. For single finger devices, the ESD performance of the device was largely unchanged but for 4 and 16 finger devices the mean performance deteriorated significantly and the variation from chip to chip increased significantly thus making the correlation worse. An interesting additional conclusion can be drawn from this data that a slow rise time TLP pulse was unable to trigger all the ggNMOS fingers in a multi finger device whereas this was not a problem in a single finger device. This would suggest that designers should aim to make all ggNMOS ESD protection single finger in construction to ensure efficient clamp designs.

2.6 Other Designs of TLP Testers

(i) Substrate current measurement

Maloney et al. [120] also explored variations on the basic TLP design. The design shown in Figure 2.10 can be used to measure the substrate current separately from the source current. This can aid understanding of the circuit operation.

(ii)TLP and EMMI

Zwol et al. [136], Gieser et al. [91] and Pogany et al. [137] demonstrated that the TLP tester can be integrated with photo emission microscopy to assist understanding of the circuit operation during an ESD event. It is well known that the recombination

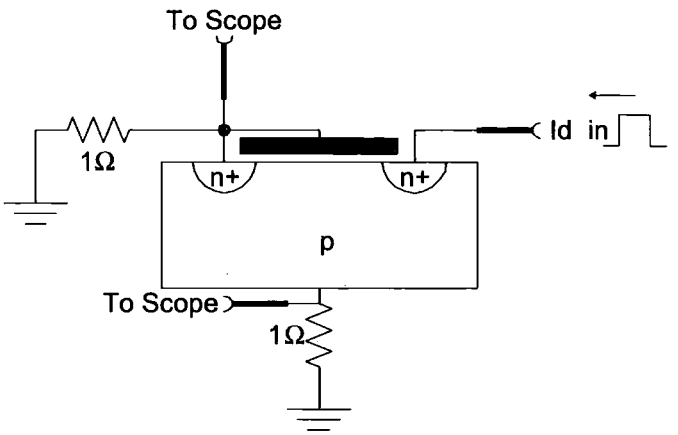

Figure 2.10: The use of a TLP tester to measure the substrate currents during an ESD event

of carriers produces photons but, as silicon is an indirect bandgap material, the number of photons emitted is small. Therefore a sensitive microscope is needed to detect these photons. If the capture period of the microscope is synchronised with the TLP triggering, it should be possible to locate areas of high current density and understand how the circuit is operating. Similar experiments were carried out by Litzenberger et al.[26][27] using laser interferometry.

(iii) Very-Fast Transmission Line Pulse (VF-TLP)testing

H.Gieser et al. [121] has proposed and demonstrated the use of a VF-TLP aimed at analyzing and solving CDM problems. The VF-TLP is of a similar construction to a normal TLP but the rise time is around $200ps$ and the pulse duration is 3-10ns.

2.7 TLP Design and Construction

For our research a constant impedance TLP design was chosen as the construction was relatively simple and a $200V$ power supply would be more than sufficient. The other TLP testers required power supplies capable of generating more than $1000V$.

2.7.1 The Circuit Operation

The basic construction of a TLP is shown in Figure 2.11. The Transmission Line (TL) is charged via R2 to the test voltage and then it is discharged, using SW1, through the DUT to ground. The length of the TL and the speed of the electrical pulse through the TL determines the length of the electrical pulse. In this design the TL was approximately $10m$ long so as to produce a $100ns$ pulse. During the test an oscilloscope is connected across the DUT, via SW3, so that the voltage across the device during the test can be measured. Given that the TL has a 50Ω impedance, the current flowing through the DUT can also be calculated.

The mismatch in impedance between the 50Ω coaxial cable and the DUT causes reflections to propagate back up the TL and unless these are suppressed they will keep reflecting back and forwards in the TL. In almost all cases the resistance of the DUT is less than 50Ω and therefore the reflected pulse will have a negative polarity. The purpose of $D1$ and $R1$ is to absorb the negative polarity pulses preventing them from being reflected back to the DUT.

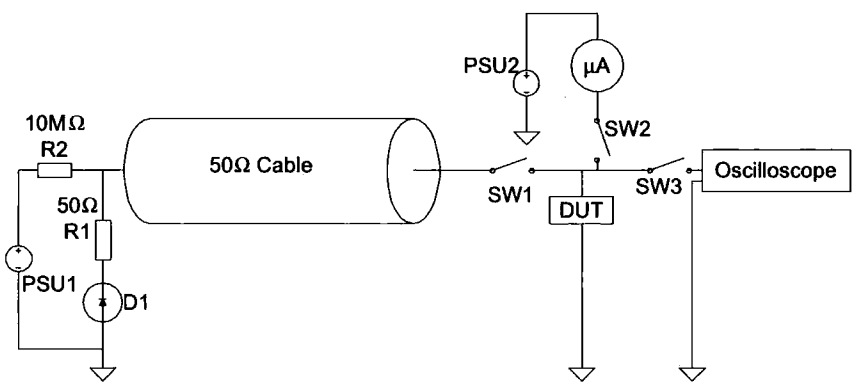

Figure 2.11: TLP tester schematic

Once the pulse has been applied, the DUT is biased by PSU2 and the leakage current, via SW2, is measured using an ammeter (Figure 2.11). The leakage current is measured before and after each TLP test so that any variation in leakage current can be detected. Often the leakage current of ESD devices will increase sharply when

the device is damaged. However it must be borne in mind that the strict definition of *device failure* is when the device does not meet the published specifications. The leakage test is used as it is quicker than a full test to see if the device meets the published specifications.

2.7.2 The Physical Circuit Design

The TLP tester is essentially an RF circuit; consideration has to be given to the physical design of the PCB.

The TLP pulse amplitude, duration and rise time need to be defined so that they stress the DUT in a similar way as an HBM test. Assuming that the peak TLP current is to be the same as the peak HBM current and, given that the TLP impedance is 30 times less than the HBM impedance (50Ω vs 1500Ω), the TLP pulse peak voltage needs to be 30 times less than the peak HBM voltage. It has been demonstrated that there is good correlation between $100ns$ TLP tests and standard HBM pulses in terms of the current required to cause a device failure. It has also been shown mathematically that the energy in an HBM pulse is equivalent to a $\sim 100ns$ TLP pulse assuming the same peak current. The length of the pulse is determined by the length of the transmission line and, given that the speed of a pulse in a TL is around $5ns/m$, the cable needs to be about $10m$ long (the pulse has to travel from the DUT to the other end of the TL and back again).

It is well known that the rise time of the TLP pulse affects the triggering voltage and performance of a ggNMOS. Barth et al. [119] argued that for a TLP to be equivalent to an HBM pulse the initial dV/dt should be the same and one should not concentrate on the rise time. The reasoning was that, as any switching in a ggNMOS happens in the first 20V of an ESD signal and that part of the triggering is capacitive, both HBM and TLP signals should reach 20V in the same amount of

	Target TLP specification	HBM equivalent
Rise time	$66pS - 333ps$	$2ns - 10ns$
Peak voltage	$0V - 66V$	$0V - 2000V$
Pulse length	$100ns$	RC=$150ns$

Table 2.1: Comparison of target TLP specification with JEDEC [6] HBM specification

time. Given that the JEDEC HBM specification requires a rise time of $2ns - 10ns$ and that, as previously explained, the HBM voltage is 30 times greater than the TLP voltage, the TLP rise time should be $66pS - 333ps$ to ensure dV/dt equivalence.

Some aspects of the ideal TLP specification detailed in Table 2.1 are easy to achieve in practice. The pulse length is determined by the length of the transmission line and this part can be calibrated by using a TL that is too long and shortening it until the pulse length is $100ns$. The peak TLP voltage is controlled by PSU1 (Figure 2.11). A power supply with a tolerance of $\pm 1V$ is more than adequate.

The difficult part is getting a $66ps$ rising edge that is fast and smooth. To achieve this the design has to be such that there is a consistent 50Ω path from the cable, through SW1, past the leakage current measurement and the oscilloscope to the DUT. Any discontinuities in impedance will cause reflections and make the rising edge slower and noisy. To ensure that the switch was a constant 50Ω impedance a reed switch was soldered across a break in a 50Ω PCB transmission line (Figure 2.12). A reed relay was also used to connect the 50Ω track to the leakage current measurement as it is a low capacitance and low leakage structure when in the open circuit state. It was also found that the reed relay was the type of switch with the least occurrence of contact bounce ($> 95\%$ of switch events were bounce free). The only significant variation in impedance is due to the oscilloscope probe which is placed as close to the DUT as possible to minimize the reflections between the probe and the DUT. It was found that the fastest rise times occurred when the track length between the switch and the DUT was as short as possible. This is due to the attenuation of

	Actual TLP specification	HBM Equivalent	Target TLP specification	HBM equivalent
Min Rise time	$< 1ns$	$< 30ns$	$66pS - 333ps$	$2ns - 10ns$
Peak voltage	$160V$	$4800V$	$> 66V$	$2000V$
Pulse length	$100ns$	RC=$150\mu s$	$100ns$	RC=$150\mu s$

Table 2.2: Comparison of target specification and achieved TLP specification

high frequency signals by the transmission line. The easiest way to create a slow rise time TLP signal was to add a few meters of 50Ω cable between the switch and the DUT. A $10m$ length of coaxial cable, between the switch and the DUT, would cause the rise time to increase from $1ns$ to $30ns$ because of the attenuation of the high frequency components of the TLP pulse.

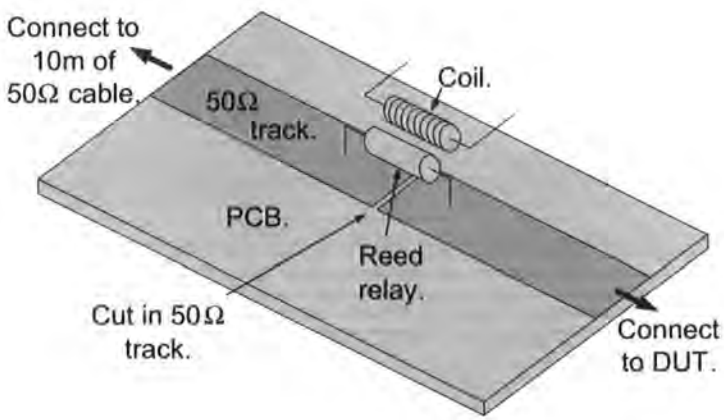

Figure 2.12: The physical construction of the output switch of the TLP tester

The actual TLP specification achieved is detailed in Table 2.2. All aspects of the target specification were achieved except for the rise time of the signal. This was due to the capacitive loading ($1pF$) of the oscilloscope probe. An experiment was carried out using a $4GHz$ oscilloscope and a low capacitance probe and it was found that the measured rise time was around $140ps$. As this is close to the rise time limit of the oscilloscope, there is no way of knowing if the rise time could be faster still. The $4GHz$ oscilloscope was not used in the final design as the dynamic range of the scope is only $4V$ making it unsuitable for ggNMOS measurements.

2.8 Calibration

The TLP tester required calibration to remove the effects of parasitic components in the circuit. This is done by TLP testing known loads including a short circuit and an open circuit. Correction factors can then be calculated so that when an open circuit is measured, the load current is $0mA$ and when it is a short circuit, the load voltage is $0V$.

2.9 Conclusions

This Chapter has presented the TLP tester and shown that it can be constructed at relatively low cost, assuming that 50Ω cable, an oscilloscope, a power supply and an ammeter are readily available in the lab. The TLP provides considerable information as to the operation of the ESD protection clamp under ESD conditions (high current short duration) without damaging the circuit. In Chapter 3 the value of the TLP tester is demonstrated by showing how various components in an ESD clamp operate and therefore why the circuit fails to protect the output despite the ESD clamp being sufficiently large to conduct an ESD safely to ground.

2.10 Summary of contribution to knowledge

The TLP tester has been shown many times to be a valuable tool. In this chapter it has been demonstrated that a good quality TLP tester with performance parameters rivalling commercial TLP test equipment, can be constructed at very low cost as long as care is taken to ensure RF signal integrity at the critical points. The only expensive component required is an oscilloscope with at least $500Mhz$ bandwidth and it is commonly found in any research lab. Some TLP tester designs use compli-

cated corrective techniques to remove, by mathematical post processing, artefacts due to reflections between the TLP switch and the device under test (DUT). The TLP design presented here eliminates the need for these correction techniques by reducing the distance between the TLP switch and the DUT to less than $50mm$; this also has the advantage of decreasing the t_R to $< 140ps$ without using any special high bandwidth components.

Chapter 3

Investigation of the BC01b

In this Chapter the TLP test is assessed and the correlation between the HBM and TLP tests is investigated. Previous research has shown that the TLP test can correlate well to the HBM test [122][123][124][125][126][127]. In the case of the BC01b it has been confirmed that there is good correlation.

The difficulties of protecting RF CMOS IO circuits are also analysed. Further TLP tests were carried out on the BC01b which suggested that there were inefficiencies in the design. The IO circuit was investigated using a Focused Ion Beam (FIB) apparatus. The FIB was used to cut tracks in the metal layer of the chip so that individual components could be TLP tested in isolation and thus improve understanding of how the complete circuit operates under ESD conditions.

3.1 Context of the BC01b

The BC01b is a single chip mixed-mode CMOS Bluetooth solution made by Cambridge Silicon Radio (CSR) which requires only a few passive components to complete the Bluetooth circuit and has low power consumption. Industry standard ESD

testing has been carried out by a commercial test house on the BC01b [138]. The results of these tests showed that the RF outputs were protected up to $600V$ HBM [6]. This level of ESD protection is poor compared to standard IO pins and the aim in this Chapter is to investigate why this is.

The role of ESD protection structures is to absorb or divert the energy dissipated during an ESD event that is conducted through an IC but at the same time allow 'legitimate' signals to pass through the IO pins within the limits set in the specification. Other requirements have to be met including using the minimum of silicon area as well as using current process technologies for manufacture. RF circuits place an extra emphasis on the requirement to allow 'legitimate' signals through the IO pins. The BC01b RF outputs operate at $2.4GHz$ [139][140]; they are therefore prone to losses due to the lossy parasitic capacitance of ESD protection structures. For this reason the structures have to be as small as possible whilst still providing adequate protection.

It would be beneficial to increase the level of ESD protection to $2KV$ HBM which is the common level of protection for most IO pins. In this Chapter we analyse the existing protection structures and consider how these may be improved.

3.2 Description of the Protection Structures

The BC01b's TX_A and TX_B use grounded gate NMOS (ggNMOS) protection structures (Figure 3.1). D1 provides negative ESD protection up to $1000V$ HBM and D2, ggNMOS1 and ggNMOS2 provide $600V$ HBM ESD protection [138]. NMOS3 are the output driver transistors. All NMOS transistors are of standard design. The grounded gate devices play no part in normal operation of the chip, they are only active during an ESD event. The ESD current is conducted to ground using the snapback mechanism by either ggNMOS1 or ggNMOS2 and D2 or both.

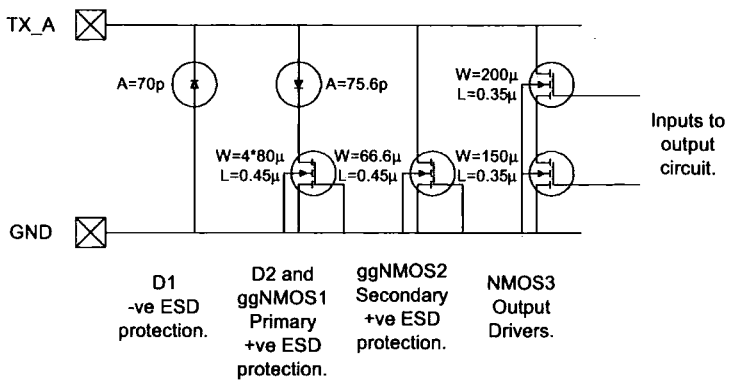

Figure 3.1: Output stage of the BC01b's TX_A which is identical to TX_B

3.3 Testing of the BC01b RF Outputs

The RF IO's IV profile during the high currents typically found in ESD events was measured using a TLP tester described in Chapter 2. Parts of the IO circuit were isolated with the FIB so that an individual section of it could be characterised by the TLP tester.

3.3.1 FIB Operation

Part of this research program uses a FIB. A FIB is very similar to a Scanning Electron Microscope (SEM), the only major difference is that it accelerates Ga^+ ions instead of electrons. With a low intensity beam it is capable of imaging ICs in much the same way as SEMs (Figure 3.2). However, high intensity beams can be used to mill selected areas of the chip to a desired depth on a sub micron scale (Figure 3.3). This allows cross sections of the chip to be made for inspection; it also enables the user to modify the circuit on the chip by cutting the metal tracks (Figure 3.4). The kinetic energy of the ions displaces the material found in its path causing both the Ga^+ ions and the milled material to be spluttered on to adjacent surfaces. Most of the time this spluttering is harmless though, in some cases (as the spluttered material is conductive) metal tracks can be shorted out thus adversely

affecting the properties of the circuit. Because of this, when the FIB is used to mill tracks care has to be taken to ensure that all exposed surfaces of metal are milled last, thus ensuring that metal surfaces are clear of spluttered material. The FIB has been described in detail by Reyntjens et al. [4].

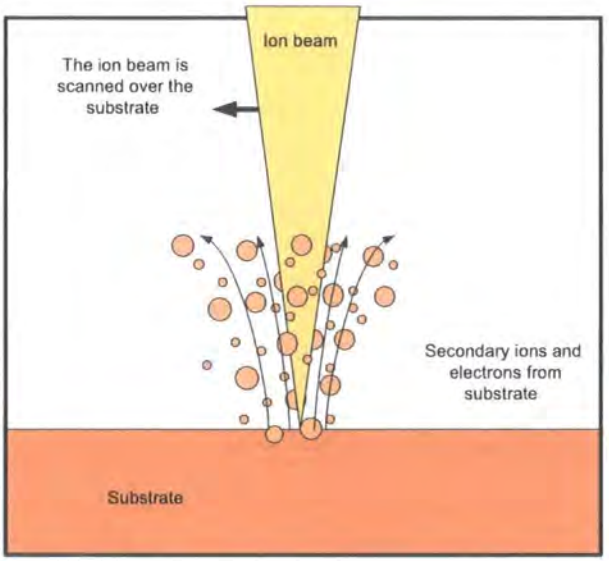

Figure 3.2: The principle of FIB imaging [4]

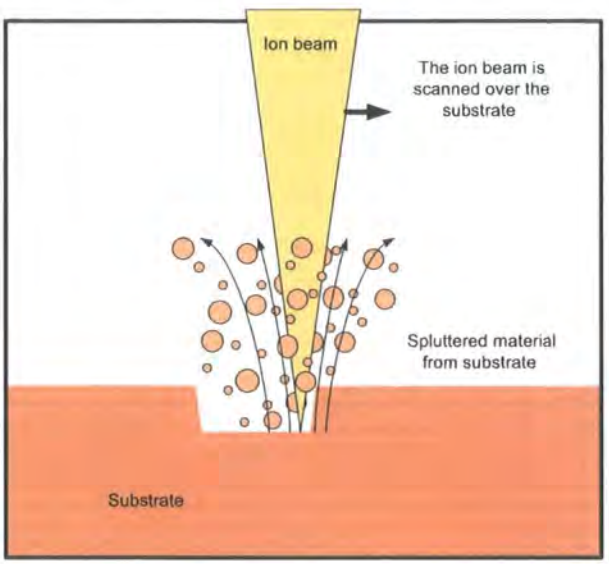

Figure 3.3: The principle of FIB milling [4]

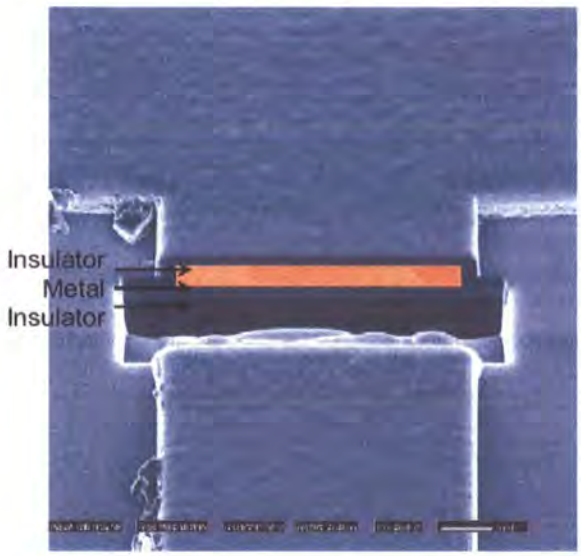

Figure 3.4: An example of milling where a metal track is cut [4]

3.3.2 TLP Test Equipment

A constant impedance TLP tester was built to the design described in Chapter 2 and illustrated in Figure 2.5. The design was chosen because it only required relatively low voltage power supply of around $5V - 150V$. Other designs (constant impedance TLP tester with a 20db attenuator as shown in Figure 2.6 and constant current TLP tester as shown in Figure 2.7) require a $2KV$ power supply to achieve reasonable currents. The TLP tester produces $100ns$ pulses up to $150V$ and currents of up to $2.5A$; it is computer controlled so that entire IV traces can be automatically generated. The rise time is normally $1ns$ but can be increased to $30ns$.

3.3.3 TLP Testing Method

The TLP set up was used to test the BC01b RF output's ESD performance. The output pins were positive voltage ESD tested to destruction using the TLP. The negative ESD performance was ignored as it is known to be significantly better than the positive performance [138].

A plot of V_{DUT} vs *time* for a TLP_{20V} test is shown in Figure 3.5. At (1) the voltage rises sharply to the snapback trigger voltage and overshoots to (2). Once the device is triggered V_{DUT} snaps back to (3) and is fixed until the end of the test pulse (4). Figures of merit are the rise time which is dependent on the tester, the time taken to snapback (5) and the snapback voltage (4), V_{hold}, all of which are dependent on the DUT. V_{hold} is calculated by taking a mean of all the voltage readings from $50ns$ to $90ns$ after the initial rising edge (6) as the circuit voltage is stable.

For the first test the TLP was charged to $5V$ and for each subsequent test the TLP charge voltage was increased by $1V$. Between each test I_{leak} was measured using a $3.3V$ bias and a pico-ammeter. When $I_{leak} > 1uA$ the test was stopped as the device was deemed to have sustained permanent damage.

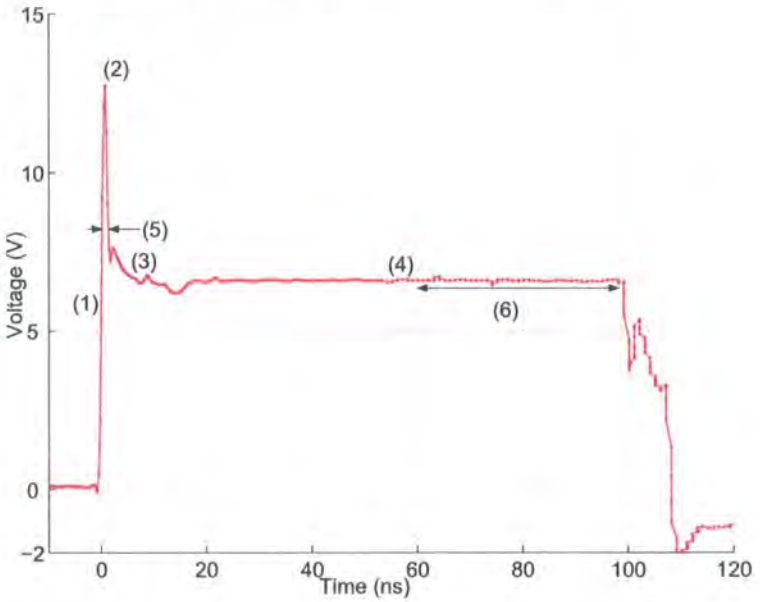

Figure 3.5: The voltage across the DUT vs time for a TLP-50Ω tester

For each TLP discharge the voltage across the device was measured using a $500Mhz$ oscilloscope sampling at $2GS/s$ and the sampled data was used to derive the IV characteristics of the device. The current through the device is calculated using Equation 3.1.

$$I_{hold} = \frac{V_{TLP} - V_{hold}}{50\Omega} \tag{3.1}$$

An example IV plot is shown in Figure 3.6. The graph shows that up to $11V$ (1), I_{TLP} is less than a few mA (actually the data shows that $I < 2mA$). Once the tester is charged to $12V$, snapback is triggered (2), V_{hold} is $7.7V$ and I_{hold} is $86mA$. This result shows that the V_{t1} for these test conditions is between $11V$ and $12V$. Subsequent pulses at increasing TLP voltages (3) show that as the TLP energy increases, I_{hold} increases significantly and V_{hold} only increases slightly. At high TLP energies the device undergoes permanent damage. This can be detected by measuring, between TLP tests, the I_{leak} of the DUT whilst biased at the normal operating voltage which, in this case, is $3.3V$. I_{leak} was measured after every test and plotted against I_{hold} in the form $log_{10}(I_{leak}) vs I_{hold}$. The graph in Figure 3.6 shows that I_{leak} is constant at around $4nA$ (4) for most of the TLP measurements. The final TLP test causes it to increase to $200uA$ (5) at which point the device is considered to have undergone irreversible damage.

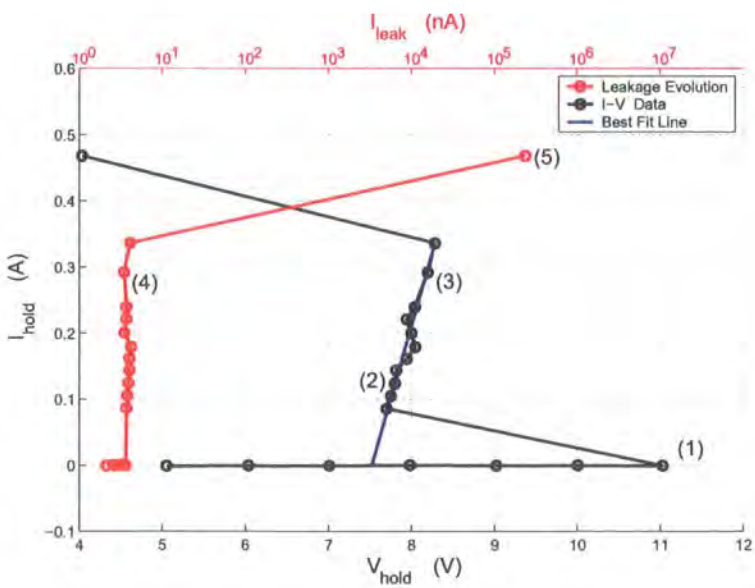

Figure 3.6: Example of a ggNMOS IV plot derived from TLP testing

3.3.4 TLP to HBM Test Fidelity

Barth et al. [123] and Stadler et al. [126] provide examples to show that there is a correlation between the holding current at failure for a TLP test and the peak failure current of an HBM test. Both papers also provided examples where this correlation did not occur. Sometimes the miscorrelation could not be explained although Barth et al. [123] show that miscorrelation could be attributed to incorrect TLP rise time.

Twelve devices were tested to determine the correlation between the TLP tester that was built by the author and a commercial HBM tester previously used to test the BC01b. The mean TLP I_{t2} was $507mA$ and the minimum and maximum I_{t2}s were $382mA$ and $680mA$ respectively. The JEDEC HBM test specification [6] requires the HBM test to be applied to three chips. For it to pass the JEDEC test, all three chips must conform to the published specification after the HBM test has been applied. For the BC01b, the HBM pass voltage was found to be 600V equivalent to a peak current of $400mA(600V/1500\Omega = 400mA)$. This is a worst case figure as all three chips have to be able to sustain this level of ESD energy. The worst case HBM failure current of $400mA$ is reasonably close to the worst case TLP failure current of $382mA$ showing that for the BC01b RF IO circuits, TLP/HBM correlation is good.

3.4 Initial TLP Tests of the BC01b RF Outputs

A BC01b sample was tested to destruction using the TLP. The IV characteristic was measured; test results are shown in Figure 3.7.

These results show that when the TLP is charged to $10V$ or less the DUT current is so small that it can be assumed to be zero (1) as all these measurements $I < 2mA$. After the TLP is charged to $11V$, the device undergoes snapback and draws $91mA$ and holds the TLP output to $6.43V$ (2). As the TLP charge voltage is increased

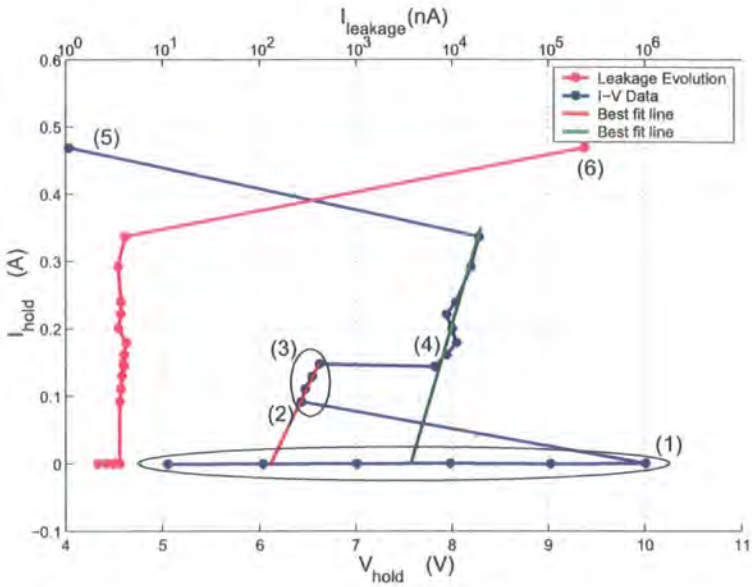

Figure 3.7: IV curve and I_{hold} vs I_{leak} for the BC01b derived from a TLP 50Ω tester

the current is increased but the hold voltage only rises slightly (3); this increase in V_{hold} is due to the series resistance on the protection device. At (4) the hold voltage rises significantly to $7.81V$, which naturally decreases the hold current as the TLP impedance is 50Ω. Again, any further increases in TLP energy cause small increases in V_{hold} and the series resistance of the protection structure. The data point at (5) indicates that permanent damage has occurred to the output. This can be confirmed by looking at the leakage current which increased significantly after the last TLP pulse (6).

The step change in V_{hold}, between (3) and (4), is clearly not due to the series resistance of the protection circuit. The red and green best fit lines in Figure 3.7 are used to calculate the V_{hold} (at $I = 0$) and R_{on} of the snapback devices. At low TLP energies the snapback voltage is $6.1V$ with a series resistance of 3.4Ω (red line); at high TLP energies the snapback voltage is $7.5V$ with a series resistance of 2.45Ω (green line). It shows that there are two distinctly different V_{hold} and R_{on} in this circuit. This suggests that there are two transistors in the circuit capable of undergoing snapback; one of them snaps back at low TLP energies and the other at high TLP energies. This correlates well with the circuit design in Figure 3.1 which shows that

there are two separate ESD protection devices. It would seem reasonable to assume that the two transistors are responsible for each of the two values of V_{hold}, but it is not clear why one transistor appears to trigger with high energy TLP pulses and the other with low energy TLP pulses. It is known [123] that the trigger mechanism is dependent on rise time (t_R) of the ESD pulse. Barth et al. [123] state that snapback can be triggered by a capacitively coupled charge which would imply that the slew rate (dV/dt) was more important than t_R. The most obvious difference between a low energy TLP pulse and a high energy TLP pulse is the slew rate (dV/dt) of the rising edge. All the TLP pulses had the same rise time but a 50V TLP pulse will have twice the slew rate of a $25V$ pulse.

A further test was carried out on another BC01b sample. Two TLP pulses were used, one with a fast rise time ($1ns$) and the other with a slow rise time ($20ns$). In this experiment, it was expected that for the fast rise time the results would be the same as Figure 3.7, but for the slow rise time, the TLP pulse should induce snapback at around 6V for both high and low energy pulses. The explanation for this result is that the dV/dt of the low rise time TLP pulse is low for *both* high and low energy pulses (Figure 3.8).

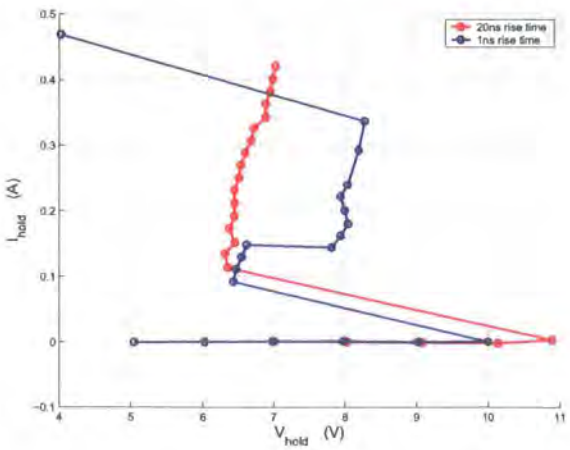

Figure 3.8: IV characteristics of the BC01b for 1ns and 20ns rise times

The blue trace in Figure 3.8 is the IV trace shown in Figure 3.7 where t_R is $1ns$. The red trace is the IV data for the same type of device when TLP tested with a pulse rise

time of $20ns$. The reader should note that the hold voltage is the same for *both* high and low energy $20ns$ rise time TLP pulses. The $1ns$ test was done using a different device to the $20ns$ device which explains the small difference in V_{hold}. To confirm the hypothesis that the two protection structures undergo snapback separately in this IC, further investigations were carried out. They are reported in next section.

3.5 Isolated Testing of the Protection Structures

The aim in this series of experiments was to determine the performance of ggNMOS1 and ggNMOS2 (Figure 3.1) under ESD conditions when operating independently of each other. The BC01b was chemically depackaged in acid so that the internal circuit could be modified using a FIB to isolate individual transistors. The transistors were characterised using the TLP.

3.5.1 Preparation of the BC01b to Allow Modification with the FIB

An IC was placed in hot ($70C°$) HNO_3 for around 15 minutes to remove the chip packing whilst leaving the metal layers and the die intact. The result was a fully operational BC01b die including the gold wire bonds. The die was glued on a glass substrate with TX_B and RF_Ground wire bonds glued to two terminals on the glass substrate. These terminals allowed quick and easy connection to the TLP tester. An example glass substrate is shown in Figure 3.9.

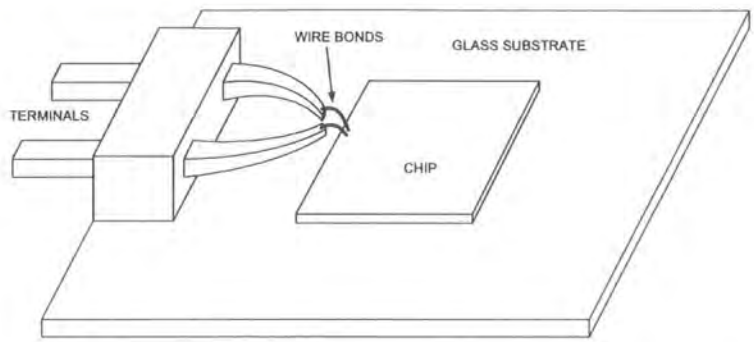

Figure 3.9: Depackaged chip mounted on a glass substrate

3.5.2 Testing for TLP Correlation between a Packaged and Depackaged Chip

The ESD performance of a chip was tested using the TLP to verify that the depackaging process did not affect the electrical characteristics of the chip. The resultant IV trace of the depackaged device is compared to the IV trace of a packaged device, taken from Figure 3.7, in Figure 3.10 which clearly shows that they are very similar.

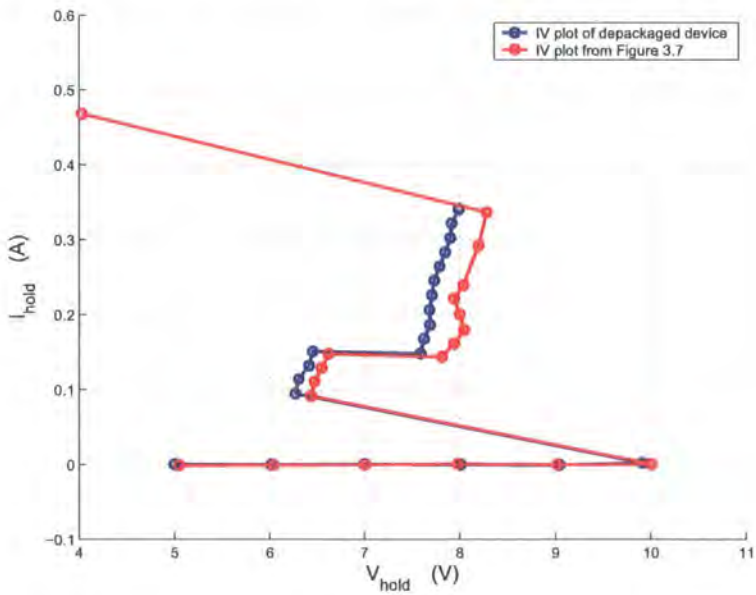

Figure 3.10: IV plot of a depackaged device

Four chips were depackaged and tested to destruction using the TLP and these results were then compared to the TLP failure voltages of packaged devices. Figure 3.11 shows the two sets of results. As only 13 devices were tested we have

insufficient data points to produce statistically significant results. Ignoring the data point at 61V there appears to be two clusters of failure voltages, one at around 28V and the other at around 41V. The lower voltage failure mode has only been measured for packaged devices. This result would suggest that there are two different failure modes or competing triggering mechanisms. More data points and further investigation is required. The graph does show that the failure voltages of the packaged and depackaged devices are broadly similar.

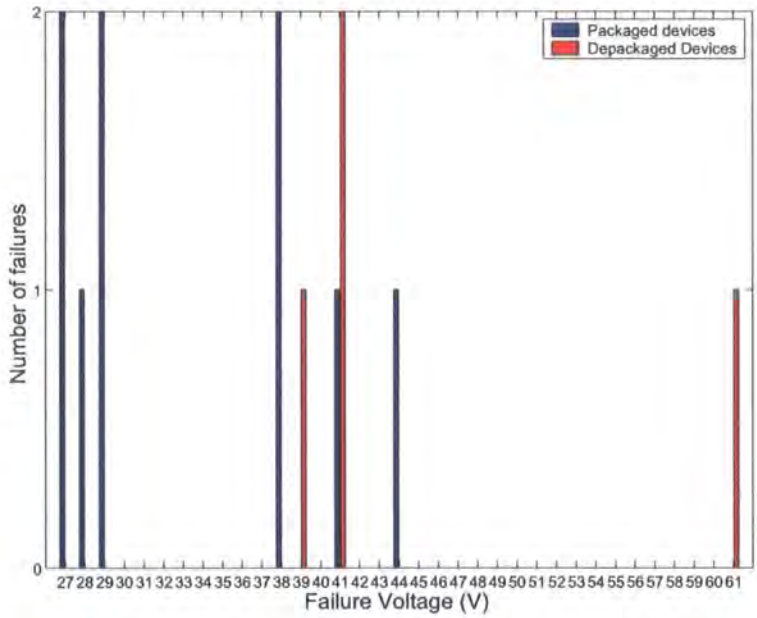

Figure 3.11: Number of failures vs Failure voltage of packaged and depackaged devices

The fact that depackaged devices did not fail at 27V-29V is probably due to the decreased parasitic capacitance in the depackaged devices as this would increase the overall t_R. Barth et al. [123] and Litzenberger et al. [26] have shown that this would increase overall ESD tolerance as it would decrease the effective TLP rise time.

3.5.3 Overall Test Method

Once the die is mounted on the glass substrate, it is tested using the TLP. The initial TLP testing is limited to 25V to prevent permanent damage. The chip is

then modified using the FIB and retested to destruction using the TLP. Test cuts were carried out using the FIB on a number of devices where a portion of the metal track was removed without altering the IO circuit (i.e. without creating any open circuits). The test devices were TLP tested before and after the FIB test cut to verify that the action of cutting the metal did not impact the electrical performance of the individual components.

The layout of the metal layers of the BC01b RF outputs is shown in Figure 3.12. Figures 3.13 and 3.14 show a schematic and a FIB image of how the circuit was modified using the FIB so that ggNMOS1 could be TLP tested in isolation. Figures 3.15 and 3.16 also show a schematic and a FIB image of how ggNMOS2 and the output drivers were isolated using the FIB. The output drivers and the ggNMOS2 protection structure cannot be isolated from each other using the FIB as they are connected together using M1 i.e. the metal layer closest to the silicon and therefore the most difficult layer to access with the FIB.

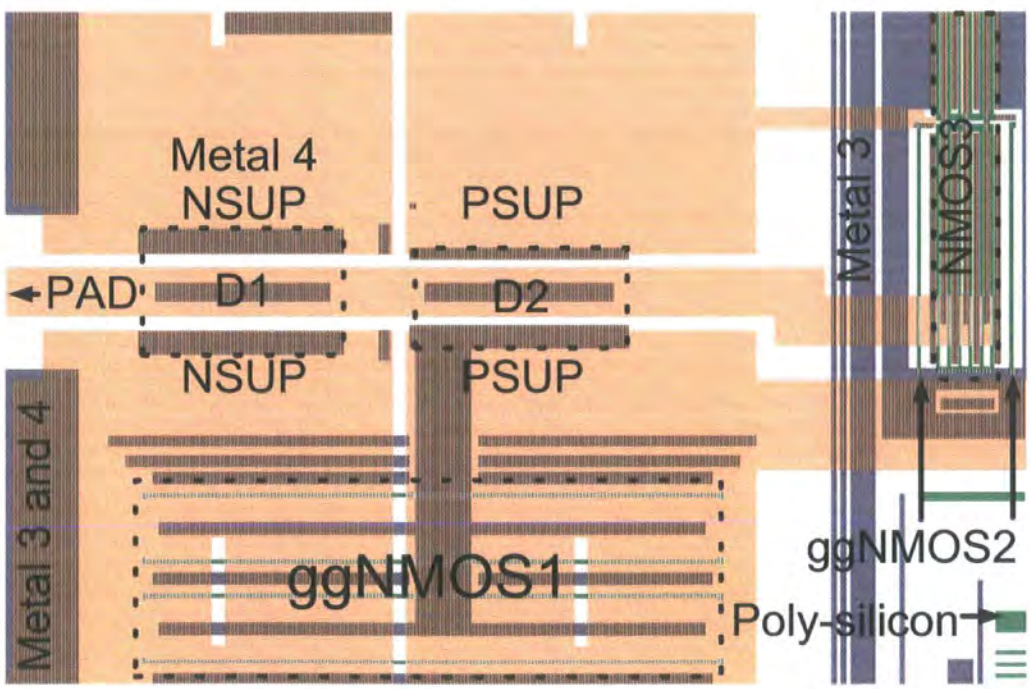

Figure 3.12: The layout of the metal layers in the output driver

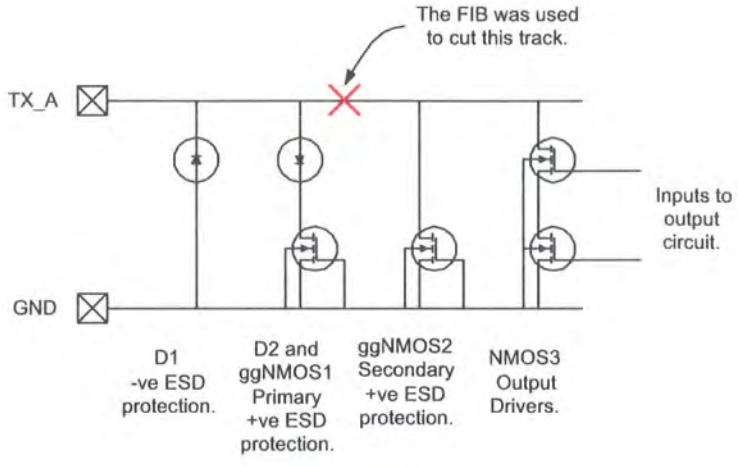

Figure 3.13: The position of the FIB cut to remove ggNMOS2 and the output drivers

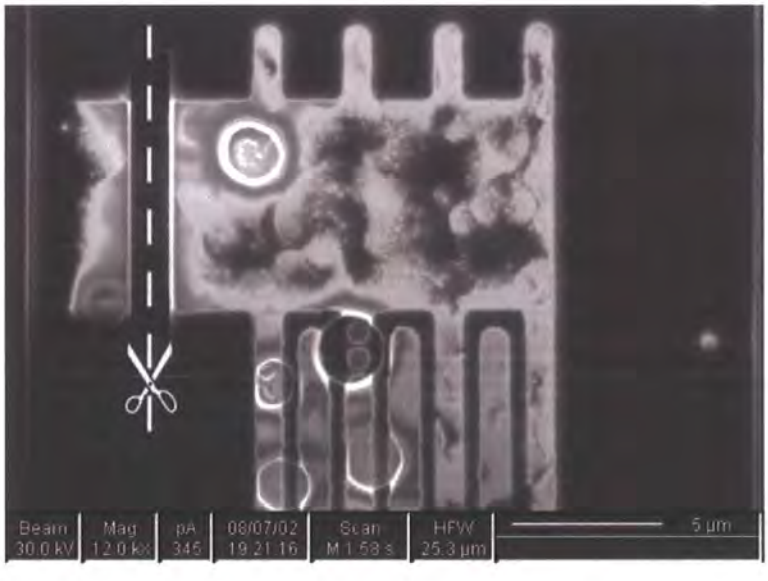

Figure 3.14: An image of the FIB cut to remove ggNMOS2 and the output drivers

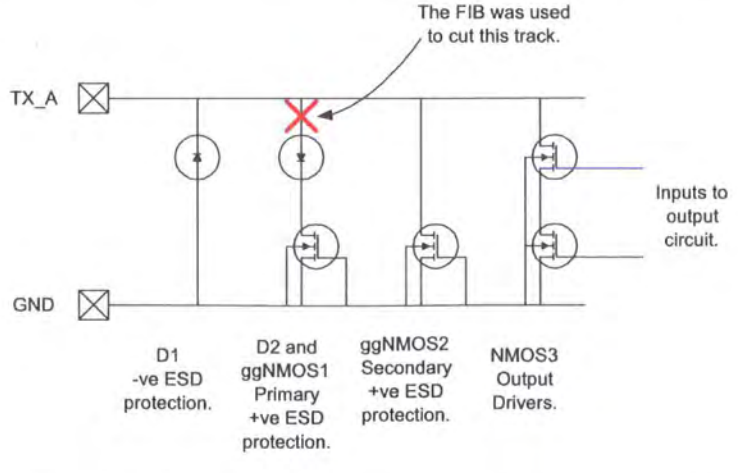

Figure 3.15: The position of the FIB cut to remove ggNMOS1

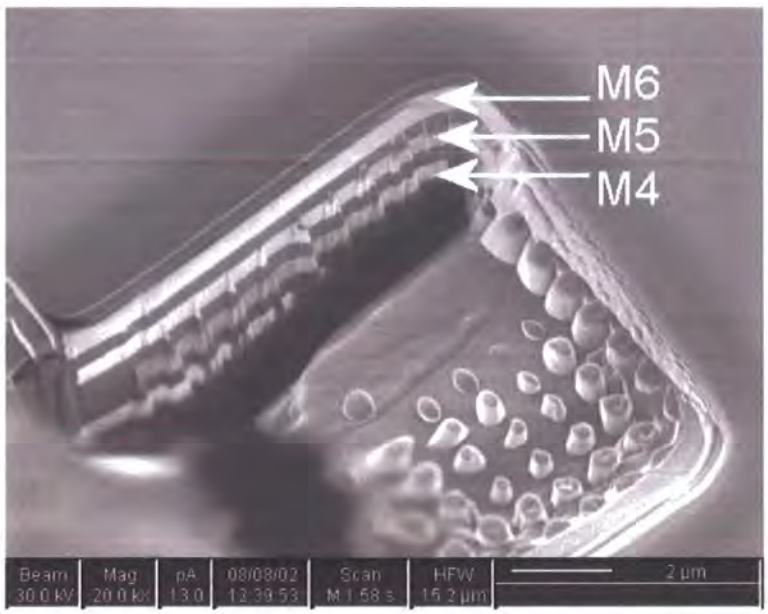

Figure 3.16: An image of the FIB cut to remove ggNMOS1; M4 is used to connect ggNMOS1 to the output

3.5.4 Testing the TLP Performance of ggNMOS1

It was found that neither samples failed at the maximum automated test voltage of $125V$ and when both were retested manually at $165V$, again neither failed. The TLP IV plot of both devices before and after FIB modification is shown in Figures 3.17, 3.18 and 3.19.

From the results it would appear that the failure voltage/current (V_{t2} and I_{t2}) of the ggNMOS1 is far in excess of the failure voltage/current of the circuit as a whole. This indicates that this component is not the weakest point in the circuit and that it should be possible to increase significantly the overall performance of the circuit by ensuring that this device triggers during all ESD events.

Figures 3.18 and 3.19 show that before modification the hold voltage is around $6V$ for low TLP voltages (A) and then as the TLP voltage is increased the hold voltage jumps to around $7.5V$ (B). After the devices were modified the V_{hold} increased to $\sim 7.5V$ for low and high voltage TLP pulses (B).

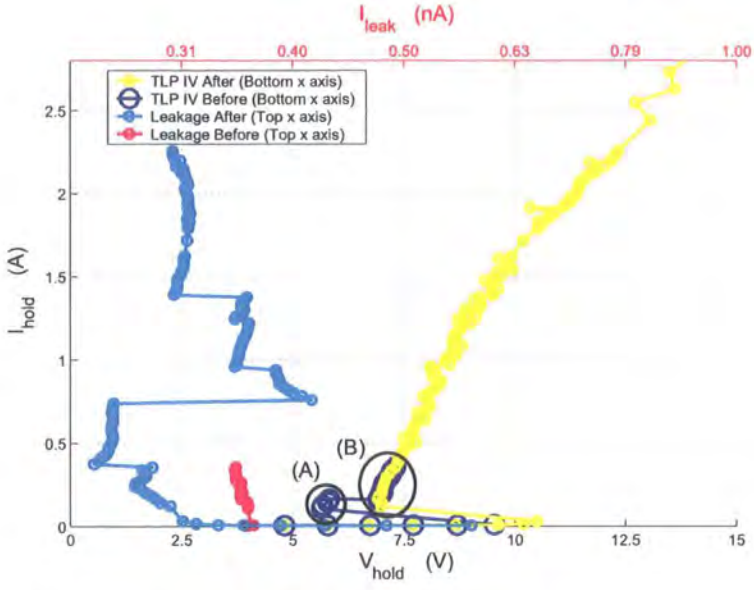

Figure 3.17: TLP IV curve of TX_A before and after ggNMOS2 and the output drivers were removed (first sample)

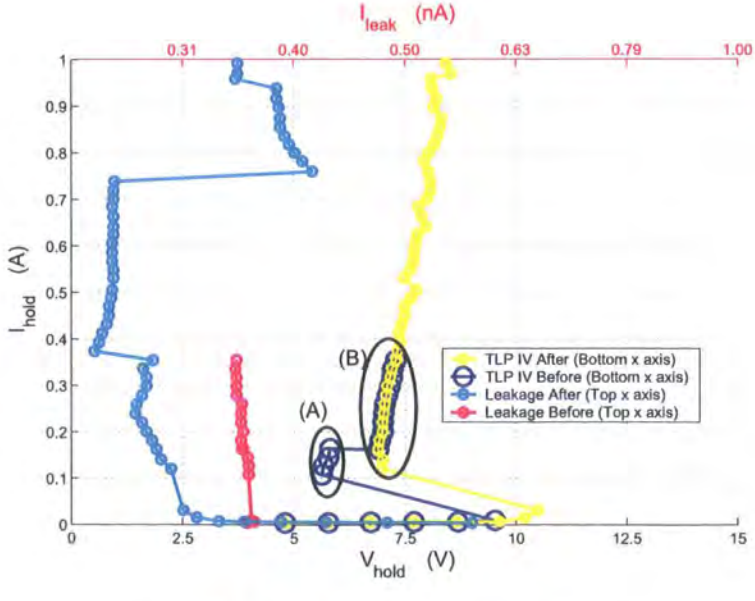

Figure 3.18: Detail from Figure 3.17

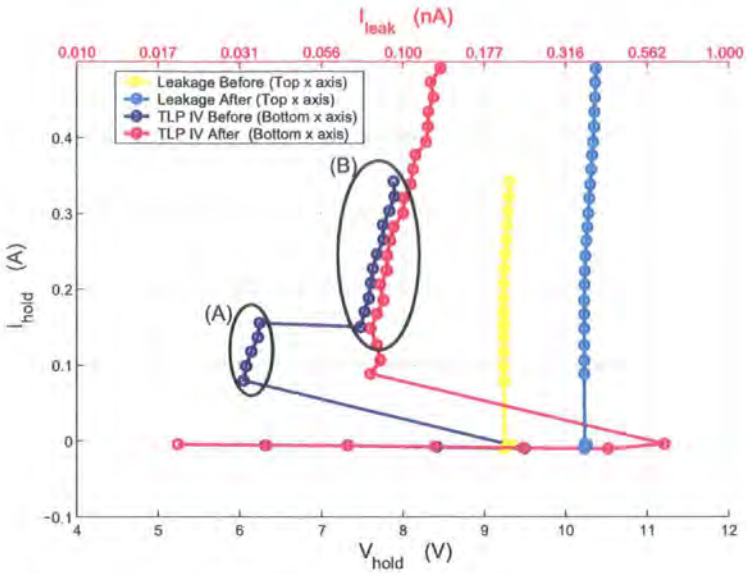

Figure 3.19: TLP IV curve of TX_A before and after ggNMOS2 and the output drivers were removed (second sample)

Note that Figure 3.17 shows that the failure voltage is $> 13V$ which is higher than the trigger voltage. It is well known that for a ggNMOS design to work well, the V_{t2} has to be greater than V_{t1} as this means that the whole width will be triggered and the transistor will operate efficiently.

3.5.5 Testing the Triggering Characteristics of ggNMOS1

The same circuit was built and modified as described in section 3.5.4 using new chips so that triggering voltage (V_{t1}) could be measured. This was achieved by applying TLP pulses to the chips starting at $10V$ and increasing the voltage by $10mV$ until the device triggered. This test was repeated using three different t_R of $1ns$, $2.3ns$ and $16ns$ (Table 3.1).

These results show that V_{t1} for ggNMOS1 is around $14V$ for slow and medium t_R and that the V_{t1} decreases significantly for fast t_R. All the measured values of V_{t1} of ggNMOS1 are significantly higher than the V_{t1} of an unmodified chip. This suggests that, in the original circuit, the ggNMOS1 does not turn on when the TLP

	Before FIB cut (the whole output circuit)			After FIB cut (just ggNMOS1 and D2)		
t_R	1ns	2.5ns	16ns	1ns	2.5ns	16ns
Device1	10.24V	10.63V	10.68V	12.65V	14.18V	14.09V
Device2	10.20V	10.61V	10.67V	12.68V	14.20V	14.20V
Device3	10.23V	10.62V	10.68V	12.68V	14.09V	14.05V
Mean	**10.22V**	**10.62V**	**10.68V**	**12.67V**	**14.15V**	**14.11V**

Table 3.1: Comparison of triggering voltages for ggNMOS1 and an unmodified device

is operating at around the trigger voltage (this is inherently a slow dV/dt condition) as the secondary protection will turn on first. This also explains why the hold voltage of the unmodified circuit at low TLP energy is lower than the hold voltage for the modified circuit. Figures 3.18 and 3.19 show that at higher TLP energies V_{hold} remains unchanged for modified and unmodified devices. Given that cutting out ggNMOS2 and the output drivers does not change the IV curve for high energies, this result would suggest that for high TLP energies, ggNMOS1 triggers. A possible explanation is that the V_{t1} of ggNMOS1 decreases by $1.48V$ if t_R is increased from $2.5ns$ to $1ns$ whereas the unmodified device V_{t1} decreases by $400mV$. It may be the case that further increases in dV/dt, which occur during higher energy TLP pulses, may allow the V_{t1} of ggNMOS1 to fall below that of ggNMOS2 and therefore trigger before ggNMOS2. This would result in ggNMOS2 having no effect on V_{hold} at high TLP energies.

3.5.6 Testing the TLP Performance of ggNMOS2 in Parallel with the Output Drivers

Ideally, the output drivers and ggNMOS2 should be tested separately. However, it is not possible to isolate these two transistors using the FIB as the drains of the two devices are connected together in M1 (these are the metal tracks closest to the silicon surface) and therefore cannot be cut by the FIB. So for this experiment the FIB was used to disconnect ggNMOS1 in three chips so that the performance of

the ggNMOS2 and the output drivers could be tested in isolation (Figure 3.15 and 3.16). These chips were then tested and all failed at $42V$. The IV curves are shown in Figures 3.20-3.22.

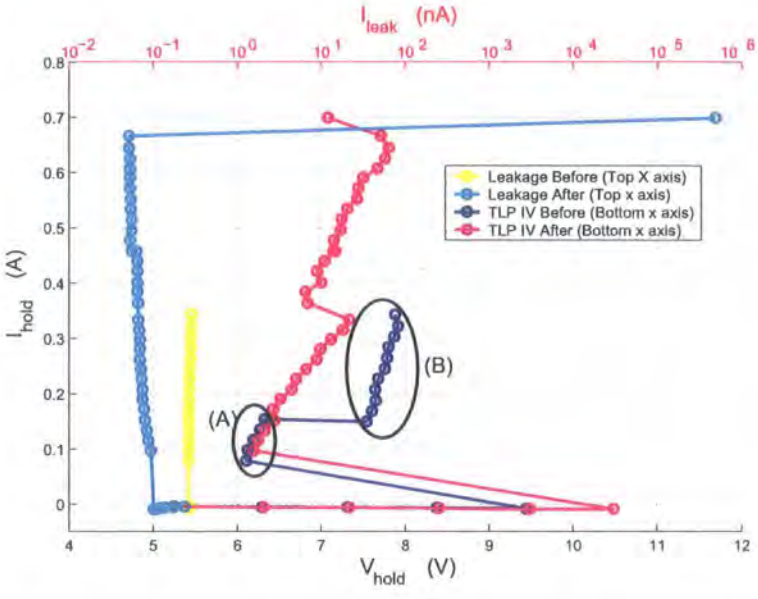

Figure 3.20: First IV curve of ggNMOS2 in parallel with the output drivers

All three graphs show that for low TLP energies the unmodified device triggers at the same voltage as the modified transistor (A). For high TLP energies the hold voltage is higher in the unmodified device (B) than it is for the ggNMOS2, the modified device.

3.5.7 Testing the Triggering Characteristic of ggNMOS2 in Parallel with the Output Drivers

A further set of devices were modified as in section 3.5.6 using new chips so that the V_{t1} could be measured. The triggering voltage for each device was measured in the same way as in section 3.5.5 (Table 3.2).

The data shows that V_{t1} does not change if the chip is modified. This indicates that when an unmodified chip is subjected to a TLP voltage of around V_{t1}, only the

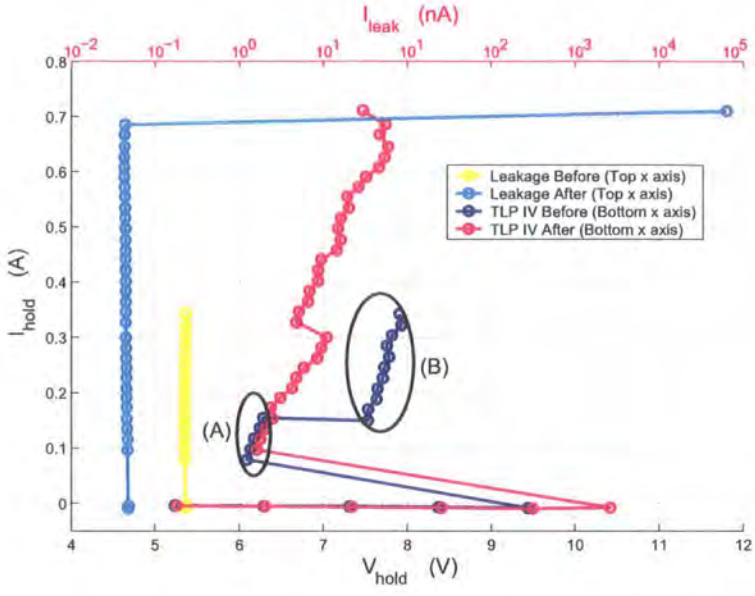

Figure 3.21: Second IV curve of ggNMOS2 in parallel with the output drivers

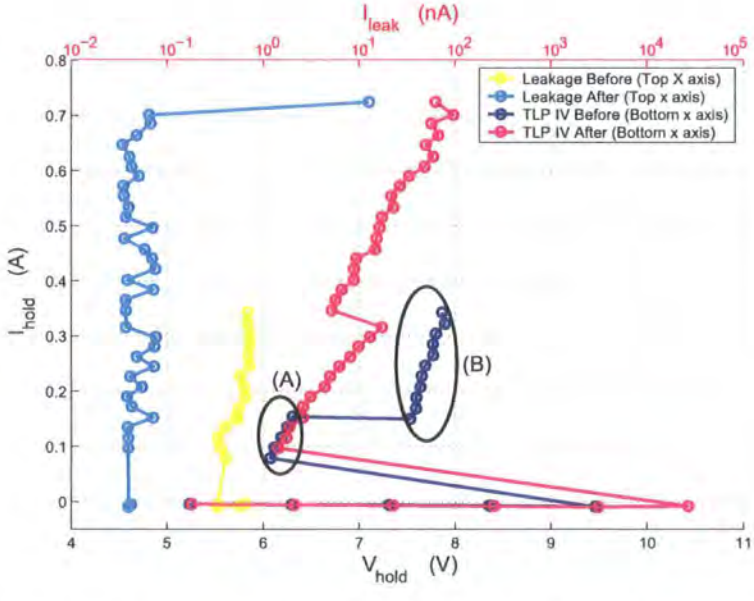

Figure 3.22: Third IV curve of ggNMOS2 in parallel with the output drivers

	Before FIB cut (the whole output circuit)			After FIB cut (just ggNMOS2 and NMOS3)		
t_R	1ns	2.5ns	16ns	1ns	2.5ns	16ns
Device1	10.39V	10.81V	10.88V	10.35V	10.79V	10.82V
Device2	10.30V	10.75V	10.75V	10.31V	10.74V	10.79V
Device3	10.22V	10.70V	10.75V	10.25V	10.68V	10.74V
Mean	**10.30V**	**10.75V**	**10.79V**	**10.30V**	**10.74V**	**10.78V**

Table 3.2: Comparison of V_{t1} for ggNMOS2 and an unmodified device

ggNMOS2 triggers.

3.6 Failure Analysis

ESD damage can be detected by measuring the leakage current. The BC01b outputs normally have a $I_{leak} < 10nA$ but after a high energy TLP pulse the leakage can rise by many orders of magnitude. This increase in leakage current can be used to signal that permanent ESD damage has occurred. It would be useful to know where in the output circuit the current is leaking as this would be the weakest point in the circuit. Figure 3.1 suggests that there are only two ggNMOS transistors in the output circuit, but in practice each ggNMOS transistor represents many transistors working in parallel. This allows the circuit to be compact whilst still having a large gate width. Figure 3.23 shows the complete output circuit of the BC01b including identical transistors wired in parallel.

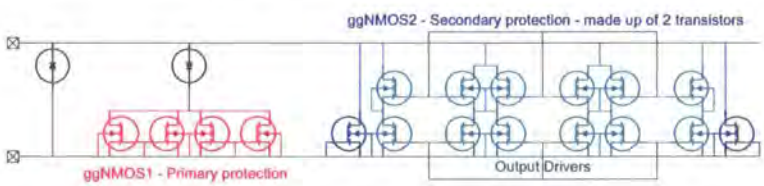

Figure 3.23: Output circuit of the BC01b showing each individual transistor including identical instances of transistors connected in parallel

Five devices (three packaged and two unpackaged) were TLP tested to destruction. The aim at this point is to find through which transistor the current is leaking. This was done by selectively cutting transistors out of the circuit, using the FIB, and measuring I_{leak} after every cut was completed. The location of the cuts is shown in Figure 3.24 and the numbering denotes the order of the cuts (1: first and 4: last).

After the first cut, I_{leak} was still high for all the devices which means that the leakage was not through ggNMOS1. When a cut was made at location 2, (Figure 3.24) I_{leak} in four of the devices fell to $< 10nA$. On the remaining device, when location 3 was

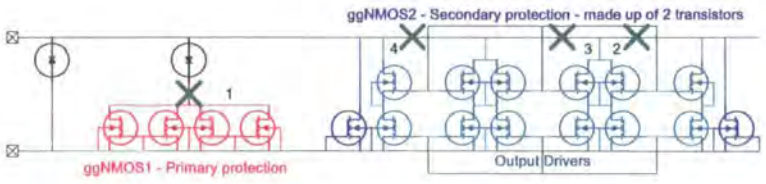

Figure 3.24: Output circuit of the BC01b showing the location of the FIB cuts to be done

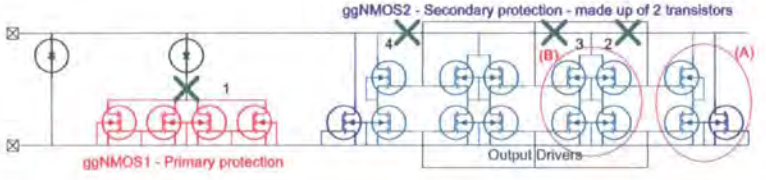

Figure 3.25: Circuit of the BC01b output. The circles indicate the location of circuit failure

cut I_{leak} also fell to $< 10nA$. This means that in four cases I_{leak} was due to the failure of transistors in circle A and in one case due to failure of transistors in circle B (Figure 3.25).

This would suggest that the weak point of the circuit is likely to be the ggNMOS2 transistor on the far right hand side in Figure 3.25. This result is not conclusive as the transistors in (A) also include two output drivers. Alternative techniques such as photo emission microscopy could be used to identify which of the two transistors is failing; on the other hand the circuit could be redesigned such that the output drivers and ggNMOS2 can be isolated using the FIB.

3.7 Conclusions

The results presented in this Chapter clearly demonstrate how a TLP tester can be used to understand how the circuit operates under ESD conditions. It would have not been possible to gain the same understanding using an HBM tester. We have also demonstrated how the FIB can assist investigations into a circuit's ESD performance without altering the performance of individual components.

The FIB has been able to provide conclusive evidence that there are two separate ggNMOS protection devices working within the output stage. ggNMOS1 has been proven to be able to divert high energy TLP pulses and therefore is capable of protecting the output drivers. When ggNMOS1 is removed from the circuit, the failure current (I_{t2}) is unchanged though we still do not know with absolute certainty that it is the ggNMOS2 is failing, as it could be the output drivers.

These experiments show that the performance of individual structures in isolation needs to be given careful consideration before they are placed in parallel. ESD protection structures must be designed in such a way that they, collectively, can conduct a high current (absorb a high voltage ESD event) before any individual component fails. It is likely that if the circuit were redesigned without ggNMOS2, it would have a significantly higher ESD performance.

3.8 Summary of contribution to knowledge

In this chapter it was shown that disconnecting components in a circuit using a Focused Ion Beam (FIB) is a valuable technique to help understand how individual components behave under ESD conditions and how they interact to produce the overall circuit performance. In this particular instance it was found that there were two possible current paths through the circuit. One was able to tolerate over $3A$ in ESD current whereas the other was limited to $\sim 300mA$. The eventual current path through the circuit was dependant on the ESD slew rate. For slow slew rate ESD events the weakest current path was activated, for moderate slew rate ESD events the strong current path was activated and for fast slew rate it is probable that both current paths were activated. It was possible to determine that the weak current path had a lower impedance than the strong current path. Therefore when both current paths were activated the ESD current would flow through the weak current

path causing the circuit to be damaged. The use of the FIB on the circuit did not appear to have any unintended effects on the circuit operation.

Chapter 4

SPICE Simulation - Analysis of Required Rise Times for Good Correlation Between TLP and HBM Simulations or Measurements

4.1 Introduction

When designing ESD protection structures the aim is to protect IO circuits from ESD discharges. In practice this means that the clamp circuit must be able to protect an IO circuit from a pseudo ESD discharge as defined by the standards bodies such as JEDEC in its Human Body Model (HBM) test [6]. As explained in section 2.2 the HBM test is designed to replicate real ESD events and is the test that the chip is required to pass. The HBM test only provides pass/fail information whereas a

Transmission Line Pulse (TLP) tester is designed to throw light on the internal operation of an ESD circuit during ESD discharge. However, as the current/voltage profile of a TLP tester differs significantly from that of an HBM tester, care is required when interpreting TLP data and comparing them to HBM results.

Results presented in section 3.3.4 and references discussed in section 2.5 have demonstrated that good correlation between HBM peak current and TLP average current can be demonstrated. In this section, the speed of a TLP pulse to produce good correlation with an HBM pulse is considered. Barth et al. [119] presented an argument stating that the initial rise time was the important factor (i.e. initial $\frac{dV}{dt}$) for good correlation; they also demonstrated that good correlation can be dependent on the rise time or $\frac{dV}{dt}$ of the ESD test equipment. Musshoff et al. [128] carried out similar investigations, came to similar conclusions though added that correlation should not be taken for granted as it is highly dependent on the ESD clamp architecture. In this section, this theory will be explored and SPICE simulations will be used to illustrate a possible solution.

4.2 Initial Rise Time

It is known that, for the purposes of good correlation, the rise time of an ESD test pulse is important [26][27][123] as many ESD protection structures (and inadvertently IO circuits) can be triggered by either DC voltages or capacitively coupled AC currents. Typically, it is a combination of both. For the purposes of this investigation, it will be assumed that the time taken for the voltage across the test circuit to reach the triggering voltage is the critical factor in assuring good correlation between different test methods (e.g. the average dV/dt between $0V$ and the trigger voltage). The JEDEC specification [6] states that the rise time of an HBM ESD discharge *current* must be between 2ns and 10ns (defined as 10% to 90%) though it

does not give any further details on the dV/dt that is required of the HBM pulse or the time taken to reach the triggering voltage.

4.3 Initial Rise Time of an HBM and TLP Tester

In this section, it is assumed that the peak HBM current is the same as the average $100ns$ TLP current as this has been shown to provide good correlation. For this assumption to be true, the TLP test voltage needs to be 30 times lower than the HBM test voltage, since the impedance of the TLP tester is 30 times less (50Ω vs 1500Ω). If it is also assumed that the shape of the rising edge of the TLP and HBM testers is a constant slope (Equation 4.1 where t_R is the rise time, and Figure 4.1: the blue line) then the *rise time* of the TLP tester needs to be 30 times faster than the rise time of the HBM tester (i.e. in the range $66ps - 333ps$) to ensure that the trigger voltage is reached in the same amount of time.

$$V_{OUT} = Amplitude * ((t > 0) * (t < 1.25) * \frac{t_R * t}{1.25} + (t \leq 1.25)) \tag{4.1}$$

$$V_{OUT} = Amplitude * ((t > 0) * (t < 1.7) * (0.5 - 0.5 * cos(\frac{1.85 * t}{t_R})) + (t \leq 1.7)) \tag{4.2}$$

Clearly ESD phenomena do not have a constant slope as a rising edge. It is likely to have a gaussian profile [128] which can be approximated by Equation 4.2 where t_R is the rise time (green line in Figure 4.1). For this type of rising edge, it is not possible simply to increase the rising rise time by a factor of 30 as this will not produce the same dV/dt between $0V$ and the trigger voltage for a TLP pulse. Equations 4.3 and 4.4 show the relevant portion of Equation 4.2 for both a TLP

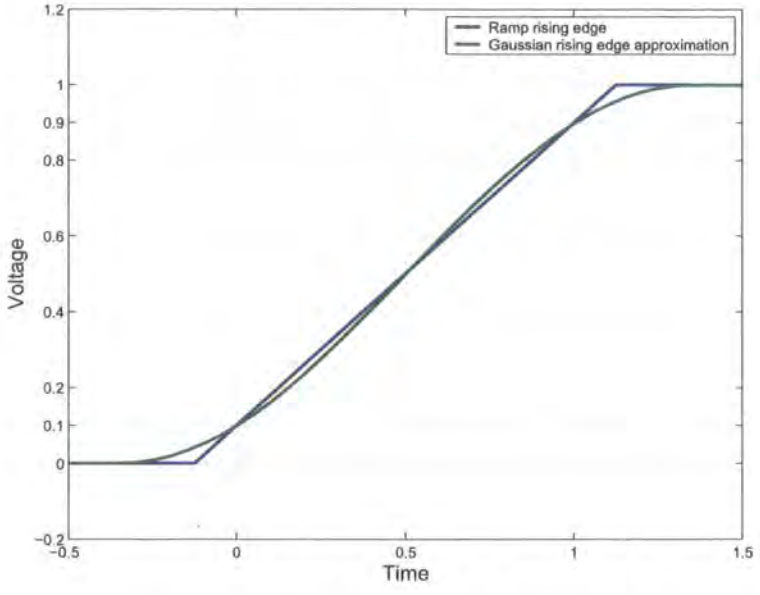

Figure 4.1: Two possible normalised rising edge profiles for a TLP or HBM pulse

and an HBM pulse respectively. t_{6V-TLP} and t_{6V-HBM} are the time taken for a $66V$ TLP pulse and a $2KV$ HBM pulse respectively to reach $6V$. The HBM rise time (t_{R-HBM}), as defined by JEDEC [6], ranges between $2ns$ and $10ns$ and the TLP rise time (t_{R-TLP}) needs to be calculated such that they both reach 6V at the same time (i.e. $t_{6V-TLP} = t_{6V-HBM}$).

$$(0.5 - 0.5 * cos(\frac{1.85 * t_{6V-TLP}}{t_{R-TLP}})) * 66.6V = 6V \tag{4.3}$$

$$(0.5 - 0.5 * cos(\frac{1.85 * t_{6V-HBM}}{t_{R-HBM}})) * 2KV = 6V \tag{4.4}$$

A $2KV$ $2ns$ HBM rising edge takes $118ps$ to reach $6V$ (a $10ns$ rise time takes $592ps$); for an equivalent energy TLP (66V) rising edge to reach $6V$ in the same amount of time it has to have a rise time of $360ps$ (or a rise time of $1.8ns$ for a $10ns$ rise time HBM pulse). For smaller amplitudes the time taken to reach $6V$ is longer but the required rise time of TLP pulse to be equivalent to the HBM pulse is broadly constant (Figure 4.2).

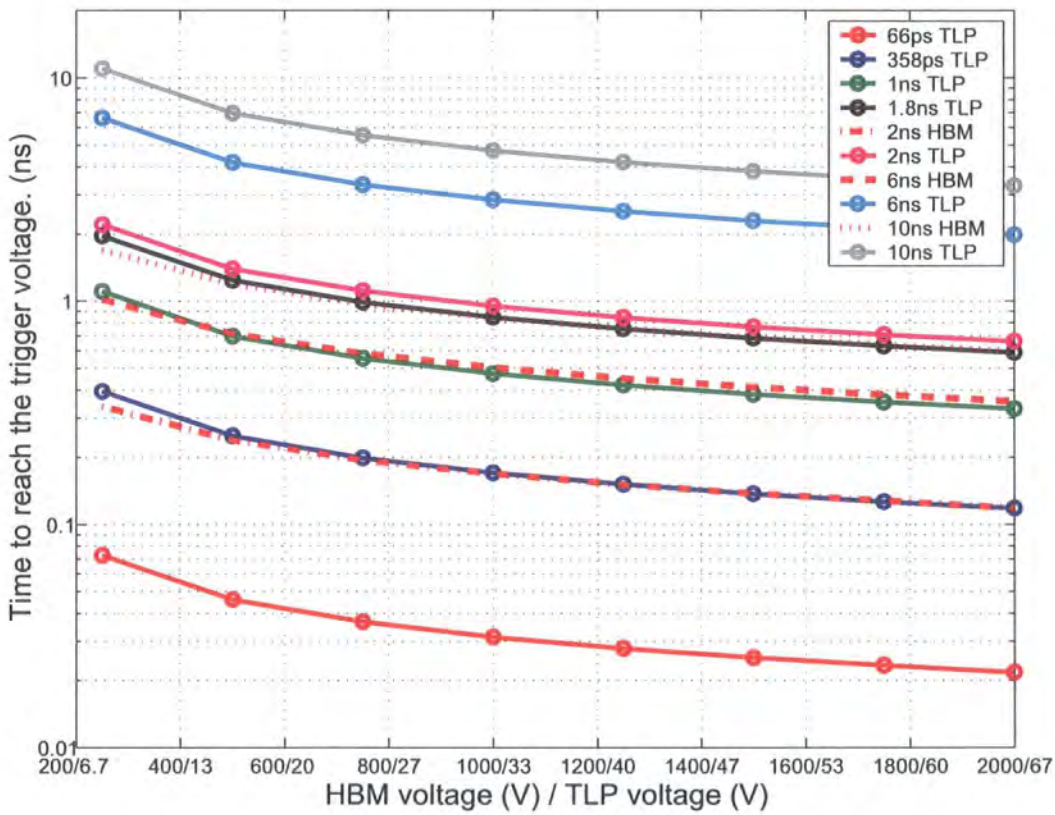

Figure 4.2: The time taken for an HBM or TLP source to reach $6V$ for a variety of rise times and amplitudes

Though this is a further refinement to the argument, it does not take into account the fact that an HBM pulse source has a higher impedance (1500Ω) than a TLP source (50Ω) and this will have an effect on the voltage source's ability to charge the parasitic capacitance of an ESD test structure. The initial rise time across the circuit being tested is of primary interest as ultimately this is the parameter that affects the test circuit's performance.

4.4 Device Initial Rise Time Under HBM and TLP Test Conditions

To determine how long it would take for the voltage across the ESD protection circuit (in this case, a 2KV ESD tolerant ggNMOS protection structure has been chosen) to reach 6V for a TLP tester and an HBM tester, a SPICE simulation was set up with a circuit as shown in Figure 4.3. The voltage source was programmed with the formula in Equation 4.2. The circuit was simulated over a range of HBM voltages, from 250V to 2000V, with a source impedance of 1500Ω and rise times of $2ns$, $6ns$ and $10ns$. The equivalent simulation was carried out for a TLP source: amplitude between 8V and 66V, source impedance of 50Ω and rise times of $360ps$, $500ps$, $1ns$, $1.8ns$, $2ns$, $6ns$ and $10ns$. The time taken for V_{DS} to reach 6V is shown in Figure 4.4.

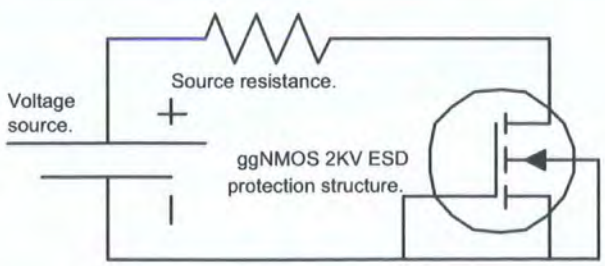

Figure 4.3: The SPICE simulation circuit so that the actual device $\frac{dV}{dt}$ can be derived

Using SPICE simulations for ESD may be questionable as the SPICE models are not designed for, and are not calibrated for, the voltages and currents that occur during ESD events. This is a valid point but the only figure of interest here is the capacitance up to the trigger voltage; the snapback operation is not relevant in this simulation as we are only interested in the time it takes for each test type, HBM or TLP, to reach the triggering voltage.

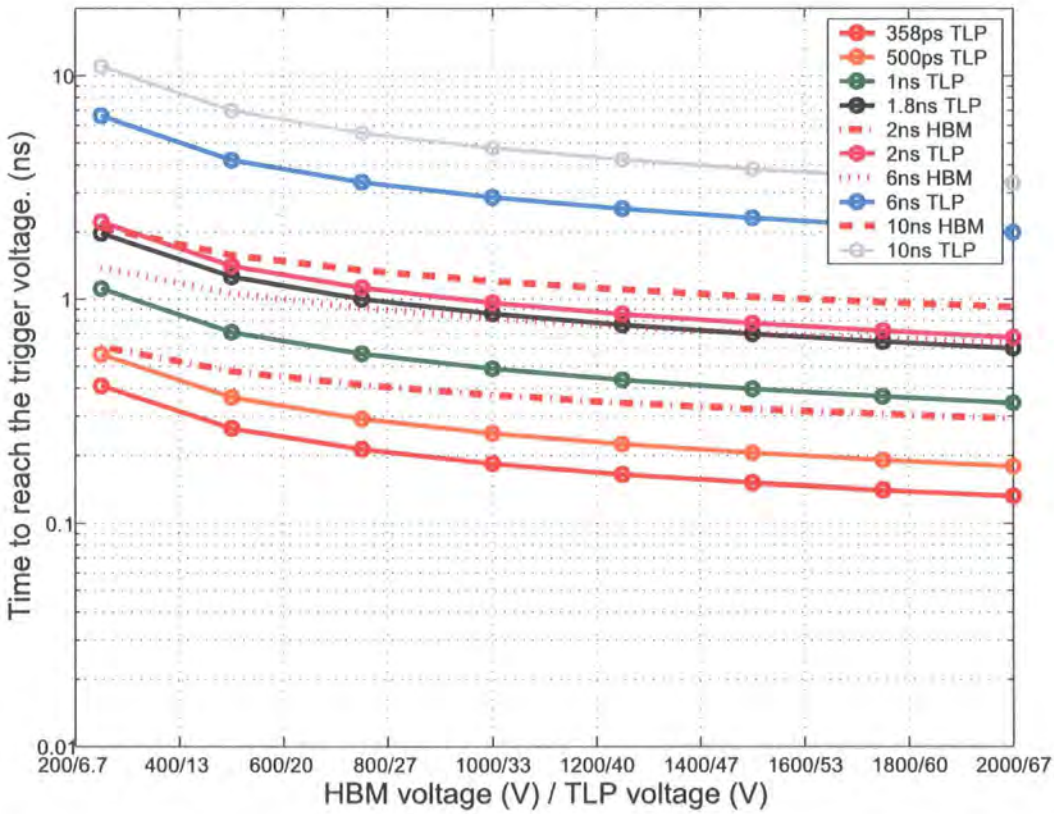

Figure 4.4: The time taken for a ggNMOS V_{DS} to reach $6V$ under TLP and HBM test conditions for a variety of rise times and amplitudes

4.5 Conclusions

Figure 4.3 shows that a TLP rise time in the range $1ns - 2ns$ is, at a circuit level, equivalent to an HBM rise time of $2ns - 10ns$ for a 2KV ESD tolerant ggNMOS protection structure. This result is dependent on the parasitic capacitance of the ggNMOS (or other DUT). As the parasitic capacitance is increased, the rise time

of the TLP needs to be increased to ensure good initial rise time correlation. This needs to be taken into account when setting the specification requirements of a TLP tester. This experiment did not take into account the other capacitances typically found in a circuit protected by the ggNMOS such as the output driver or the parasitic capacitance of the pad/packaging.

4.6 Summary of contribution to knowledge

TLP manufacturer Barth Electronics has published a number of papers extolling the virtues of TLP testers with a $t_R = 200ns$ [115][123][132][119]. They have certainly shown that the rise time is an important factor in the triggering voltage of ggNMOS but they assume that the voltage profile seen by the ESD protection circuit is identical to the voltage profile measured at the output of the test equipment. They failed to take into account the difference in source impedance of a TLP tester (50Ω) and an HBM tester (1500Ω) and the parasitic capacitance of a DUT. In this chapter, the time taken for an ESD source to increase the voltage across an ESD clamp circuit to the triggering voltage was SPICE simulated. It was shown that once the parasitic capacitance and tester source impedance are factored into the slew rate calculations the required rise time of a TLP tester is significantly increased, when compared to Barth's calculations, as the TLP circuit has a significantly lower source impedance than the HBM tester.

Chapter 5

Finite Element Electrothermal Simulation

5.1 Introduction

This Chapter describes how finite element electrothermal simulation was used to understand further how ESD protection structures and circuits behave during ESD events and also to investigate new protection structure designs. The advantage of this type of simulator, when compared to circuit simulators such as SPICE, is that they can, in theory, model any variation in the device parameters as well as any device that has yet to be built or a device under conditions that have yet to be measured. These tools are finding increasing use in industry due to the greater availability of computing resources. The disadvantage is that the solution given is prone to errors or even no results at all as the simulator fails to converge. This contrasts with SPICE models which are derived from curve fitting real world measurement data to a standard set of formulas and are only accurate during the measured operating modes of a standard transistor. These modes do not include the high current/voltage and fast transient conditions common during ESD events, nor

do they cover non standard layout geometries. Multi project wafers could be used to test any possible solutions in silicon though the transistors' geometry could only be varied in compliance with the design rules. The doping profiles of the transistors could not be altered as this would affect the performance of all the transistors on the wafer and also altering the doping is a complicated process that impacts many chip performance parameters. Test chips require a significant amount of effort to design, more than ten weeks to get them back from the foundry and they are also a costly test method. As a consequence, simulation tools are used to provide a means of rapidly assessing design ideas and options for ESD protection methodologies.

5.2 Aims and Objectives - Understanding the Fundamentals of ggNMOS Snapback

The aim in this Chapter is to understand how a ggNMOS operates under ESD conditions; to understand which factors affect the electrical performance of a ggNMOS and how this occurs.

5.3 Simulation

Within the ESD industry, circuits are modelled in a SPICE based simulator and devices are modelled in either a SPICE based model or a finite element electrothermal model. Sometimes the SPICE based model and the finite element models are combined in a mixed mode simulation where most of the components in the circuit derive their characteristics from the SPICE parameters and one or more components derive their operating characteristics from a finite element simulation.

For this Chapter, only mixed mode or pure finite elements simulations were carried

out. In the mixed mode simulations only simple components were used such as ideal resistors, capacitors and switches as these components can easily be accurately simulated under the high current conditions that occur during ESD events.

5.3.1 Finite Element Electrothermal Simulation

Finite element simulation is commonly used in many disciplines of engineering because it involves using a computer to break-up a complex problem into discrete parts which can easily be solved to solve a larger problem. Examples of commercial finite element semiconductor simulators include Sequoia Device Designer supplied by Sequoia [141] and Medici and Taurus supplied by Synopsys [142]. These operate by dividing the component (resistor, diode, transistor etc...) into a large number of small regions. Each region will have its own semiconductor properties (e.g. doping density, band gap, electron and hole carrier density, electric field etc...). The software carries out calculations on each cell based on the properties of the cell (e.g. carrier density which will affect the rate of recombination) and the interactions with neighbouring cells (e.g. differences in charge density giving rise to electric fields or inflows/outflows of carriers changing the carrier density).

To start the simulation the software has to solve Maxwell's Equations so that the carrier densities in all parts of the silicon model are consistent with local current densities, doping densities and the voltages at the input nodes of the silicon (such as the drain/source/bulk/gate contacts). To continue the simulation the current is allowed to flow for a short period of (simulation) time whilst the potential of the input nodes are changed (such as increasing the gate voltage if we wanted to simulate what happens when the transistor is turned on). The simulator then has to recalculate the local carrier densities using Maxwell's Equations.

Maxwell's Equations create a circular argument where the carrier density affects

charge density, which determines the electric field, which causes (by drift) current flow, which in turn changes the carrier density (Figure 5.1). The carrier density also affects the current flow by diffusion. The very nature of discretising these calculations in both time and space means that there will be errors in them. These can be reduced by decreasing the time step and/or the individual region size at the expense of computing resources. If the time steps or regions are too large, the software may fail to find a solution within the specified error bounds (arbitrarily defined by the input file). If the regions or time steps are too small, the computer may run out of memory and/or take an excessive amount of computational effort.

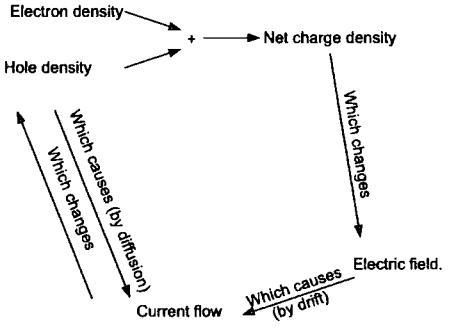

Figure 5.1: The calculations that the Finite Element Simulation software has to satisfy in order for the simulation to converge

The following Equations are the Maxwell's Equations used in the simulation software. The net charge density determines the electric field using Maxwell's Equation:

$$\nabla \cdot D = \nabla \cdot (\epsilon_s \varepsilon) = \rho(x, y, z) \tag{5.1}$$

In summary, this Equation says that the difference in effective electric field ($D = \epsilon_s \varepsilon$) between two points, is equal to the charge density difference between these two points.

The current is determined by two components: *drift current* which is set by the carrier density and the electric field; and *diffusion* which is set by the gradient of the carrier density. This is described in Equation 5.2.

$$J_n = q\mu_n.n.\varepsilon + q.D_n.\nabla n \tag{5.2}$$

$$ElectronCurrentDensity = ElectronDrift + ElectronDiffusion \tag{5.3}$$

Where q=electron charge, μ_n is the electron mobility, n is the electron carrier density, ε is the electric field strength and D_n is the electron diffusivity.

$$J_p = q\mu_p.p.\varepsilon - q.D_p.\nabla p \tag{5.4}$$

$$HoleCurrentDensity = HoleDrift + HoleDiffusion \tag{5.5}$$

The local charge density is a function of current density and of generation and recombination rates of the carriers.

$$\frac{\delta n}{\delta t} = G_n - U_n + \frac{1}{q}\nabla \cdot J_n \tag{5.6}$$

$$\frac{\delta ElectronDensity}{\delta t} = G_n - U_n + \frac{RateOfChangeOfJ_nOverDistance}{ElectronCharge} \tag{5.7}$$

Where G_n is the *rate of electron generation*, U_n is the *rate of electron recombination* and J_n is the *electron current density.*

$$\frac{\delta p}{\delta t} = G_p - U_p - \frac{1}{q}\nabla \cdot J_p \tag{5.8}$$

$$\frac{\delta HoleDensity}{\delta t} = G_p - U_p + \frac{RateOfChangeOfJ_pOverDistance}{ElectronCharge} \tag{5.9}$$

Where G_p is the *rate of hole generation*, U_p is the *rate of hole recombination* and J_p is the *hole current density.*

The previous explanation covers the top level of the simulation; in reality the simulation is considerably more complicated. For example, carrier drift reaches a maximum velocity (saturation velocity) at high electric fields. Details of all the models available in Taurus finite element simulator are described in detail in the Taurus manual [143]. Some of the models are mandatory, others are optional. Most of the models have input parameters such as electron mass, saturation velocity and bandgap temperature coefficient which can be set by the user as required.

The properties of a region are determined by the properties of the given region (e.g. recombination will be calculated from n and p carrier density) and the neighboring region (e.g. the electric field will be calculated from the difference in charge density). There will be a discrete difference in properties from region to region. In the simulation the properties for a region represent an average while, in reality, the properties would change in a continuous manner while moving through the semiconductor. Ideally, you would have lots of small regions so that the difference between them would be as small as possible. This would mean that the calculations are more likely to converge, but having a large number of regions has the disadvantage of requiring a lot of computer resources (CPU time and RAM). The software has to go through a number of iterations to converge to a solution where the error in Equations 5.1, 5.2, 5.4, 5.6 and 5.8 is below an arbitrary threshold. If there are few large regions, it will be difficult to make the simulation converge which will increase the simulation time. Alternatively in the worst case it will be impossible to achieve a low error rate and the calculation will fail. If there are lots of small regions the

simulator will find it easier to identify a route to convergence but will require more computing resources as there are more cells to converge. In the worst case there may be insufficient RAM to complete the calculations.

In practice, the regions are chosen to be small where there is a large change in parameters over a short distance (e.g. the depletion region in a PN junction), and large where the parameters are reasonably constant. For example in a MOSFET the source-bulk and drain-bulk pn junctions and the active regions of the transistor are divided up into small cells (Figure 5.2) as these regions have a high electric field, high minority carrier concentrations and large changes in carrier concentrations. All other areas of the transistor have a large cell size as the simulation parameters are fairly constant.

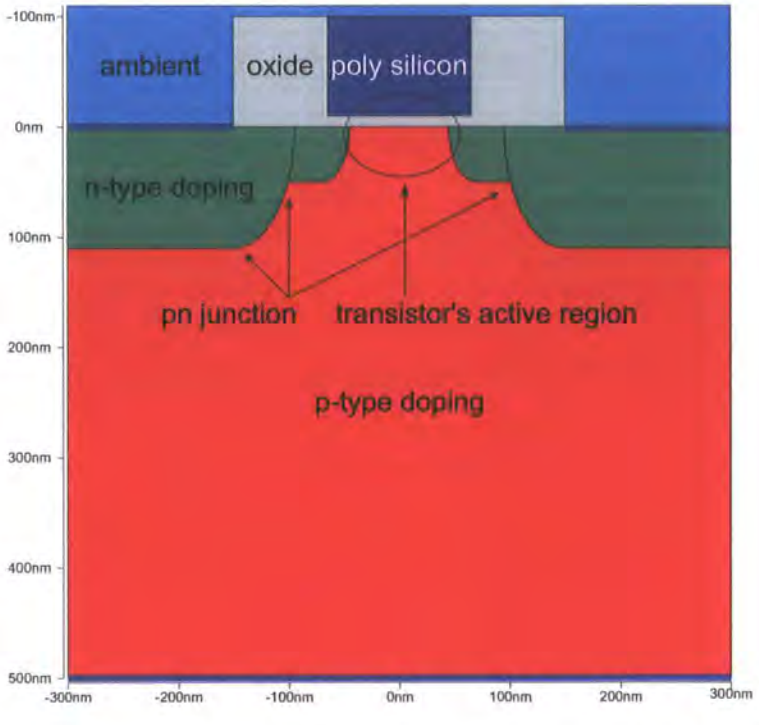

Figure 5.2: The parts of the transistor that require High Grid Density for Finite Element Simulation

In both simulators it is possible to have variable size regions (e.g the regions below the gate would be much smaller than the regions below the source and drain contacts). Both simulators require the user to specify the region size in any given

point of the silicon. The Taurus simulator [142] also provides some tools that allow automatic cell size refinement based on the variation in simulation or material parameters (e.g. electric potential, charge density or doping density).

The final component that can be added to the calculation are the thermal properties: the temperature of each region, the inter-region heat flow and heat generation within a region. Temperature plays a significant part in most of the formulas and any movement of charge will give rise to IR heating.

There are many examples in the literature where electrothermal simulation has been used, though they give little detail on how do to it. The consensus is that it is difficult to achieve accurate simulations [86] as the evolution of semiconductor technologies is moving the goalposts [144] though a few authors claim to have managed it. Esmark et al. [25] investigated whether 2D or 3D simulation was required for a ggNMOS. It was found that 3D gave the most accurate answer and 2D simulation gave optimistic results as it assumed a constant current density across the whole width of the protection structure.

5.4 Presentation of the Tools

In this work, simulation software from two suppliers was used: *Taurus* from *Synopsys* [142] and *Sequoia Device Designer* from *Sequoia* [141]. Both packages operate along similar principles, the only major difference between the two is that Taurus is able to simulate 2D and 3D structures whereas *Sequoia Device Designer* is limited to 2D. These two tools will now be briefly described before the investigations are discussed.

5.4.1 Taurus

Taurus is a command line and script file simulator that runs on UNIX/Linux which is capable of finite element structure definition and electrothermal simulation of both 2D and 3D structures. Other programs supplied by *Synopsys* are used to view the currents and voltages of each port during the simulation, *Synopsys* also provides a graphical viewer where the properties of any part of the 2D or 3D simulation mesh (such as the current density or direction) can be viewed. Detailed program operation of *Taurus* is described in detail in Appendix:A.

5.4.2 Sequoia

The *Sequioa* package has a Graphical User Interface (GUI) defined by TCL scripts and the *Sequoia* software. When the program is run, it presents the user with the GUI shown in Figure 5.3. Within this spreadsheet individual experiments are listed alongside basic details of the experiment.

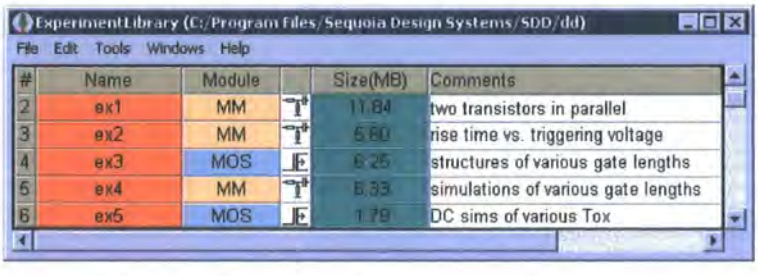

Figure 5.3: The main menu on the Sequoia program where all the experiments are catalogued

Double clicking on an individual experiment loads the TCL code within a module and presents the user with a spreadsheet (Figure 5.4). The rows list individual experiments and the columns contain the variables governing each individual experiment. *Sequoia* have written a number of different TCL files which allow the user to simulate different structures including diodes, MOSFETs, bipolar transistors, poly resistors and bulk resistors. The GUI (Figure 5.4) allows the user to vary some of

the physical parameters such as the MOSFET's source/drain doping density or oxide thickness, simulation parameters such as the grid density, as well as sweep or fix some of the electrical parameters such as V_{DS} or V_{GS}. A sample of these parameters is shown in Table 5.1.

Figure 5.4: The GUI that is used to set up each individual experiment

When using a TCL file designed to simulate MOSFETs, most common MOSFET parameters can be altered but, if the user wishes to simulate alternative structures such as a cascoded transistor or other arbitrary structures, then the TCL file has to be edited manually. This procedure is discouraged by *Sequoia*. To assist the simulator in the creation of the simulation grid the user has to specify the finite element grid density at all locations (in contrast *Taurus* allows the user to instruct the simulator to refine the grid in locations with high electric fields, doping gradients, carrier gradients etc...). In this respect, the *Sequoia* software is less efficient than *Taurus* as it is difficult to refine the mesh in the regions where it is required without causing adjacent regions to have a grid which is unnecessarily fine. This slows down the simulation and increases the memory requirements. The other disadvantage is that manual mesh refinement means that it is possible to overlook accidentally areas where mesh refinement is required.

There are two stages to simulations. First the transistor structure is created (Figure 5.5) and stored in a file, then voltage sources are applied and the electrothermal simulation begins.

Sequoia have also written a TCL file that allows the user to SPICE simulate one or

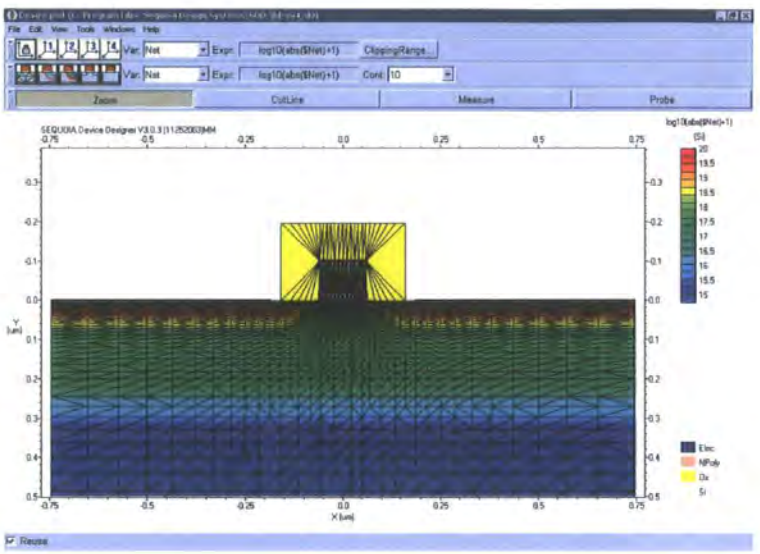

Figure 5.5: A typical MOSFET structure with the finite element grid overlaid

Parameter	Description
Lpoly	Length of the gate electrode (μm)
Tox	Oxide thickness (Å)
Wsp	Spacer width (μm)
W	Device width (μm)
Ns	Substrate doping (cm^{-3})
LDD	Peak LDD doping (cm^{-3})
LDDYch	LDD range (μm)
LDDXY	LDD ratio of x to y diffusion ranges
DyMin	Vertical mesh spacing at the Si-Ox interface (μm)
Fy	Vertical mesh spacing ratio
NGox	Number of horizontal mesh lines in the gate oxide
Vgmax	Maximum Gate bias for IdVg measurement (V)
Vgstep	Gate bias step size for IdVg measurement (V)
IdVg_Vd	Drain bias for linear region IdVg measurement (V)

Table 5.1: A sample of the variables used to define MOSFET structure creation and simulation

more structures that have been generated by the other simulations. This allows the user to apply any voltage or current source via a capacitor, resistor or inductor to any pin on the device(s) thus enabling complete flexibility in the type of measurements that can be made.

Sequoia does provide a limited manual which only gives just enough information to get the system up and running. Unfortunately, it does not provide any information on how the transistor structure is created, simulated, on which models were used or on the input parameters for the semiconductor models. The only factors that the user can control in the finite element calculations are whether impact ionisation is taken into account and whether temperature effects from joule heating are factored into the simulation.

It is possible to simulate an arbitrary user-defined semiconductor structure, such as a diode, by editing the TCL scripts (in the *.itcl files) though this is discouraged by the *Sequoia.* A diode, cascoded transistor and other novel structures were simulated for the purposes of exploring the tools capabilities. The TCL files also provide further insight on how the standard transistor is created and simulated, but it does not provide any information on the finite element semiconductor simulation calculations except for whether to include impact ionization and/or joule heating into the calculations.

5.4.3 Sequoia Finite Element Structure Voltage Stimulus

In the *Sequoia* simulator, three types of applied voltage stimulus were used. Some simulations were done with a slow ramp, so that the DC operation of the transistors could be derived. As a consequence all capacitive effects are ignored. For other simulations reproduction of real world conditions was attempted by either applying a TLP pulse waveform or by generating (using an RC filter in SPICE) an HBM

pulse. The aim was to understand how the AC component of ESD affects the device performance. The simulation method also allows joule heating to be added.

5.4.3.1 TLP Emulation:

This type of voltage stimulus is designed to replicate TLP test conditions, as described in Chapter 2. A TLP voltage source ($100ns$ long square pulse as shown in Figure 5.6) is applied across the finite element structure and $16K\Omega$ resistor in series (Figure 5.7). The rationale behind this resistance is that a TLP test is usually carried out on a complete ggNMOS clamp which, for CSR's $0.18\mu m$ chips, was $320\mu m$ wide. The simulator assumes that the 2D transistor being simulated is $1\mu m$ wide. Therefore to ensure good correlation the resistor needs to be 320 times greater than the impedance of the TLP tester (50Ω). The rising and falling edges are approximated using a SINE wave as they are assumed to have a gaussian profile. Full details on this stimulus is given in Appendix:B.

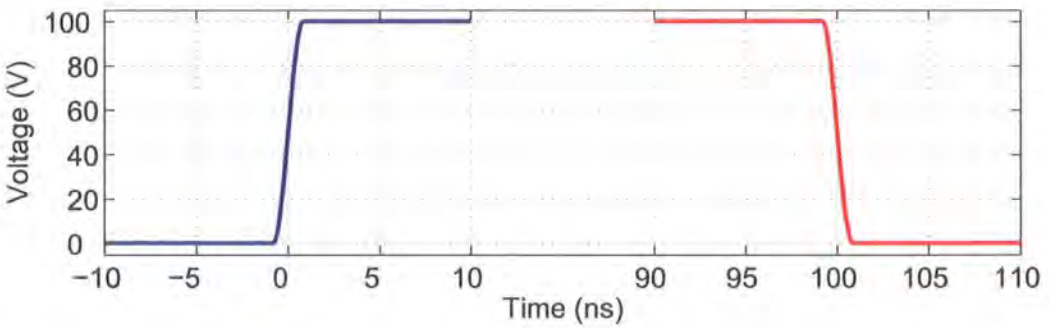

Figure 5.6: The voltage characteristics of a TLP waveform with t_R and t_F of $1ns$ and a duration of $100ns$

5.4.3.2 Ramp:

A slow voltage ramp, as described in Equation 5.10, can be applied to a structure with a $16K\Omega$ resistor in series to determine the DC IV characteristics on the finite element structure. This has the advantage of producing an IV curve directly from

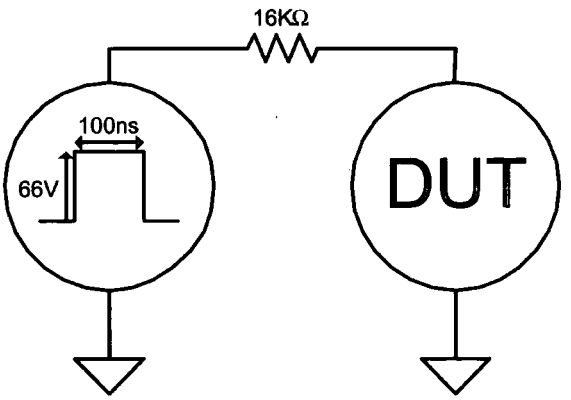

Figure 5.7: Graphical illustration of the application of a TLP pulse in a simulation

the simulation whilst at the same time, given that the rise time of the applied voltage is much slower than that of an HBM or TLP pulse, any capacitive effects are significantly reduced. A disadvantage is that self heating effects cannot be modelled as the applied voltage is not the same, in waveform shape or duration, as that of a TLP or HBM pulse. In practice, it was found that as the rise time was decreased the simulation time increased dramatically. If the rise time was too fast the electrical performance of the transistor would start to change, though the simulation times would decrease significantly. It was found that for rise times in the order of $10^8 V/s$ the simulation time was around 1 hour on an standard desktop 3GHz PC and the electrical results were largely identical to those of a simulation running with a rise time of $10^7 V/s$.

$$V = 10^8 * t \tag{5.10}$$

5.4.3.3 HBM Emulation:

The final type of voltage stimulus presented here is designed to replicate HBM test conditions (as described in section 1.4.3). The HBM test has been defined by JEDEC [6] as a charged 100pF capacitor which is discharged through a 1500Ω resistor and through the DUT to ground. In real world tests, the transistor would

typically be $320\mu m$ wide and as the simulator assumes that the transistor is $1\mu m$ wide, the resistance needs to be increased to $480K\Omega$ (=$320 * 1500\Omega$) and the capacitance needs to be reduced to $312.5fF$ (=$100pF/320$). This can easily be replicated using the SPICE circuit shown in Figure 5.8. The voltage source is then programmed to produce a gaussian rising step, approximated using a SINE wave, with a rise time of $6ns$.

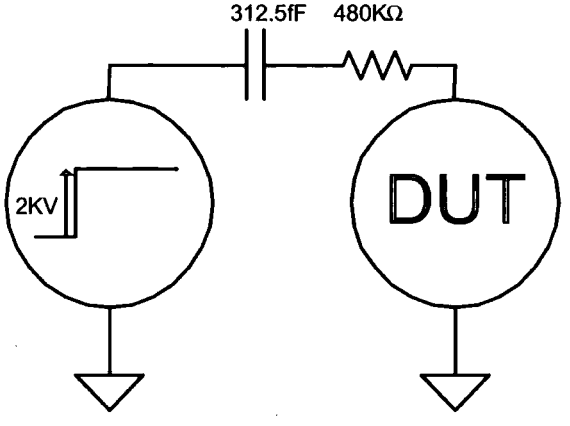

Figure 5.8: Graphical illustration of the application of an HBM pulse in a simulation

5.4.4 Choice of Simulation Program

Both tools were used to simulate a ggNMOS in snapback whilst varying a wide range of parameters. The Taurus simulator's documentation was found to be superior to the Sequoia's, particularly at the device physics level. It also allowed greater flexibility in varying any model parameter and adding or substituting models whereas the Sequoia simulator was very rigid in this respect. On the other hand if you just want to set up a basic simulation, Sequoia's GUI makes things considerably simpler than running a Taurus simulation using a script file. Despite the Taurus simulator being more flexible and capable of 2D/3D simulations, the Sequoia simulator (which is limited to 2D simulations) was used for most experiments as it seemed to have less difficulty converging. Convergence was a problem for both simulators; even the smallest change in a seemingly trivial parameter caused convergence problems. For

our experimental use, the interface and operation of both programs placed unnecessary constraints on how they were used. This often meant that it was not possible to understand why a simulation failed to converge and it also limited the parameters that could be altered to assist convergence. Further on in this Chapter, an example is given where the simulator misses out on the peak voltage due to the way in which the simulator is implemented. There is no fundamental mathematical reason why the simulator should fail to produce this result accurately.

5.5 Creating the Structures

5.5.1 Introduction

The ultimate aim in this Chapter is to use finite element simulation to understand the snapback mechanism as this will assist design improved devices. Of particular interest is how ggNMOS transistors in CSR's $0.18\mu m$ chips would behave under ESD conditions since this was the current process at the time. Due to commercial sensitivities, only limited data on the doping profiles for the process used to manufacture the CSR's $0.18\mu m$ chip was available. The gaps in the recipe were filled in using information provided by Wolf's book: *Silicon Processing for the VLSI Era*[18] and using the example files available with the simulators. The doping profile used is shown in Figure 5.9. The author does not believe that it is possible to achieve simulation results that are close to measured values given that, even if the correct doping data were available, it has been shown many times in the literature [145][146] that it is still very hard to achieve good correlation between simulation and measurements. Yet the author does believe that it is still possible to gain a greater understanding of the mechanisms involved in snapback and determine how variations in device parameters affect ggNMOS operation.

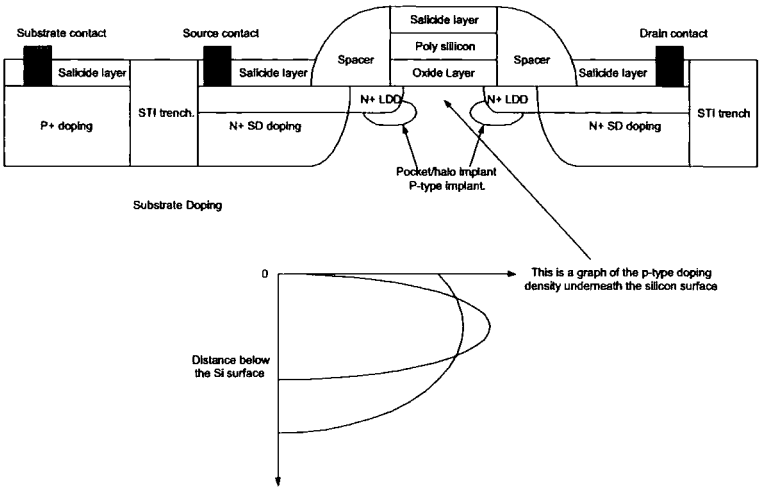

Figure 5.9: The structure of an nMOS transistor

A number of simplifications have been made which do not affect the simulation fidelity.

- The substrate contact is placed at the bottom of the cell so that all the lowest nodes in the mesh are at the substrate potential. This allows a reduction in the simulation grid size without significantly altering the resistance or distance between the transistor and the substrate tap.
- The STI trenches are not included in the mesh as there is no significant electrical field or current that passes through them.
- The oxide and the spacers are drawn as a single 'U' shaped oxide (Figure 5.10).
- The silicide layer is not included in the finite element mesh. Instead, the nodes between the silicon edge and the spacers are assigned to the source and drain electrodes (Figure 5.10). Given that the silicide layer is a tungsten implant designed to reduce significantly the surface resistivity of the silicon, replacing the silicide depositions with contacts will have a very similar effect.
- The poly-silicon of the gate is omitted and replaced with a contact along the top surface of the gate oxide. In certain cases, where V_{GS} is high, a

depletion region can be formed in the poly silicon causing the gate capacitance to decrease and the poly resistance to increase. Given that the V_{GS} of a ggNMOS, by definition, is $0V$ this effect can be ignored.

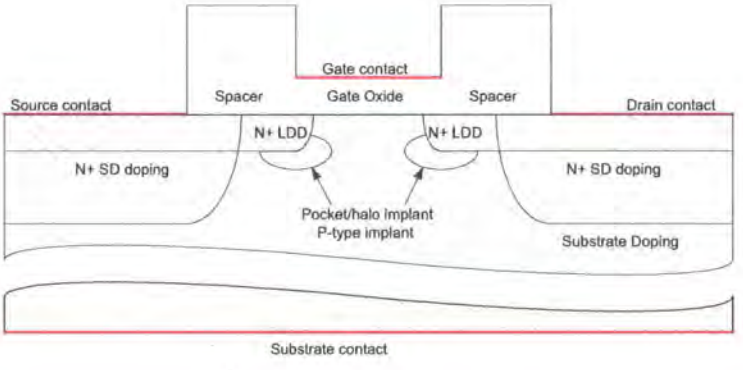

Figure 5.10: The structure of an nMOS transistor as defined by the Finite Element Simulation

p-type doping is added just under the silicon surface (Figure 5.9) to prevent punch-through; it achieves this by decreasing the depletion width of the source and drain pn junctions. It is well known that increasing the doping density decreases the drift velocity [5]. Therefore the dopant is implanted in such a way that there is a high doping density in the region as a whole yet on the silicon surface, where most of the current flows during normal transistor operation, the doping density is lower. This is achieved by ion implantation so that the peak doping density of $1.5 * 10^{17} ions/cm^3$ is below the silicon surface and the surface doping density is around $2 * 10^{16} ions/cm^3$. All the doping densities and doping depths are shown in Table 5.2.

Item	Typ	Units
T_{ox}	$32 - 80$	Å
Substrate Doping	1×10^{12}	$ions/cm^3$
LDD doping	1×10^{19}	$ions/cm^3$
LDD depth	60	nm
SD doping	1×10^{20}	$ions/cm^3$
SD depth	160	nm
Halo implant	1×10^{18}	$ions/cm^3$
Halo depth	60	nm
P-type1 implant	1.5×10^{17}	$ions/cm^3$
P-type1 depth	80	nm

Table 5.2: The doping profiles used for the Sequoia simulations

5.6 Simulating the Structures in the Linear (non ESD) Mode of Operation - Basic Model Verification

The transistor structures were simulated using Sequoia Device Designer in their standard operating modes (i.e. not as ESD protection devices) and the results were compared to generally accepted theory. If theory and the simulation do not match then the discrepancy should be resolved or explained before ESD simulations are started. The results of the simulation were in line with expectations [5] and are shown in Figures 5.11-5.13. Figure 5.11 shows that when $V_{DS} << V_{GS}$, the transistor is in the linear mode and behaves as a resistor whereas when $V_{DS} >> V_{GS}$, the transistor is saturated and behaves as a current source. It is clear though that the drain impedance in the saturated mode is not infinite ($\Delta I_{DS} \neq 0$) and this is due to short channel effects. Figure 5.12 shows V_{GS} vs I_{DS} and it can be seen that V_T is around 0.4V. The graph appears to show a form of saturation; this is due to the series resistance of the transistors which is about $2.5K\Omega$. Figure 5.13 shows V_{GS} vs I_{DS} over Temperature. As expected, the graph shows that as temperature increases, V_T decreases and the leakage current increases. This is due to the fact that the temperature increase causes the semiconductor material to become more

intrinsic. As a result the gate voltage required to turn on the transistor (V_T) reduces and minority carrier density increases which causes the leakage current to increase. The temperature increase also reduces the carrier mobility which therefore results in a decrease I_{DS} and G_M.

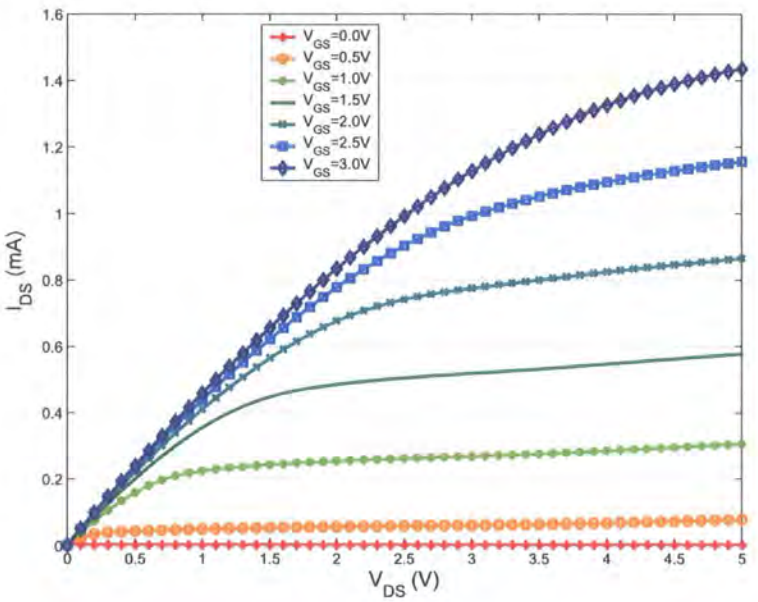

Figure 5.11: V_{DS} vs I_{DS} over a range of V_{GS}

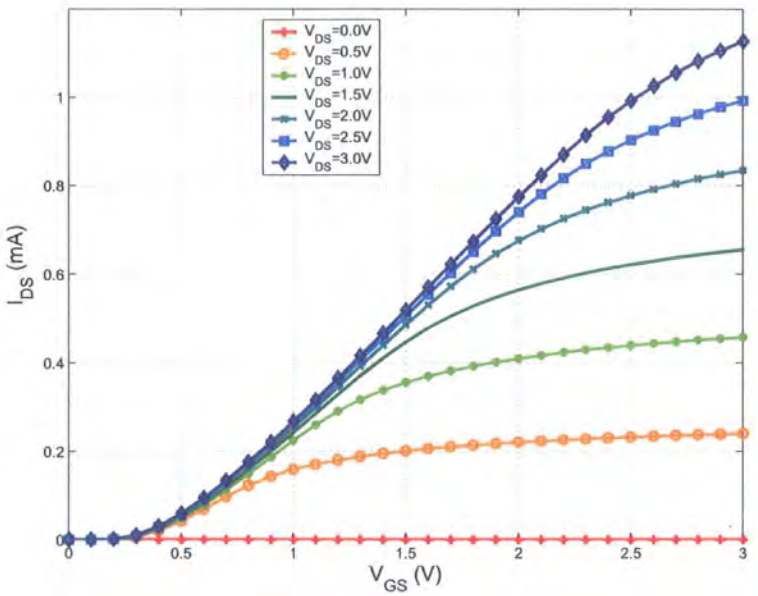

Figure 5.12: V_{GS} vs I_{DS} over a range of V_{DS}

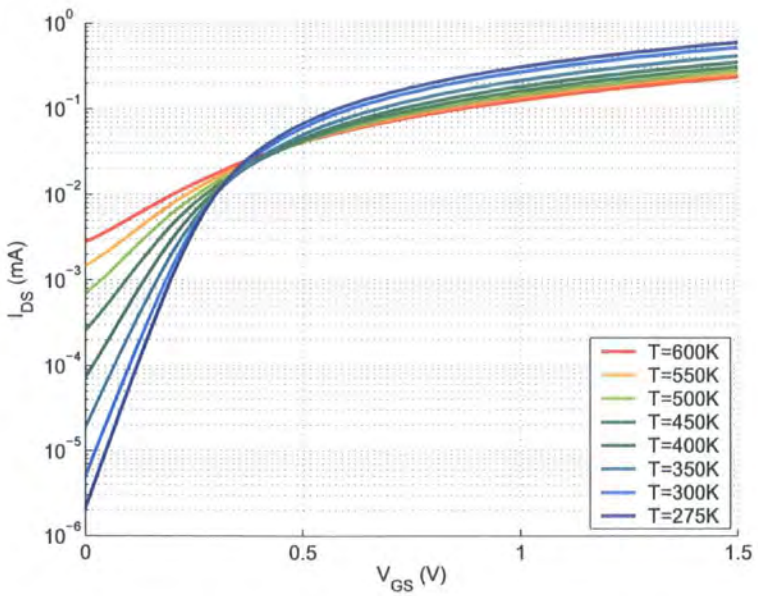

Figure 5.13: V_{GS} vs I_{DS} over Temperature

5.7 Simulating the Structures under ESD Conditions - Triggering during ESD

5.7.1 Introduction

The triggering voltage, V_{t1} (Figure 1.5), is an important factor in the device operation as it will affect the maximum voltage that will occur across the circuits that the ggNMOS is protecting. It will therefore give an indication as to whether it is an effective ESD clamp. The triggering mechanism is discussed in detail in section 1.2.1.4.

5.7.2 Variable L_{gate}

The ggNMOS as described in section 5.5, was simulated over a range of gate lengths in Taurus and Sequoia to determine how this would affect the triggering voltage and the hold voltage using the *ramp* stimulus described in section 5.4.3. The results presented in Figures 5.14 and 5.15 broadly show that as the gate length is increased, the trigger voltage and the hold voltage increase (see Figure 1.5 for a reminder of the definition of trigger and hold voltage). As explained in the literature review (Chapter 1), one of the components of the snapback triggering and holding mechanism is the parasitic bipolar transistor illustrated in Figure 5.16. Increasing the gate length decreases the β of the bipolar transistor which increases the base current required to ensure that I_{DS} is maintained. As the base current is supplied by the avalanche multiplication within the drain-substrate diode, the voltage across the diode will have to increase thus increasing V_{DS} both for triggering the ggNMOS and sustaining the snapback current. Figure 5.17 summarizes the data in Figure 5.14 and 5.15 and it shows that there is a trend for the trigger voltage to increase as the gate length is increased.

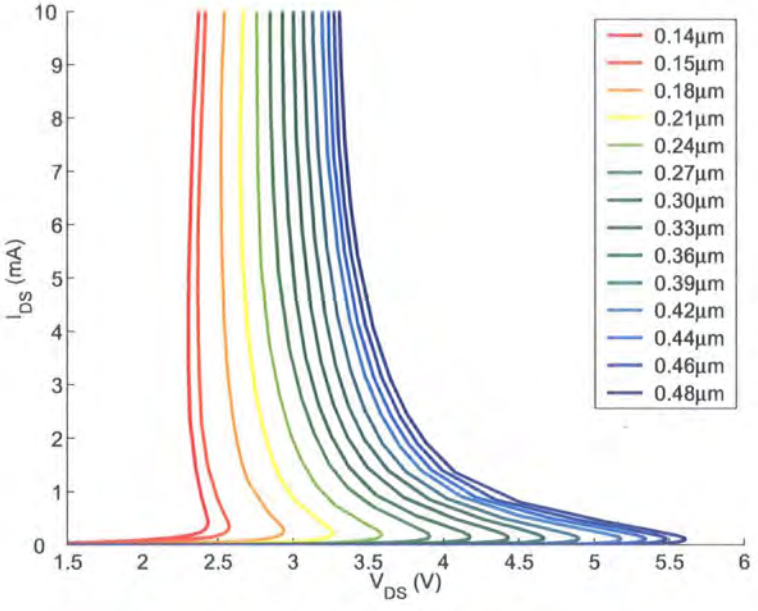

Figure 5.14: Taurus snapback simulation of V_{DS} vs I_{DS} over a range of L_{gate}

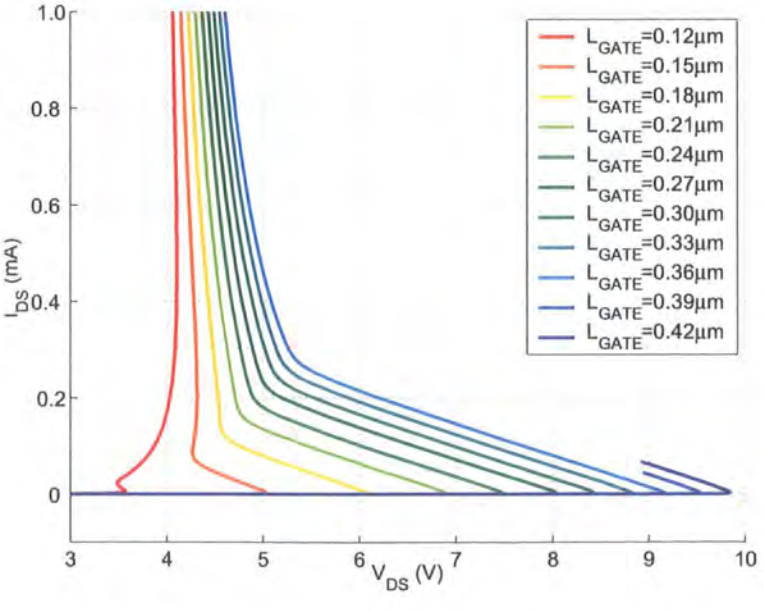

Figure 5.15: Sequoia simulation of V_{DS} vs I_{DS} over a range of L_{gate}

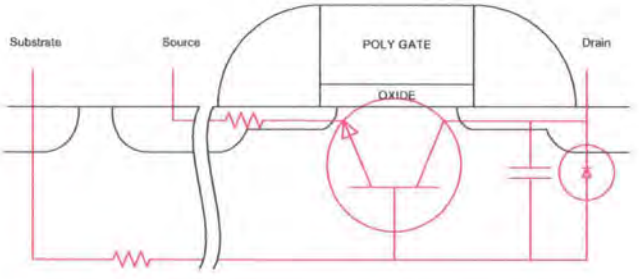

Figure 5.16: The parasitic components in a ggNMOS required to cause snapback

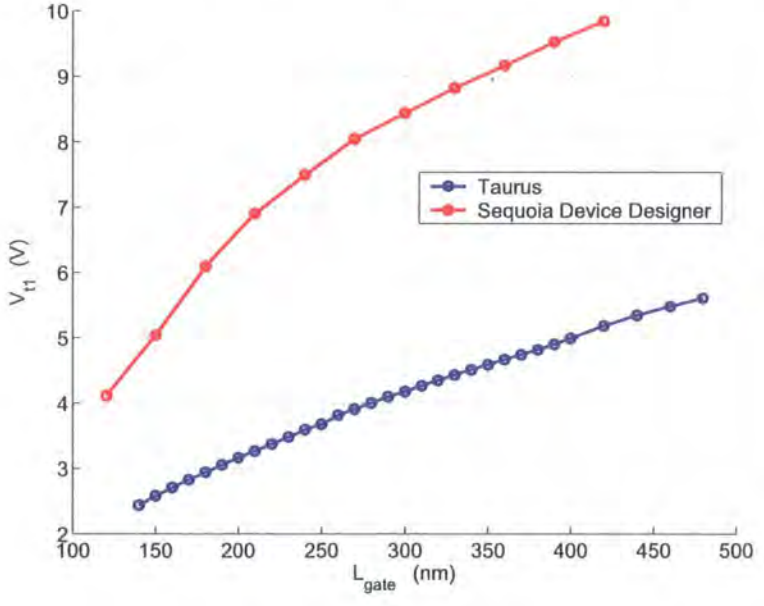

Figure 5.17: The V_{t1} vs L_{gate}

5.7.3 Variable Oxide Thickness

Oxide thicknesses of 32Å and 82Å are used for $1.8V$ and $3.6V$ transistors respectively, in the $0.18\mu m$ process being considered. There are also other differences in the doping profiles of both transistors though here the aim is to find out and understand the effect T_{ox} alone has on the ggNMOS performance. Using the standard ggNMOS 32Å doping profiles, transistors with a range of gate oxide thicknesses were simulated over the range 12Å$< T_{ox} <$ 72Å.

The most notable difference in the ESD performance of the ggNMOS transistors is the increase in trigger voltage as the oxide becomes thicker (Figure 5.18). To

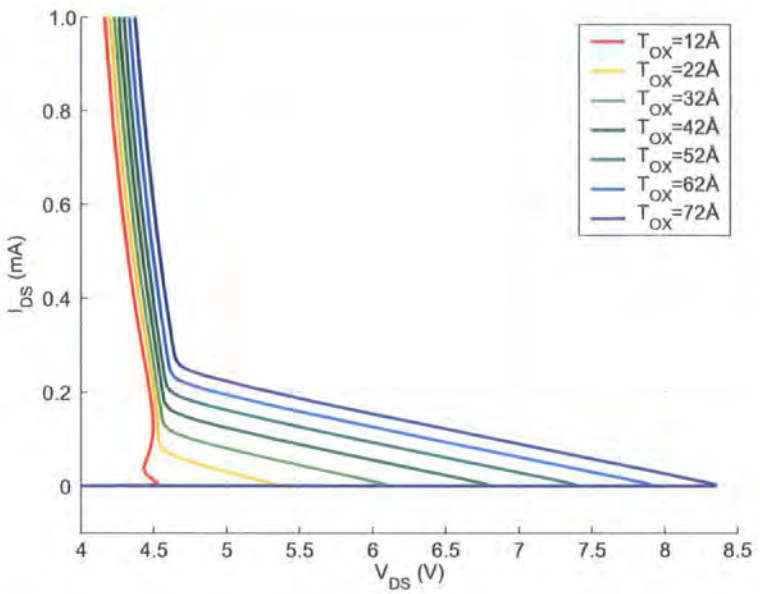

Figure 5.18: Sequoia simulation of V_{DS} vs I_{DS} over a range of T_{ox}

understand why this is happening, the triggering mechanism needs to be revisited. When $V_{DS} << V_{t1}$, the ggNMOS behaves as a reverse biased diode, as any current generation due to impact ionization or bipolar transistor current gain is negligible. The leakage current of a reverse biased diode is dependent on the minority carrier concentration on both sides of the diodes [5] so long as the reverse voltage across the diode is not too high or the temperature has changed. At the silicon surface, just under the gate oxide, the conduction and valence bands are bent due to the difference in doping levels of the p-type silicon substrate and the n-type poly silicon gate.

Band bending is due to the charge build up either side of the oxide which in turn induces an electric field across the gate oxide as per Maxwell's Equation (Equation 5.11 and Equation 5.12).

$$\nabla \cdot D = \nabla \cdot (\epsilon_s \varepsilon) = \rho(x, y, z) \tag{5.11}$$

$$\varepsilon = q/\epsilon_s \tag{5.12}$$

Since the potential difference between the polysilicon gate and the silicon substrate will remain constant, as T_{ox} is increased, the electric field will decrease thus reducing the charge build up either side of the gate oxide and therefore decreasing the band bending. The reduced band bending decreases the minority carrier concentration on the substrate side of the substrate-drain pn junction, thus decreasing the leakage current. A reduced leakage current means that a higher avalanche multiplication gain is required to provide the base current for the parasitic transistor which therefore means that the triggering voltage will increase to compensate.

Duvvury et al. [83] have used a similar effect to reduce the triggering voltage. In this case, the gate potential was increased to increase the band bending at the surface the silicon just underneath the gate. This methodology increases the minority carrier concentration leading to higher leakage current and therefore to a reduced triggering voltage.

5.7.4 Variable V_{GS} and Variable V_{BS}

In this series of simulations a standard $0.18\mu m$ was simulated over a range of gate voltages V_{GS} and bulk voltages V_{BS}. The IV traces are shown in Figures 5.19 and 5.20. They clearly show that as the bulk or gate voltage is increased, V_{t1} is decreased. It should be noted that the variation in V_{t1} is far greater for variations in V_{GS} (Figure 5.19) than for variations in V_{BS} (Figure 5.20). This is most likely because the depletion region within the silicon, underneath the gate oxide, is thicker than the gate oxide thickness; as a result the capacitance between the active region in the transistor and the bulk is lower than the gate oxide capacitance.

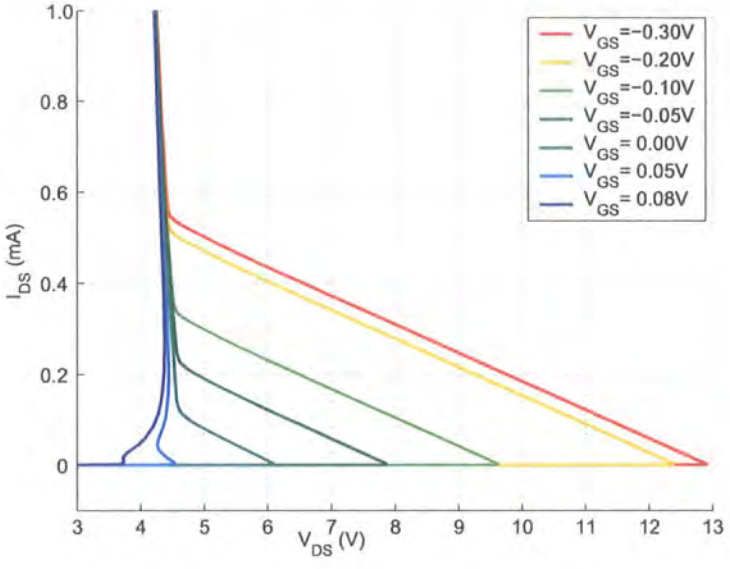

Figure 5.19: Sequoia simulation of V_{DS} vs I_{DS} over a range of V_{GS}

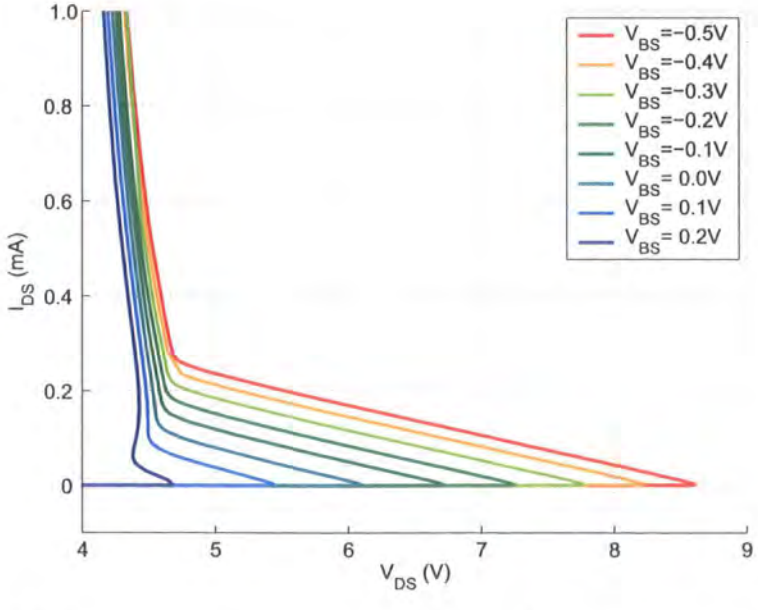

Figure 5.20: Sequoia simulation of V_{DS} vs I_{DS} over a range of V_{BS}

5.7.5 Comparison Between All Variable V_{t1} Methods and I_{leak} at $V_{DS} = 2V$

The previous subsections have shown that there is a correlation between the variation in L_{gate}, T_{ox}, V_{GS} and V_{BS} and the variation in V_{t1}. For this section the I_{DS} leakage was measured at $V_{DS} = 2V$ for each simulation set up (Figure 5.21 shows $I_{DS@2V}$ vs V_{t1}).

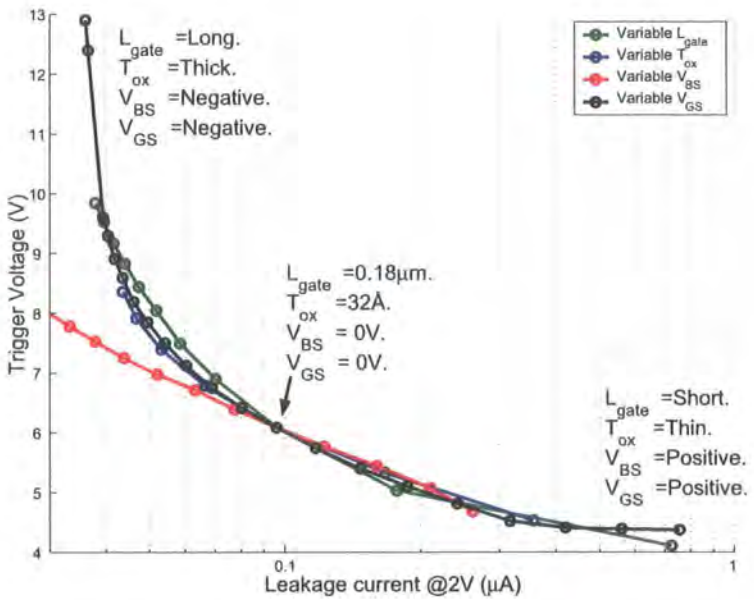

Figure 5.21: V_{t1} vs I_{leak} over a range of gate lengths, oxide thicknesses, gate biases and bulk biases

It is interesting to note that, regardless of the method used to vary the triggering voltage, the leakage current can be used to predict the triggering voltage. The experiments with negative bulk-source voltage do not follow the trend set by the other three series. For these experiments the trigger voltage is significantly lower for a given leakage current. The author is unable to explain why this phenomenon occurs and would have thought that for a given leakage current, the triggering voltage would be higher than the trend, not lower. The reasoning is that, as you decrease the bulk-source voltage, the drain leakage current decreases. As previously explained, this has the effect of increasing the triggering voltage. Furthermore,

decreasing the bulk voltage should also increase the (avalanche) current required to turn on the parasitic bipolar transistor, since the avalanche current is used to induce a current across the bulk resistor to forward bias the bulk-source diode thus turning on the bipolar transistor, thus increasing V_{t1}. In summary, decreasing the bulk voltage should increase V_{t1} though two mechanisms, (1) decreasing the leakage current and (2) making it harder to forward bias the bulk-source junction.

5.7.6 The Triggering Mechanism and its Relationship with the Rise Time

This section discusses the effect the ramp speed has on the triggering characteristics. For these simulations the voltage across the device will ramp as described in section 5.4.3. For each simulation the ramp speed is varied between 10^8 and 10^{11} V/s; V_{DS} vs Time is shown in Figure 5.22.

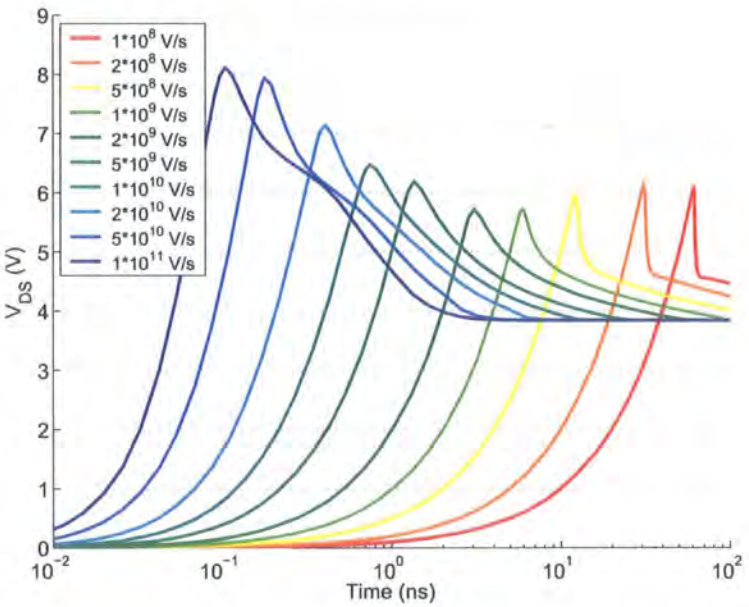

Figure 5.22: V_{DS} vs Time over a range of applied voltage ramps

Figure 5.23 shows I_{DS} vs time. As the ramp speed is increased, pre triggering I_{DS} increases, as indicated by the black arrow. This increased current is due to the

bulk-drain capacitance and has the effect of providing an alternative path through the bulk-drain diode to trigger the snapback mechanism. The voltage needed across the bulk-drain diode to trigger snapback is therefore reduced. Figure 5.22 shows that for slow ramps of 10^8 V/s, the triggering voltage is $6.08V$. As the ramp speed is increased to 10^9 V/s the triggering voltage falls to $5.71V$ due to the bulk-drain capacitance. For these experiments the triggering voltage is defined as the peak voltage across the transistor before there is any significant current flow. As the ramp speed is further increased the *apparent* triggering voltage then increases to $8.1V$ for ramp speeds of 10^{11} V/s. This may be due to the transistor's delay in snapback operation thus allowing the drain voltage to exceed significantly the snapback voltage before triggering occurs.

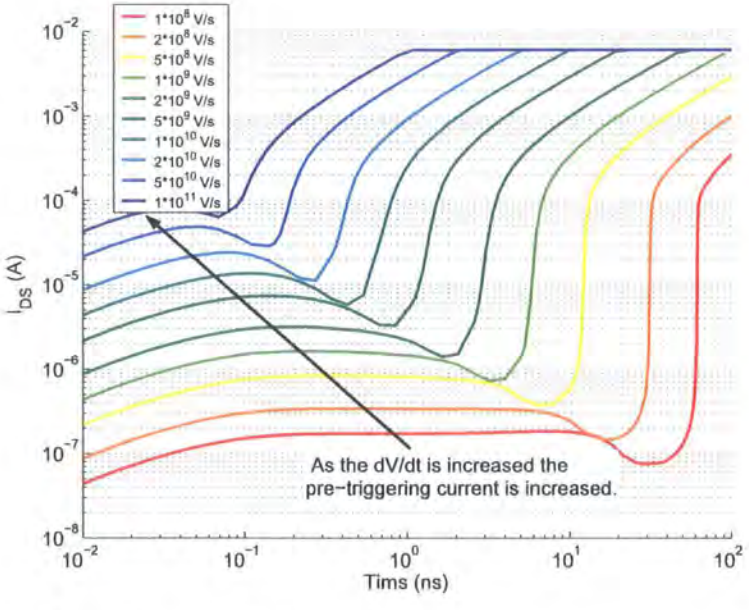

Figure 5.23: I_{DS} vs Time over a range of applied voltage ramps

To investigate further this phenomenon, the ramp voltage was limited to $4.25V$ (Figure 5.24) so that the high dV/dt of the ramp can trigger the transistor without causing a high V_{DS} to occur across the transistor due to the delay in snapback.

Figures 5.25 and 5.26 show V_{DS} vs time for a range of different ramp speeds. When the ramp speed is below $10^9 V/s$ the transistor does not trigger and the final voltage is $4.25V$. For ramps faster than 10^9, the final voltage is about $4.15V$ showing that

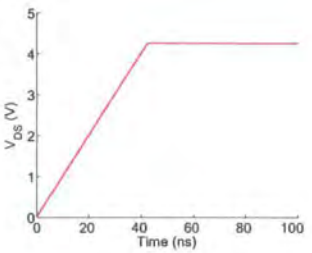

Figure 5.24: $10^8 V/s$ voltage ramp with a voltage limit of $4.25V$

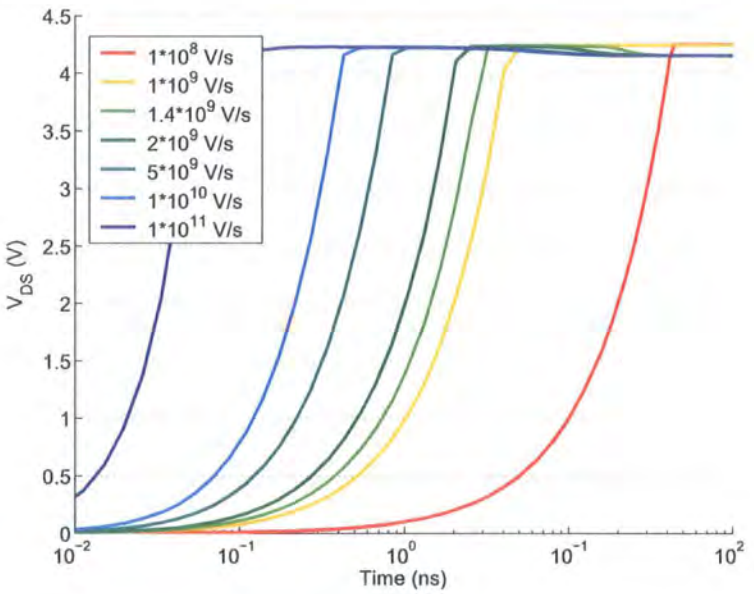

Figure 5.25: V_{DS} vs time over a range of voltage limited ramps

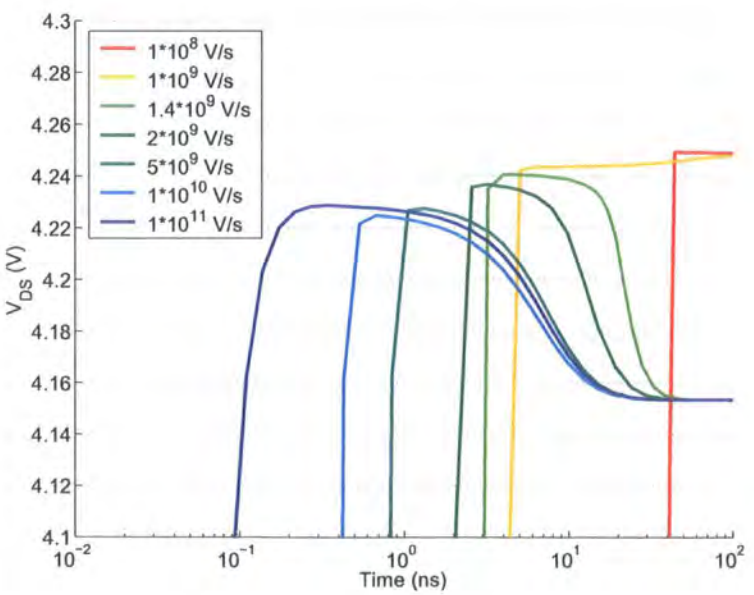

Figure 5.26: Detail from 5.25

the transistor is conducting and in snapback. In this instance, a ramp as slow as $1.4 * 10^9 V/s$ with a voltage limit of $4.25V$, is still able to force the transistor into snapback. On the other hand, the previous experiment showed that V_{DS} had reached $6.08V$ when the ramp speed was $10^8 V/s$. Therefore the rising edge of V_{DS} contributes significantly to the triggering mechanism. It is also interesting to note that for a ramp speed of $10^{11} V/s$ the transistor does undergo snapback yet the maximum V_{DS} is $4.25V$. This suggests that $4.25V$ a sufficiently high voltage to cause snapback and that any occurrence of $V_{DS} > 4.25V$ is due to the rapid rise in V_{DS} and the inability of the transistor to undergo snapback sufficiently quickly.

5.7.7 Investigating the Bipolar Effect

The advantage of simulations is that process parameters can be set to unconventional levels where the transistor may not operate correctly purely to study the mechanisms involved in the circuit. This experiment could be done in silicon but would be prohibitively expensive and time consuming to carry out on a commercial process.

In this particular experiment, the transistor was resimulated with a modified substrate doping of $10^{16} ions/cm^3$ to $10^{17} ions/cm^3$ (normally the substrate doping is $10^{12} ions/cm^3$). The effect of the increased substrate doping is to decrease the substrate resistance (Figure 5.16) which alters the IV characteristics of the ggNMOS (Figures 5.27 and 5.28). Figure 5.27 shows that, as the substrate doping is increased the hold voltage increases. This can be explained by looking at the circuit of the parasitic components in a ggNMOS (Figure 5.16). During snapback the base of the bipolar transistor needs to be kept at $\sim 0.7V$ by the holes provided by the avalanching diode. As the substrate resistance decreases the amount of current required to sustain the base voltage increases, thus requiring a higher drain voltage.

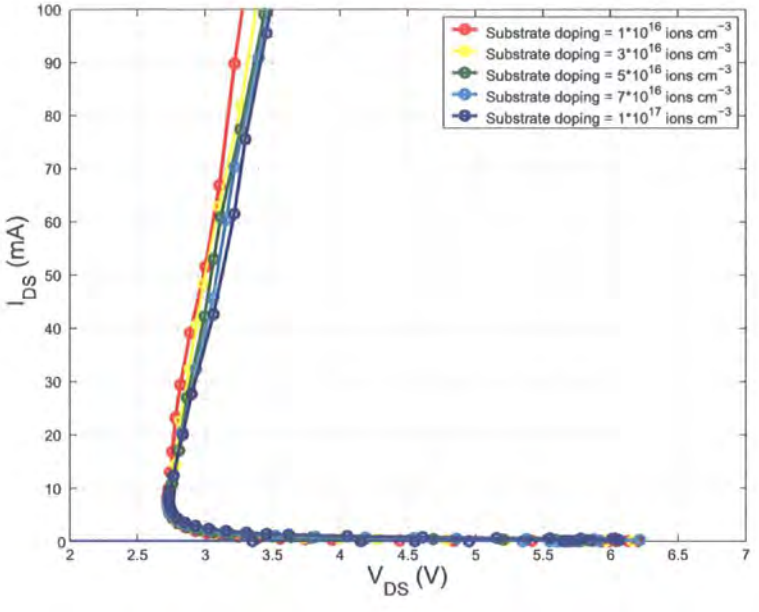

Figure 5.27: The ggNMOS IV profile for different substrate doping

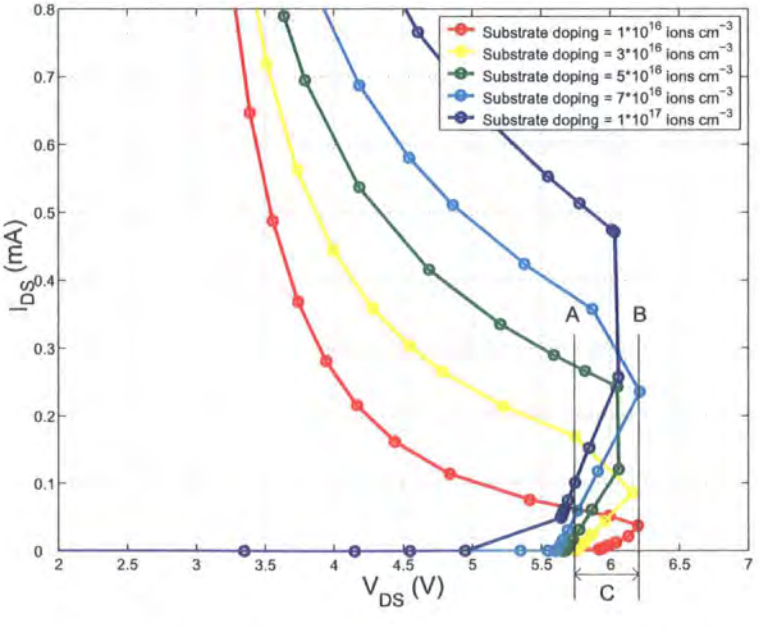

Figure 5.28: Triggering detail from Figure 5.27

Figure 5.28 shows the triggering voltage in detail: it can be seen that the current required to reach the triggering point for the substrate doping of $10^{16} ions/cm^3$ is around $37\mu A$ and as the doping level increases, the current required to trigger the device increases. This is due to the decreased resistance of the substrate which means that a higher current is required for the base of the bipolar transistor to be biased to $0.7V$. In addition to this, the actual shape of the IV curve demonstrates the operation of the parasitic transistor. Note that the transistors begin to conduct at around $5.6V$ (marked A on Figure 5.28) but snapback does not occur until the voltage has increased by a further $0.38V - 0.60V$ which is around the turn on voltage of the transistor (marked B). The explanation for this phenomenon is that the reverse biased drain substrate diode is starting avalanche breakdown in A but as the substrate resistance is low, the voltage at the base of the transistor remains low. As the impedance of the avalanching diode is low, any increase in V_{DS} will go to the base of the bipolar transistor. Once the voltage has increased by $0.4V - 0.6V$, snapback occurs (B on Figure 5.28). A similar IV curve was reported by Khazhinsky et al. [147] when they ESD tested devices with V_{GS} over a range $0V - 6V$ they found that transistor would start conducting a modest ($\sim 2mA/\mu m$) current before triggering if $V_{GS} > V_T$. If V_{GS}=$0V$ then the triggering current is very close to the leakage current. In this instance Khazhinsky explained that the increased V_{GS} suppressed the bulk voltage by supplying electrons from the source to the bulk thus decreasing the effective substrate resistance to ground. The author believes that Khazhinsky did not spot the final $0.7V$ increase in V_{DS} just before triggering where the rate of increase of I_{DS} increases sharply. The author also maintains that the sharp increase in I_{DS} is due to avalanche multiplication and the fact that it takes a further $0.7V$ increase in V_{DS} to trigger the device is evidence that the effective substrate resistance is low and therefore that it takes a higher current to trigger snapback in this transistor.

The previous explanation sheds further light on the method of operation of the

ggNMOS. It has a weakness though: the simulated voltage step between avalanching and V_{t1} (V_{ava-sb} - See on Figure 5.28:C), as the substrate doping is increased, is erratic (Figure 5.29). Given that increasing the substrate doping would decrease the substrate resistance, one would expect the triggering voltage and V_{ava-sb} to increase.

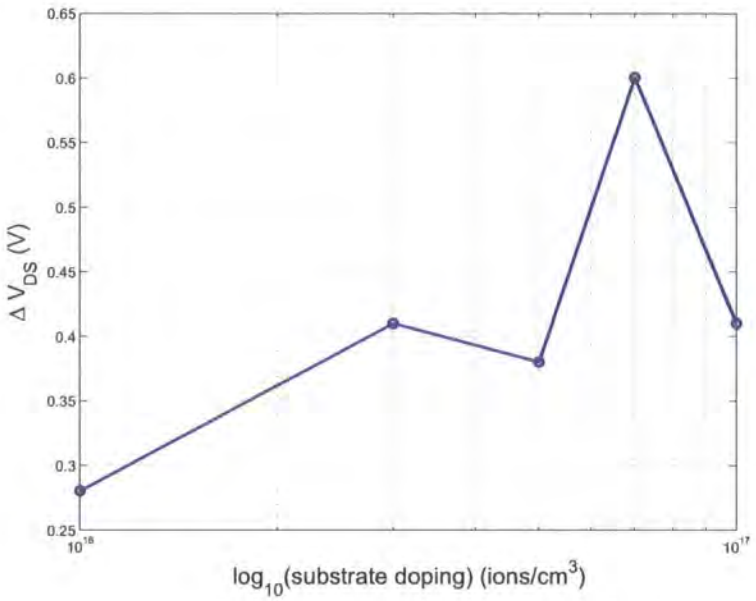

Figure 5.29: Doping density induced changes between the avalanching point and the triggering voltage

The author believes that the problem lies in the simulation step size (the ΔV and/or ΔI between individual simulations). This determines the likelihood that the difference between the peak measurement voltage and the actual peak voltage will be large. It is illustrated in Figure 5.30 where the red line shows the continuous IV trace for a ggNMOS and the other lines show the discrete measurements with the lines joined up. The green line gives a peak voltage which is near the continuous peak voltage but the blue line does not. It is a matter of chance whether a simulation follows that blue or green trace. If the step size is small then the difference between the peak measurement and the actual peak voltage is likely to be smaller (Figure 5.31).

It was not possible to control the step size in the Taurus simulations (and the level

control in the Sequoia software is limited) as the simulation step size is automatically calculated by the software. A useful addition to both simulators would be a feature allowing more control of the simulation step size so that these regions can be accurately traced.

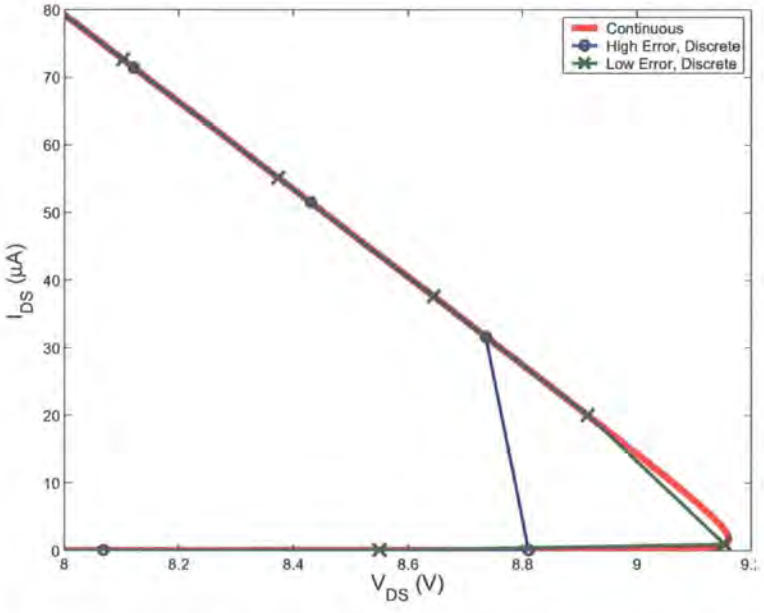

Figure 5.30: Error in peak voltage measurement caused by simulation step size

5.8 Simulating the Structures under ESD Conditions - Operation Once Triggered

5.8.1 Correlation Between an HBM and TLP Pulse

In this experiment, an HBM and TLP stimuli (as described in section 5.4.3) was applied to the transistors, in this instance joule heating was included in the calculation. The aim is to verify whether there is good correlation between the failure current for an HBM and TLP waveform. Good correlation is assumed to occur when the peak temperature for the TLP and HBM simulation is the same for waveforms with the same peak current. The simulation was set up on the basis that the ggNMOS

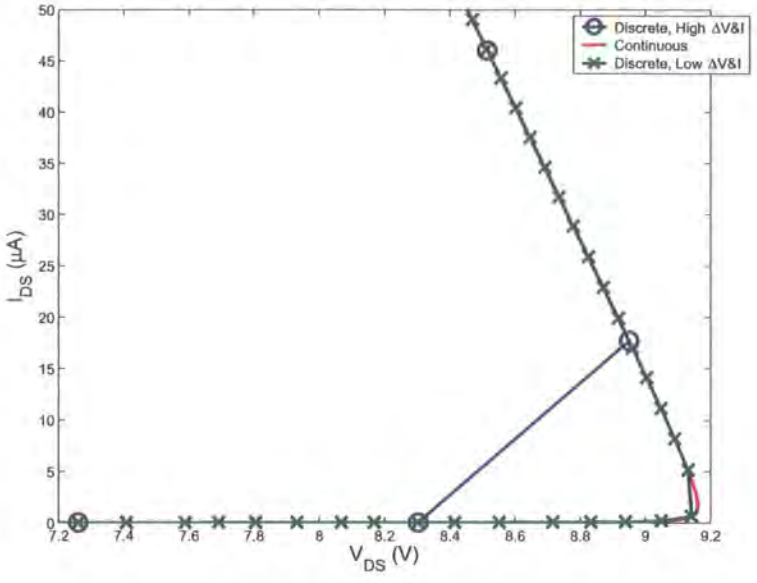

Figure 5.31: Improvement of the simulation accuracy due to the decrease in simulation step size

would be $320\mu m$ wide; therefore as the simulator assumes that the transistor is $1\mu m$ wide, the value of the resistors used in the TLP and HBM set up was increased by a factor of 320 and the value of the capacitor used in the HBM set up was decreased by a factor of 320. The circuits used in the simulation are shown in Figures 5.7 and 5.8.

Figure 5.32 plots the peak temperature of each simulation and shows that a simulation of a TLP stimulus predicts a much higher temperature than the simulation of an equivalent energy HBM simulation. This result clearly disagrees with the measurements described in section 3.3.4 and the papers reported in the literature review [123][126].

Before criticising the experiments that show that the maximum temperatures of TLP and HBM simulations do not correlate for a given ESD peak current, it is worth noting that correlation has been shown but does not always necessarily occur [123] [126][128]. It is not surprising that the temperatures reached in a TLP simulation are considerably higher than those of an HBM simulation given that the TLP pulse

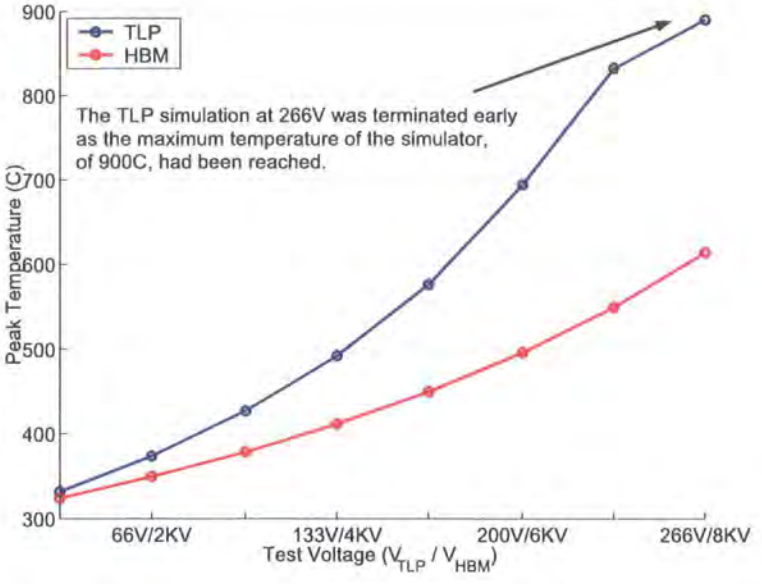

Figure 5.32: T_{MAX} for TLP and HBM simulations over a range of test energies

sustains the peak current for the duration of the test pulse ($100ns$) whereas the peak current found in an HBM pulse lasts a few ns and rapidly decays. It is important reiterate that the peak currents are the same in both tests.

The miscorrelation could be due to the possibility that these simulations do not reflect the actual conditions found within a ggNMOS during snapback; it may also be that peak temperature is a poor criterion by which failure should be defined. It is difficult to comment on the correlation in failure current between the simulations and the actual conditions found within a ggNMOS during snapback, given that it is difficult to measure the current density or temperature profiles that occur during an ESD event. The simulations results can be difficult to interpret as the designers of the simulation software have not provided full details on how the simulator works. It is for this reason that an open source semiconductor finite element simulator would be particularly beneficial.

5.8.2 Variable L_{drain}

It is well documented in the literature that an extended drain ensures that the ESD tolerance of a ggNMOS is proportional to the overall gate width of the ggNMOS transistor(s) [2][59][61][83][85][87][88]; this requirement is known as *width scaling.* To understand why an extended drain improves the performance of a ggNMOS, the reader needs to remember that a wide ggNMOS can be represented by a series of small transistors in parallel (Figure 5.33:1). During manufacture of the ggNMOS there will be process variations. As a consequence the triggering voltage of the devices will not be identical. Therefore once one device has triggered and reduced V_{DS}, it will prevent any other device from triggering. If the drain current is increased, the triggered device will fail due to excessive heat generation causing secondary breakdown. If the drain of the ggNMOS is extended, a resistor is effectively added to the drain of the ggNMOS (Figure 5.33:2). In this type of circuit when one of the ggNMOSs triggers, it will reduce V_{DS} across all the devices thus preventing them from triggering. As I_{DS} is increased, the resistor in series with the triggered device causes V_{DS} to increase. This causes all the other *small ggNMOSs in parallel* to trigger thus spreading the current out across all the devices and therefore maximising the efficiency of the device.

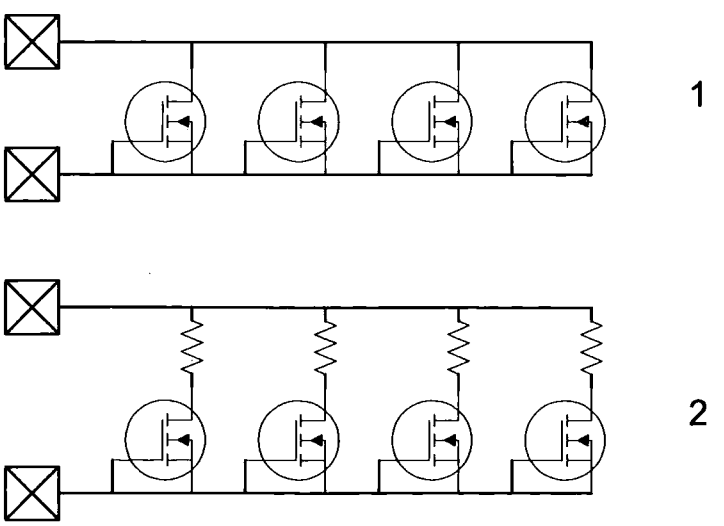

Figure 5.33: The effect an extended drain has on a ggNMOS

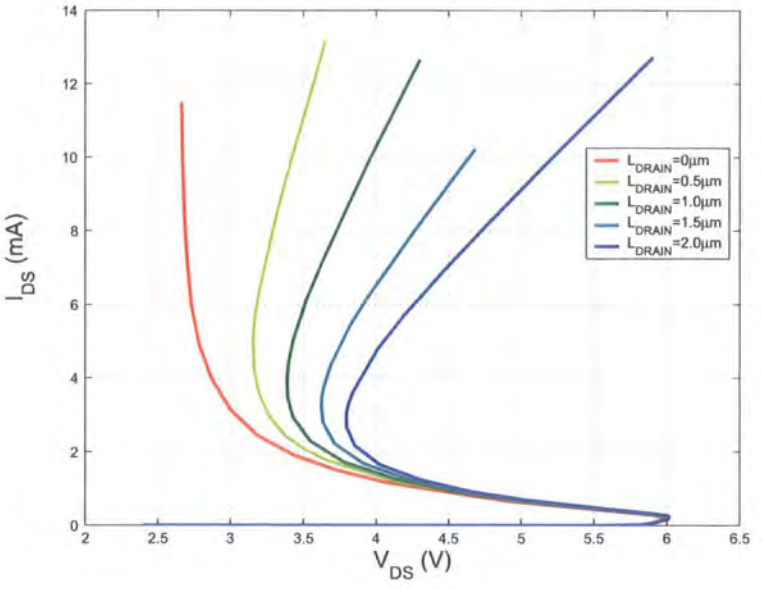

Figure 5.34: The IV characteristics of a ggNMOS for various lengths of extended drain

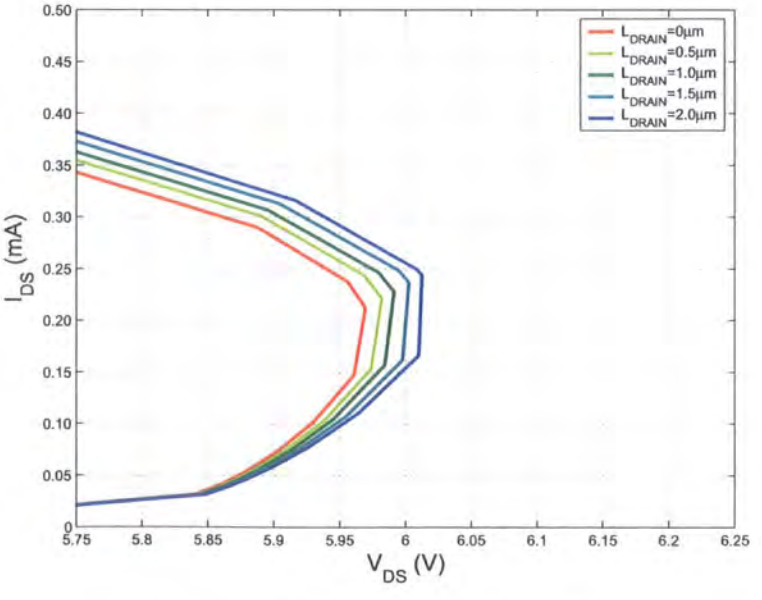

Figure 5.35: Detail from Figure 5.34

A ggNMOS with an extended drain was simulated using Taurus (Figure 5.34 and 5.35). Clearly, as the drain length increases, the voltage across the device increases, though one would not expect the triggering voltage to increase since the current at the triggering point is very low. To further our understanding, the voltage difference between each curve in Figure 5.34 was calculated: and the results are shown in Figure 5.36. The individual points are close to the linear best fit lines which would suggest that the extra drain resistance is ohmic and that velocity saturation or conduction modulation, which has been reported by Esmark et al. [25] and Liou at al. [148], is not occurring in this simulation. The slope of each best fit line is different and there does not appear to be any apparent trend in the slope of the best fit lines. It would seem obvious that each line should have the same slope, therefore this apparent anomaly in the simulation should be further investigated. This anomaly could be due to poor mesh refinement causing variation in the doping profile for a given location within the transistor when compared between transistors.

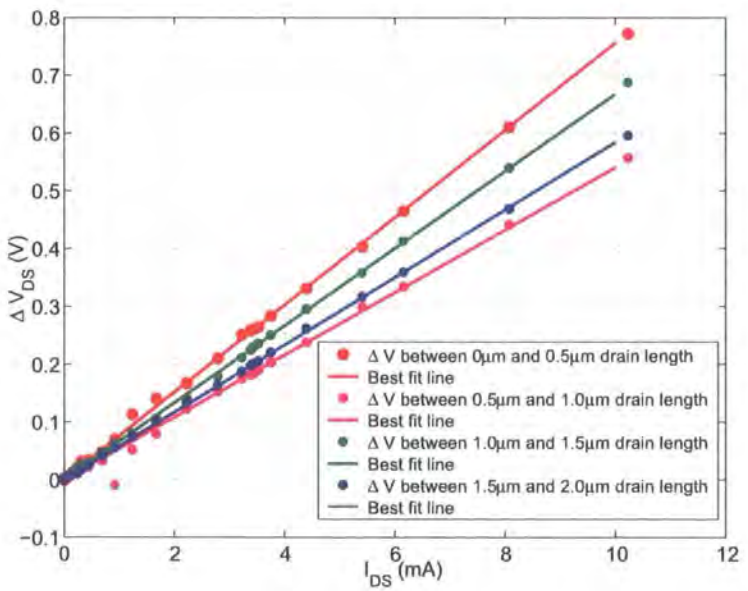

Figure 5.36: Change in V_{DS} for structures with different drain lengths

5.8.3 Investigating the Drain Resistance in Isolation

Accurate prediction of the effect that the extended drain will have on the overall operation of the ggNMOS is important as there are competing demands placed on the ggNMOS. It has been widely reported that if $V_{t2} > V_{t1}$ (i.e. the voltage at which damage occurs is greater than the trigger voltage), the ability of a structure to absorb ESD current is increased [2][59][61][83][85][87][88]. Increasing the drain length can promote this effect but at the expense of increased parasitic capacitance and increased silicon area (see section 1.2.1.1). To create the most efficient design one needs to understand the factors that will affect the resistance of the $n^{+}well$ under ESD conditions. It is well known [5] that velocity saturation and mobility temperature dependance can affect the resistivity of an n-well resistor, though in this case, we are also interested in interactions between the transistor and the drain resistor. Esmark et al. [25] and Liou et al. [148] showed that conduction modulation can occur in ggNMOS ESD protection. This is where the carriers generated in the avalanche multiplication within the reverse biased diode increases significantly the number of electrons available for conduction within the drain resistor or holes available for conduction within the substrate resistor. In this instance, the effect of temperature on mobility is not calculated as the simulator is unable to converge ggNMOS simulations that are not at room temperature.

In practice, the resistivity (in $\Omega/\square$) of an n-well resistor is a parameter that will be supplied by the foundry though the parameter will only apply to low currents and to resistors operating in isolation. The aim in this instance is to find under which conditions the foundry value for resistivity can be used and the ESD conditions in which the foundry values do not apply. The n-well resistors were simulated over a range of lengths ($300nm - 500nm$) so that the resistance of the contacts could be removed by calculation. The IV curves are shown in Figure 5.37. The IV traces are straight and show no sign of velocity saturation (the IV traces would bend to

the right if velocity saturation was occurring, i.e. the resistance would increase as the current increases). A best fit line was fitted to each trace in Figure 5.37 and the slope of each best fit line vs resistor length is shown in Figure 5.38. Fitting a least squares best fit line to the points in Figure 5.38 gives an offset resistance value of 1.96Ω and a slope of 126.2Ω/□. This resistivity value agrees with the extra resistance provided by the extended drain from the previous section. This would suggest that within the simulation there is no interaction between the operation of the transistor in snapback mode and the n-well resistor in high current operation.

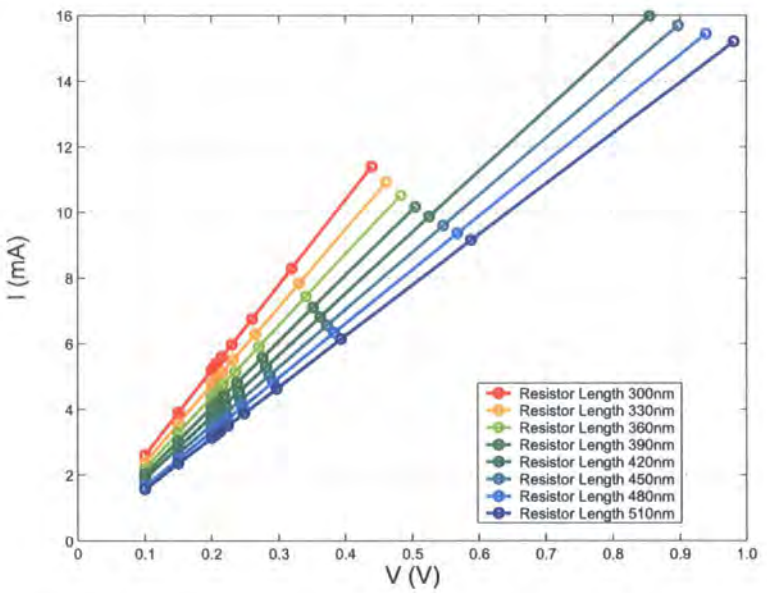

Figure 5.37: IV characteristics of an n-well resistor

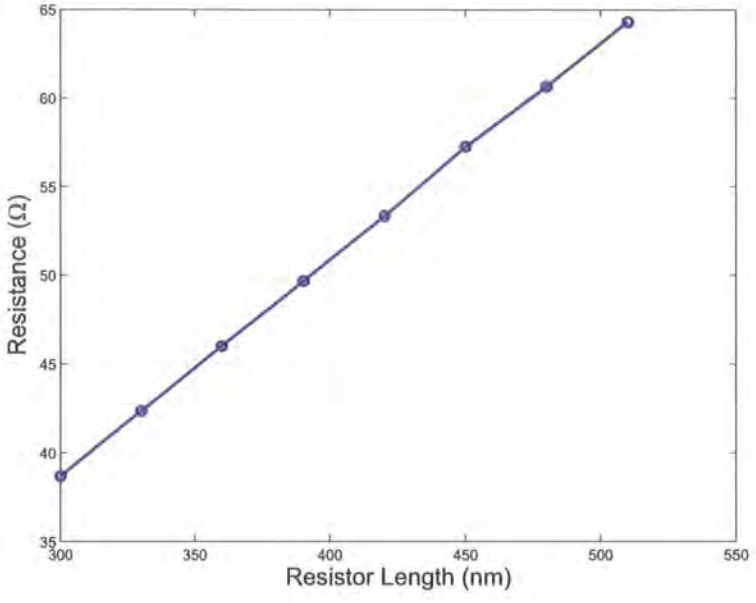

Figure 5.38: Resistance of the n-well resistors vs length

5.9 New Structures

The previous sections covered understanding gained whilst simulating the ggNMOS. This knowledge can be used to improve the electrical performance of existing designs, both during ESD events and during normal operation, as well as to create new ones.

5.9.1 Extended Source instead of Extended Drain

The fundamental difference between this new design and the standard ggNMOS design is that the drain resistor is moved to the source side of the transistor (Figure 5.39-5.40). Section 1.3.1 explained that the drain resistor is added to increase the drain voltage when I_{DS} is high which ensures that the current is shared equally across the width of a device as well as between devices in parallel. The source resistor should work in the same way as the drain resistor by increasing the potential of the parasitic bipolar transistor emitter which will decrease V_{BE} turning the transistor off resulting in an increased collector voltage. The increase in the collector

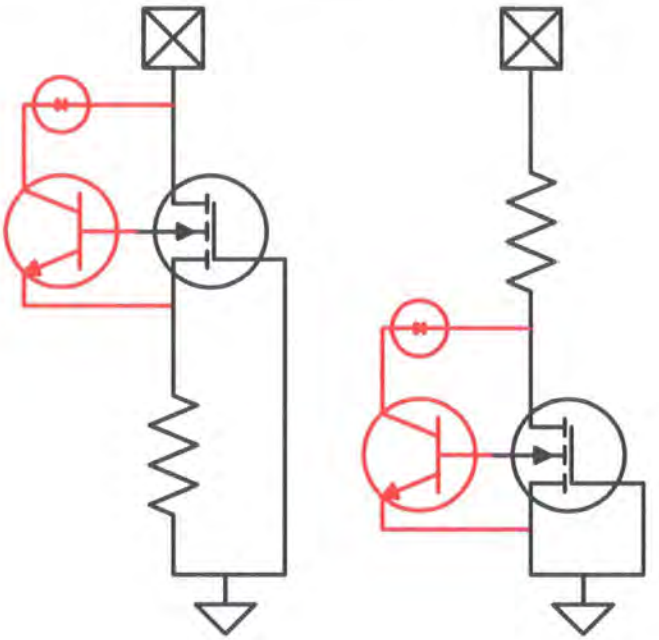

Figure 5.39: Extended source and extended drain ggNMOS schematic

voltage will in turn increase the base voltage and turn the transistor on again. Any increase in the emitter voltage should be closely repeated in the collector. The aim is to provide an alternative to the drain resistor as the long drain presents a parasitic capacitance to both the drain contact and any signal on the IO that the ggNMOS is protecting. The source resistor should provide the same function without the capacitive loading.

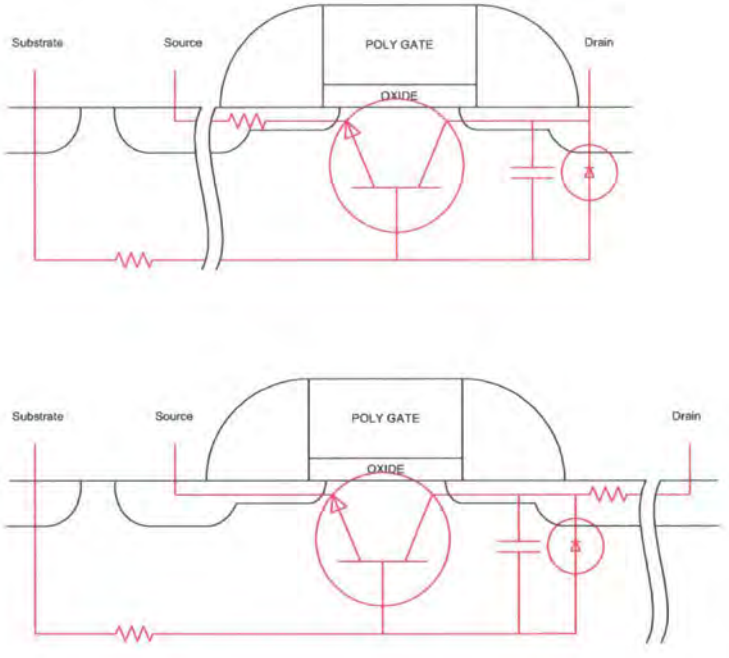

Figure 5.40: Extended source and extended drain ggNMOS layout

The IV curves for a ggNMOS with a long drain resistor, long source resistor and a ggNMOS with neither a long drain or long source are shown in Figure 5.41 and clearly the extended source resistor does not work as intended.

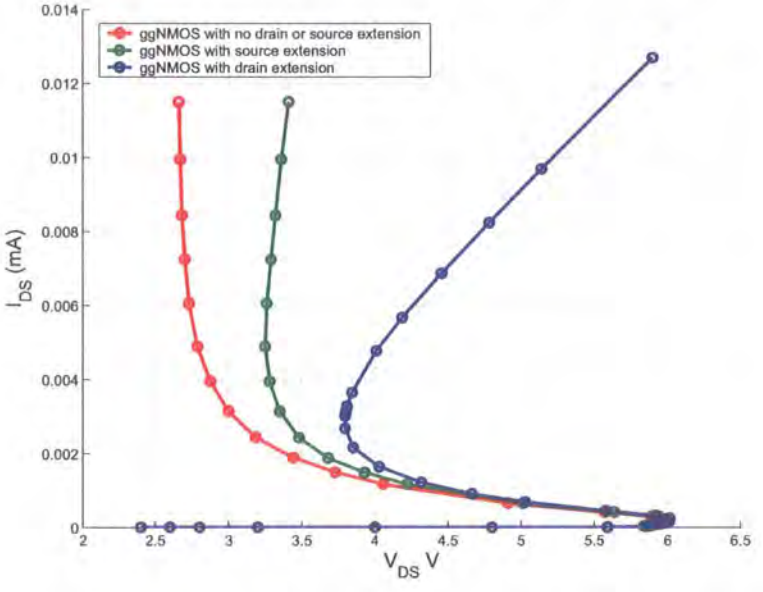

Figure 5.41: Extended source, extended drain and standard ggNMOS IV curves

The reason for this failure is that for an extended drain ggNMOS, the current stays in the drain resistor (Figure 5.42) while in the extended source resistor ggNMOS, the current *leaks out* through to the substrate (Figure 5.43) effectively decreasing the resistance of the source as the current is not confined to the n^+ doped regions.

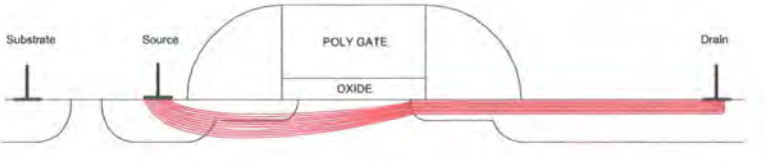

Figure 5.42: The current flow lines in a long drain ggNMOS

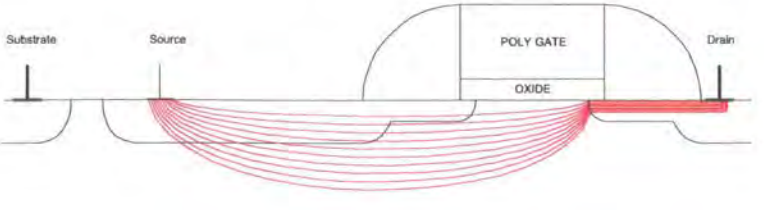

Figure 5.43: The current flow lines in a long source ggNMOS

In a triggered ggNMOS the majority of the carriers are created by the avalanche multiplication in the high electric field region found in the depletion region of the reverse biased bulk-drain pn junction. The electrons will drift/diffuse towards the drain and the holes towards the source. The electrons are confined to the drain due to the high electric field found in the drain-bulk reverse biased pn junction. As the source-bulk pn junction is forward biased, the electric field across the junction is minimal and therefore unable to confine the holes in the source n^+ well region.

Although the long source ggNMOS did not work, there is an alternative solution (Figure 5.44) which was shown to work. The source resistor is formed in polysilicon and the oxide underneath it prevents the current spreading into the substrate thus keeping the resistance high. An IV curve of this transistor is shown in Figure 5.45

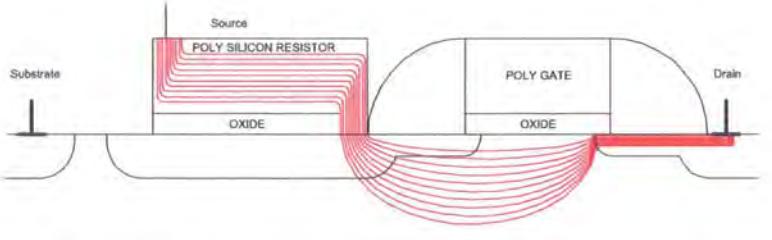

Figure 5.44: Extended source using a poly silicon resistor ggNMOS layout

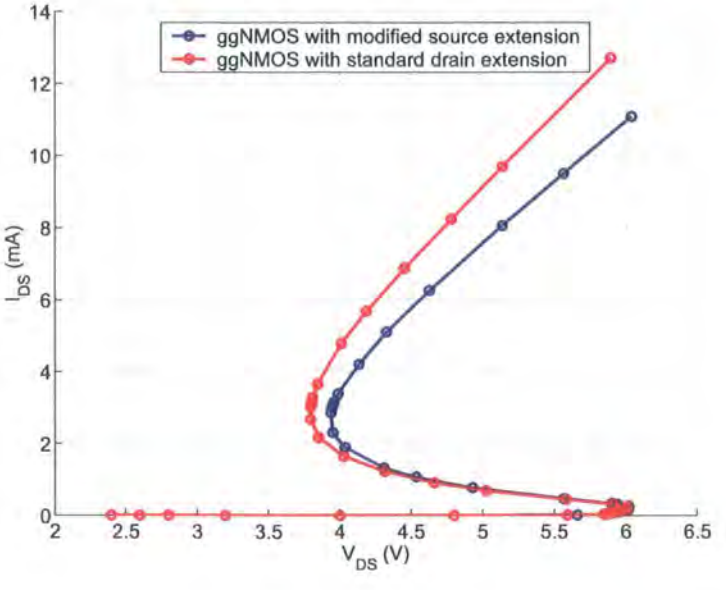

Figure 5.45: Extended source using a poly silicon resistor ggNMOS IV curves

5.9.2 Conclusions

Despite the fact that it was not possible to source a complete doping profile recipe that was used in the CSR's foundry, the author was able to construct a finite element semiconductor simulation that produced plausible results both in the standard operating modes as well as the ESD operating modes. This was done by sourcing the recipe from published data as well as some assistance from the foundry, nevertheless this is a tricky operation as the simulator results are very sensitive to variations in doping density and profiles.

The simulations supported the theory that snapback is an operation consisting of an interaction between a *parasitic* bipolar transistor, a diode across the bulk-drain junction and a resistor between the base of the transistor and ground. They also showed that drain leakage current (when $V_{DS} = 2V$) was strongly correlated to triggering voltage regardless of whether physical alterations, such as gate length or oxide thickness, or electrical stimuli alterations such as V_{GS} or V_{BS}, were used to alter the leakage current. This Chapter also explored the possible interaction between the snapback device and the extended drain resistor. No evidence of interaction between the two was shown. The simulator was able to illustrate the snapback mechanism by altering the parasitic circuit. In this instance, the resistance between the base of the bipolar transistor and ground was reduced by increasing the substrate doping and the simulator showed that a high drain current and voltage was required to trigger the device. It also separated the action of the avalanching reverse biased diode from the switching mechanism of the bipolar transistor. This series of experiments highlighted one of many flaws in the simulation software leading to inaccurate results. The conclusion is that there should be more flexibility in how the software is operated.

The simulator has also been able to simulate the ggNMOS in very fast rising V_{DS} conditions showing that as the edge speed is increased the triggering voltage is decreased. This does not mean that the peak voltage across the device is reduced

when a fast rising edge is applied as the triggering speed of the device is insufficient to prevent V_{DS} significantly overshooting the DC triggering voltage.

In the final section of the Chapter this knowledge is used to develop a new method for ensuring that the current spreads out across the whole width of the device. Initially the proposed solution is shown to perform badly but a small modification to the original concept is shown to work well.

5.10 Summary of contribution to knowledge

In Chapter 5 finite element semiconductor simulation software was used to investigate the operation of a ggNMOS under ESD conditions and the conduction mode known as *snapback*. It was shown that if the leakage current (I_{DS}) was increased by decreasing gate length, gate oxide thickness or increasing V_{GS} or V_{DS}, the triggering voltage would decrease. Regardless of the method used to increase the leakage current the variation in triggering voltage was the same, with the exception of simulations where $V_{DS} < 0V$. This would suggest that triggering voltage is a function of leakage current. It may also be the case that triggering voltage is also a function of temperature, since leakage is a function of temperature, though in this instance this hypothesis was not tested as it was not possible to get the simulator to converge.

The simulations also showed that the triggering voltage could be decreased if the rise time was decreased. The likely cause of this effect is capacitively coupled current providing an alternative current source to the base of the parasitic bipolar transistor. For very fast rise times the triggering voltage increased significantly and this may be due to the delay in triggering mechanism.

One of the advantages of finite element semiconductor simulation is that the process parameters of the transistor can be easily altered. This would be very time consum-

ing and difficult to do in silicon. In this instance it was possible to demonstrate, by increasing the substrate doping, the snapback action in its constituent components. Typically, once the drain-substrate diode starts to avalanche, the whole snapback process is triggered. However in this instance, the substrate resistance was deliberately decreased, which means that a significant avalanche current is required to increase the substrate voltage so that the parasitic bipolar transistor can be turned on. The triggering occurred approximately $0.7V$ above the initial avalanching voltage.

Interactions at high currents between the drain and the transistor in snapback have been reported where the carriers created in the avalanching junction diffuse into the extended drain resistor thus reducing the resistance of the extended drain. It was not possible to reproduce these conditions with the simulator.

A novel structure was proposed with an extended source instead of an extended drain as it was believed that the extended source would be as effective in spreading the current. Simulations showed that an extended source would not work as the current would not be confined to just the source doping in the same way as it is confined to the drain doping when an extended drain is used. The spreading of the source current has been resolved by replacing the extended source n^+ resistor with a poly resistor. This type of resistor was shown to work well as the current was confined to the poly resistor thus ensuring that the effective resistance was high.

Chapter 6

Test Structures for New Protection Devices

In the intensely competitive world of semiconductor design and manufacturing, every effort must be made to reduce chip area and this has been achieved in analog and digital circuit design by shrinking the processes/transistor size. Unfortunately within the IOs the trend is in the other direction as chip pin counts are increasing. Process shrinking does little to assist ggNMOS shrinking since a significant proportion of its area is used to layout the extended drain. Moreover the transistor width cannot be reduced as the energy dissipated within the transistor during an ESD event is broadly unchanged by process shrinking. Reducing the size of ESD protection structures not only would have the advantage of reducing silicon area, but it also would reduce the parasitic capacitive load on high speed or RF IO signals which are increasingly common on CMOS processes.

For this Chapter, 24 test structures were designed and included on a commercial $0.18\mu m$ process 4th generation CSR Bluetooth chip to provide a comparison with the simulations and to assist with future ESD protection structure design. The aim in this experiment is to verify accuracy of simulation, compare with results reported in

the literature and to measure variation in n-well resistance under low current (using an n-well resistor) and high current conditions (using an extended drain ggNMOS).

6.1 Introduction: The Test Structures

In this investigation, there are two sets of test structures: The first set of structures are derivatives of a standard ggNMOS structure design provided by the foundry and will be tested using the TLP tester at currents of $< 1A$. The second test structures are a set of resistors of various geometries which will allow the resistance of the extended drains of the first set of test structures to be measured in isolation from the whole ggNMOS. The aim is to measure the resistance at low currents ($< 1mA$) to determine if the drain resistance changes as the current increases due to effects such as velocity saturation or conduction modulation. Conduction modulation is an interaction between the ggNMOS and the extended drain. During this interaction, the number of carriers provided by the dopants in the drain is exceeded by the carriers provided by the avalanching junction and the parasitic bipolar transistor in the ggNMOS causing the drain resistance to decrease. The ggNMOS cannot be tested at very low currents because such devices need a minimum current to remain in a triggered state; equally crucial are the limitations in the operation of a TLP tester as described by Figure 2.9 in section 2.4.

6.2 Layout Design of the Test Structures

Figure 6.1 shows the top level layout of the chip onto which the test structures were added; the red box shows the position of the ggNMOS structures and the blue box highlights the resistors' location. Figure 6.2 shows the ggNMOS test circuit where all of the test structures are connected in parallel using *Metal6* tracks to a single

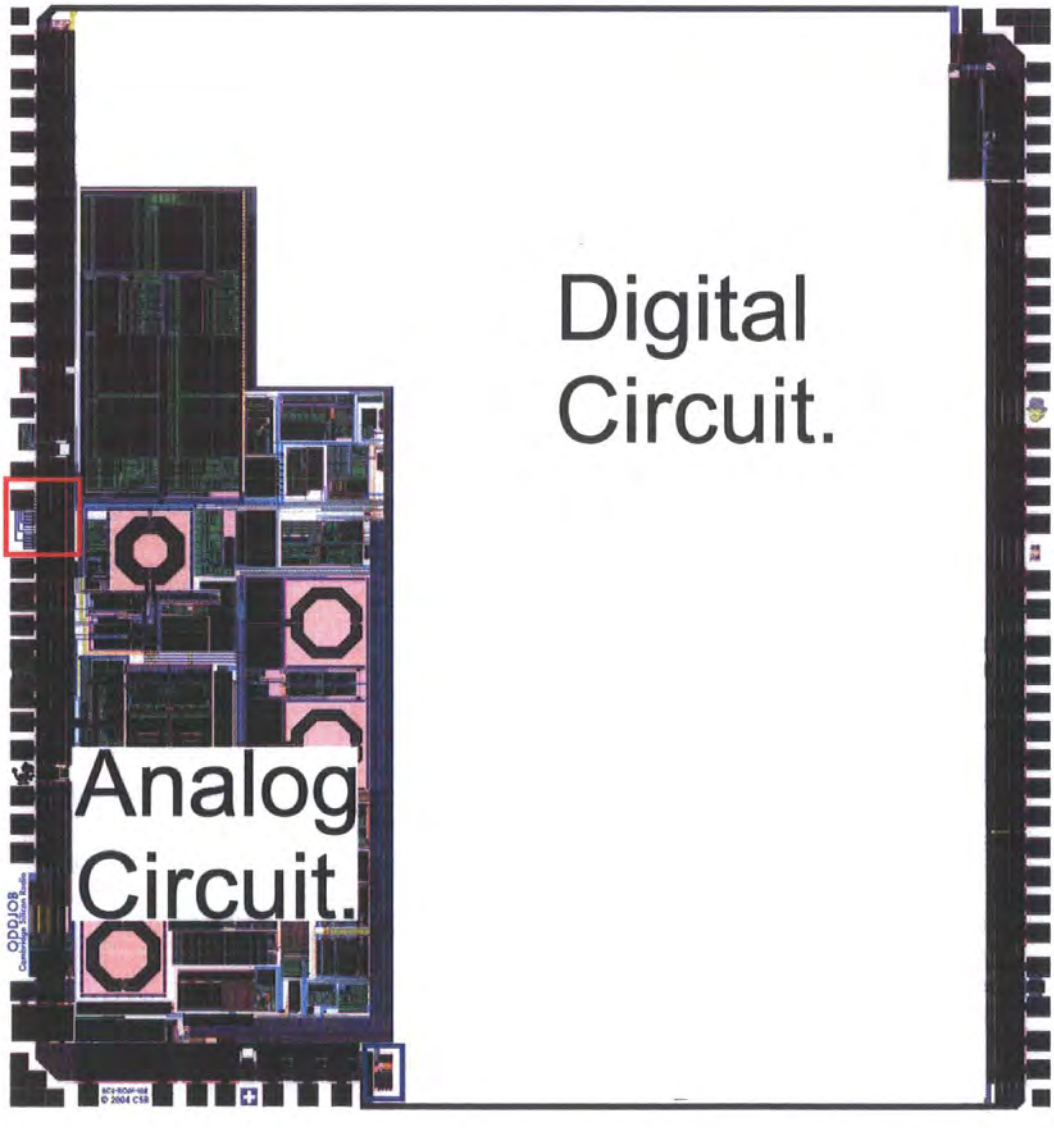

Figure 6.1: Top level layout of the CSR chip

bond pad to save silicon area. To access an individual test structure, the Focused Ion Beam (FIB) apparatus was used to cut selectively the *Metal6* tracks, to disconnect all the test structures except for the test structure that needed to be measured. The chip was then connected to and tested with a TLP tester in the same way as the methods reported in section 3.5.

The blue box in Figure 6.1 highlights the position of the second set of test structures on the chip. Figure 6.3 shows a close up of the region where the ggNMOS test structures were added. A multiplexer was used to connect a current source and voltage sense to each of the test structures as illustrated by the circuit in Figure 6.4. As the return path of the current is via the ground rail, there was also a test connection to ground to measure the ground resistance as it was expected that the ground resistance would be higher than the test structure resistance.

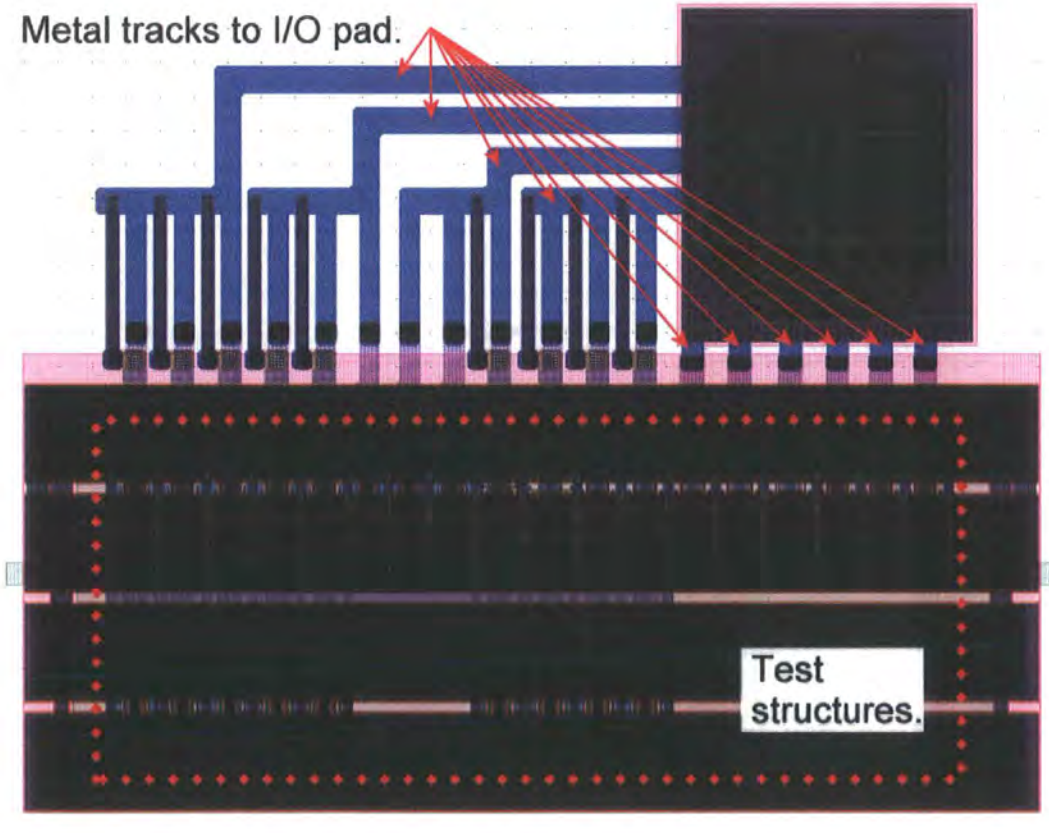

Figure 6.2: A close up of the ggNMOS test structures

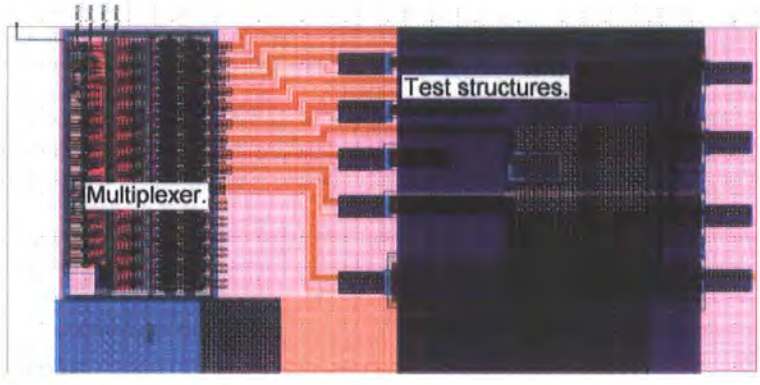

Figure 6.3: A close up of the low current test structures

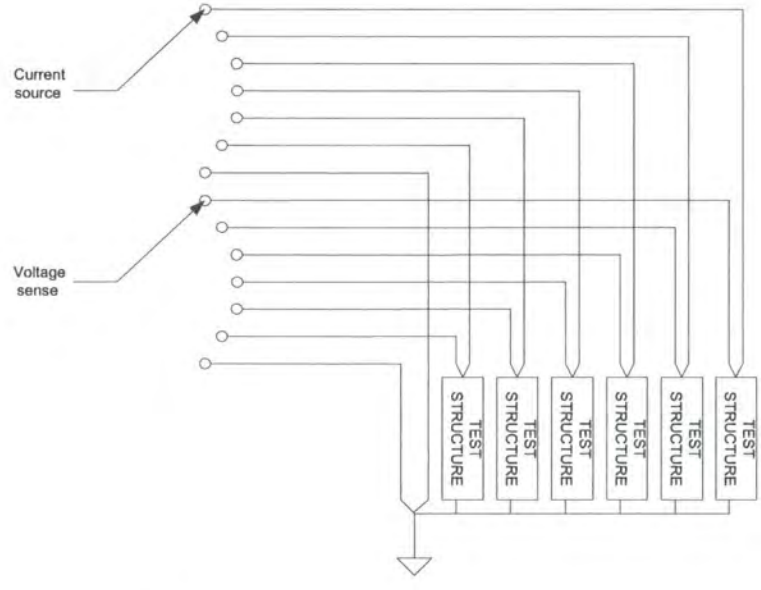

Figure 6.4: The circuit used to characterise the low current test structures

6.2.1 ggNMOS Device with a Variety of Gate Lengths

It has been reported in the literature [95][96] that the trigger and snapback voltages increase as the gate length increases and these results are supported by the simulations reported in section 5.6. Given that the papers are based on an older ($0.25\mu m$) technology, the link between gate length and trigger voltage may not be applicable on $0.18\mu m$ technology. Equally, we cannot rely on the simulations as the doping densities and profiles used were derived from publicly available data [18][142][149] for a $0.18\mu m$ process and not from the foundry doping profiles. In addition, it has been reported that it is difficult to achieve good correlation between simulation and measurements.

In this experiment the standard design supplied by the foundry for a ggNMOS was used. Five transistors were designed with gate lengths of $0.18\mu m$, $0.21\mu m$, $0.25\mu m$, $0.29\mu m$ and $0.35\mu m$ with all other feature sizes being identical and as recommended by the foundry. The $0.18\mu m$ and $0.35\mu m$ designs are shown in Figure 6.5.

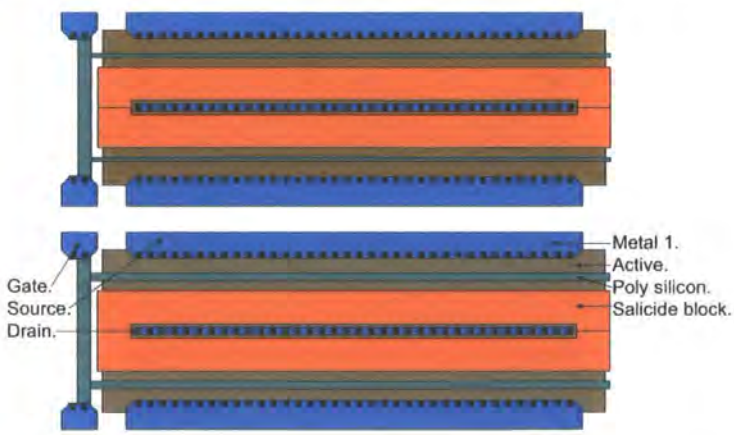

Figure 6.5: The layout of a $0.18\mu m$ (top) and $0.35\mu m$ (bottom) version of the $1st$ set of test structures with variable gate length

The test chips were prepared and tested as described in section 3.5 although in this case *Metal6* was cut to isolate individual transistors. (See Figure 6.6 for the resultant IV traces). Figures 6.7 and 6.8 show the hold voltages at low currents and the triggering voltages in detail. The data produced by the TLP tester is relatively

noisy though Figure 6.7 clearly shows that the hold voltage increases as the gate length increases. In Figure 6.8 it can be seen that the variation in trigger voltage $(6.2V - 6.4V)$ is only marginally greater than the TLP tester resolution $(100mV)$. Despite this, a trend can be seen showing a small increase in trigger voltage as the gate length increases. A summary of the results including a comparison with the simulation results is shown in Table 6.1.

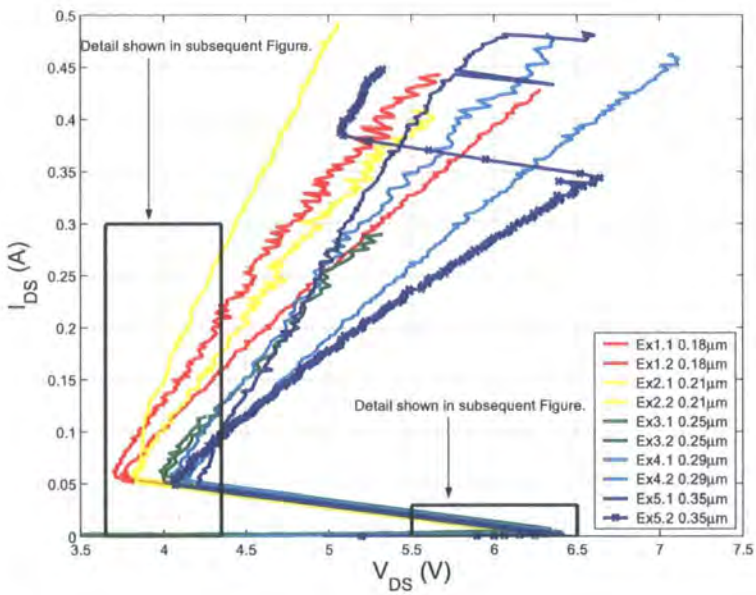

Figure 6.6: IV curves of the test ggNMOS's with variable gate lengths

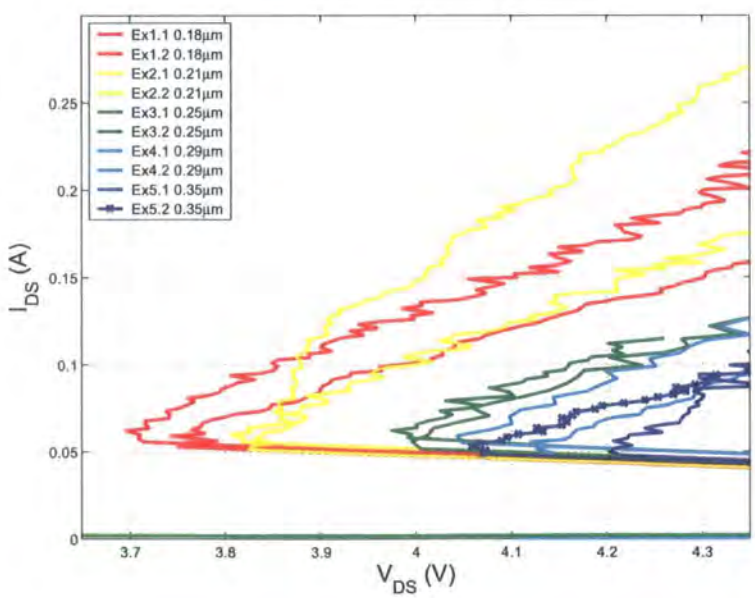

Figure 6.7: Detail of Figure 6.6 showing ggNMOS once triggered

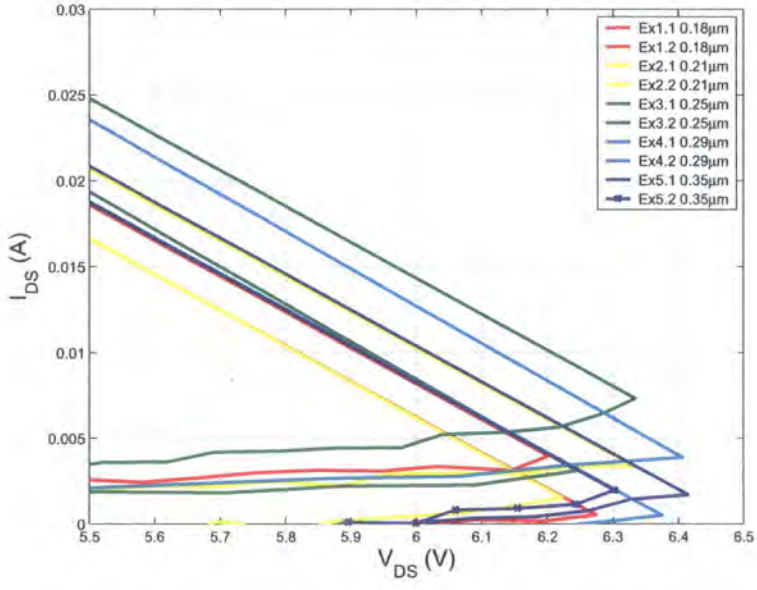

Figure 6.8: Detail of Figure 6.6 showing the triggering of the ggNMOS

Experiment	Δ Trigger Voltage	Δ Hold Voltage	Range of gate lengths
Sequoia simulation	$2.8V$	$0.362V$	$0.18\mu m$ to $0.33\mu m$
Synopsys simulation	$1.15V$	$0.373V$	$0.18\mu m$ to $0.33\mu m$
ggNMOS measurement	$\sim 0.1V - 0.2V$	$\sim 0.45V$	$0.18\mu m$ to $0.35\mu m$

Table 6.1: Trigger voltages and the hold voltages of measured and simulated devices with variable gate lengths

Overall, the direction of variation in the trigger voltage and the hold voltage is consistent with the literature and simulations but the variation in trigger voltage is about an order of magnitude less. A possible explanation is that the rise time is sufficiently fast to cause triggering by capacitive coupling therefore reducing the triggering voltage, as was demonstrated in sections 3.5.4 and 5.7.6. It could also be argued that the triggering voltage during capacitive triggering is not dependent on the gate length. Two ggNMOSs with a $0.18\mu m$ gate length and two with a $0.35\mu m$ gate length were TLP tested using both fast ($1ns$) and slow ($30ns$) rise times and it was found that the difference in triggering voltage was between $10mV$ and $60mV$ for fast and slow rise times and therefore insignificant. As a $30ns$ TLP rise time is slow in the world of $0.18\mu m$ process transistors, so it can be assumed that the device is not triggering by capacitive triggering.

Given that it has been established that it is not triggering due to capacitive effects, the snapback mechanism description from the literature review needs to be revisited (summarised from section 1.2.1.4):

The snapback mechanism is an interaction between the avalanche multiplication effect (M) and the gain of the parasitic bipolar transistor (β). The high potential across the drain-substrate junction multiplies any leakage current flowing through it which is subsequently conducted to the substrate tap. The increase in the substrate potential turns on the bipolar transistor which in turn supplies more current back to the avalanching junction where it is multiplied again. This process relies on the fact that there will always be some leakage through a reverse biased pn junction to get the snapback process started. Snapback will occur once $M * \beta > 1$. At this point, the current will rapidly increase turning the bipolar transistor fully on, which will supply a large current to the avalanching junction. Once triggered the avalanching junction only needs to supply a small current to the base of the transistor, I_{DS}/β, the avalanche multiplication required falls to $1/\beta$ and therefore the voltage across the

junction decreases sharply.

The leakage current through a reverse biased diode is directly related to the minority carrier concentration either side of the junction. If the leakage is low, then the amount of carrier multiplication required to turn the bipolar transistor on is high. This is because, for the same output current, less input current (I_{leak}) is supplied therefore requiring a greater carrier multiplication. Ultimately the triggering voltage is increased. At higher voltages, across a reverse biased pn junction, the multiplication factor increases and the $\frac{dV}{dM}$ decreases (as $M\alpha V^n$) which means that for a small change in β there will be smaller change in triggering voltage (remember that triggering occurs when $M * \beta > 1$). Therefore, as the gate length is increased, β will decrease and there will be a smaller increase in trigger voltage. Conversely, if the leakage current is high, the variation in trigger voltage will be high over a range of gate lengths.

It is worth noting here that β is a function of I_{CE} and therefore a function of the leakage current (illustrated in Figure 6.9 [5]). Therefore M will need to be even higher if the leakage current is low; consequently $\frac{dV}{dM}$ will decrease further still, as the avalanche multiplication rate is an exponential function of voltage.

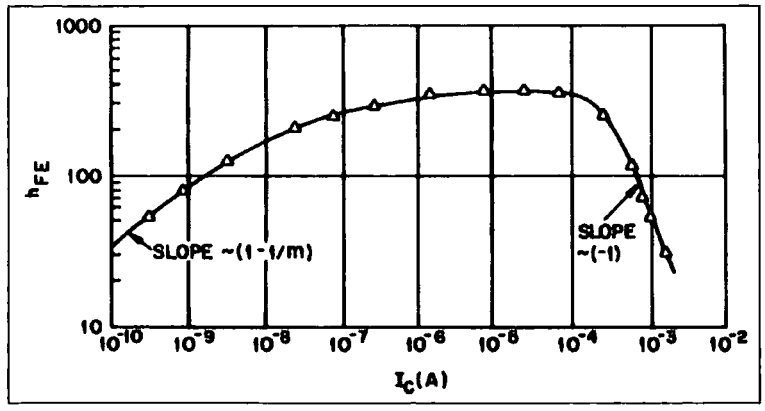

Figure 6.9: Variation in the bipolar current gain (β) as a function of I_{CE} for a bipolar transistor [5]

This theory is consistent with the results illustrated in Figures 6.6, 6.7 and 6.8 and Table 6.1 which shows that V_{hold} does change with gate length and therefore

changes with β; yet the trigger voltage is unaffected by variations in β. The leakage current, when the ggNMOS is off, should have no effect on its performance once it has undergone snapback. Figure 6.10 shows the leakage current of the Sequoia ggNMOSs and of the real ggNMOSs vs Gate length at 3V DC. This data was derived from SPICE simulations using foundry supplied models. The simulated ggNMOS's have a leakage current that is more than three orders of magnitude greater than the test structures.

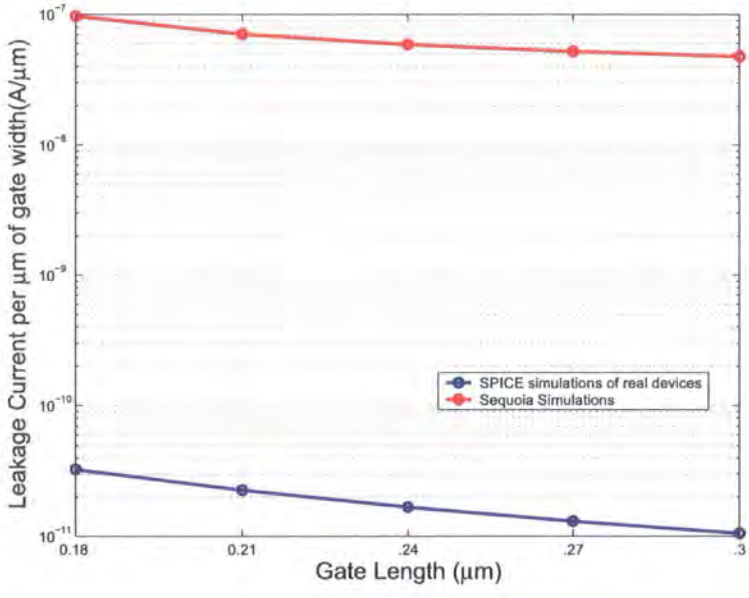

Figure 6.10: Comparison of the leakage current of a ggNMOS calculated using SPICE and Finite Element Simulation

To test this theory, a new Sequoia simulation was carried out identical to the ones described in section 5.5 except that the V_T doping density in the p-type substrate region, just below the gate oxide, was increased by an order of magnitude. This did not alter the properties of the bulk of the silicon. As the leakage current is a function of the minority carrier density either side of the drain-substrate pn junction, the increased doping density will decrease the leakage current. Figure 6.11 shows that the modification does decrease the leakage current as expected.

The IV characteristics of the modified ggNMOS and the hold voltage and triggering in detail are in Figures 6.12, 6.13 and 6.14 respectively. The variation in the

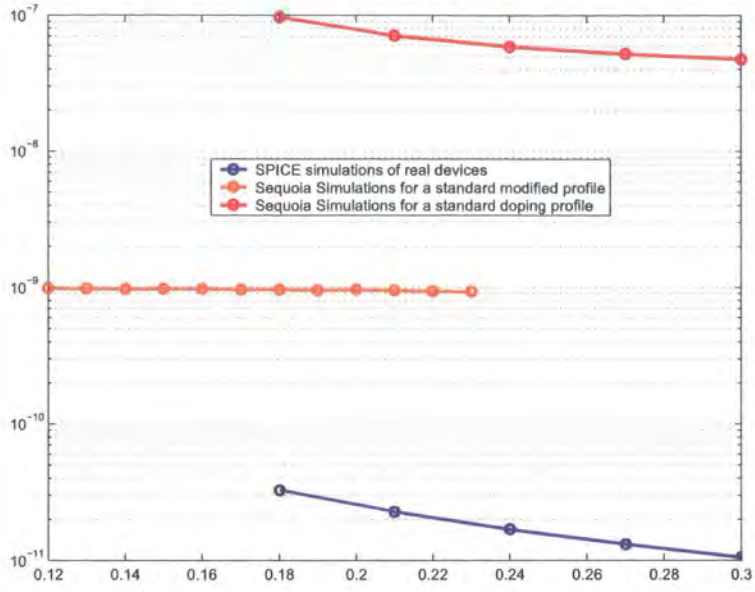

Figure 6.11: Change in I_{leak} when the V_T doping profile is changed

triggering voltage is in the order of $30mV$ and appears to be below the measurement limit of the simulator but it is clearly far less than the variation in triggering voltage seen in previous simulations (See Figures 5.14 where $\triangle V_{t1} \simeq 800mV$ and $\triangle V_{hold} \simeq 330mV$ for a $\triangle L_{gate} \simeq 90nm$ and Figure 5.15 where $\triangle V_{t1} \simeq 2.0V$ and $\triangle V_{hold} \simeq 240mV$ for a $\triangle L_{gate} \simeq 90nm$).

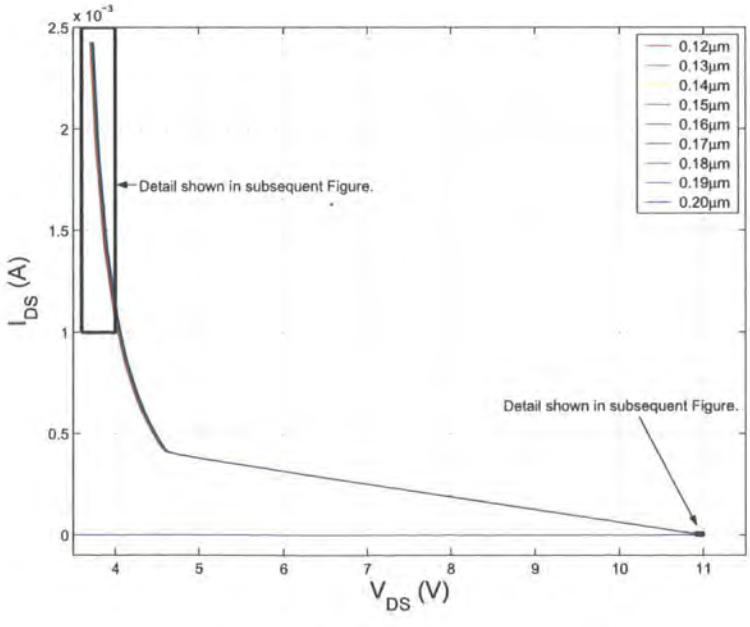

Figure 6.12: IV characteristics of a modified ggNMOS

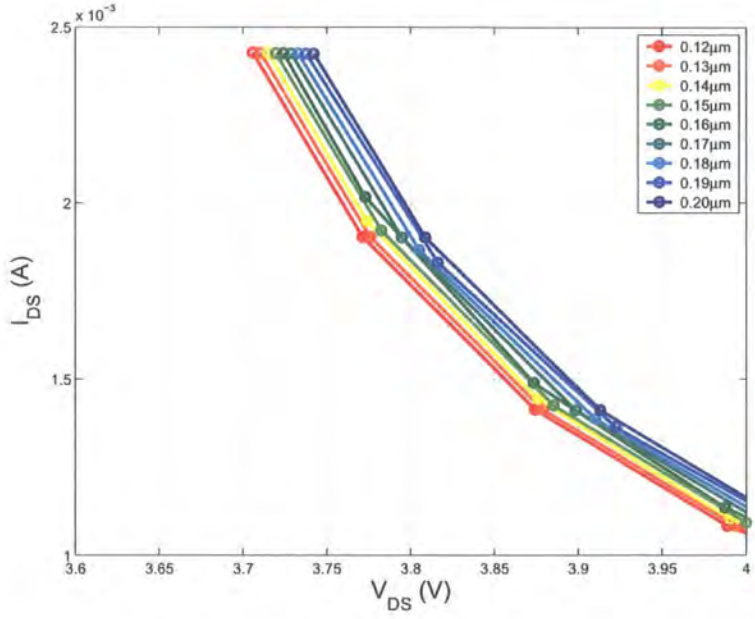

Figure 6.13: Detail from Figure 6.12 (hold voltage) of the IV characteristics of a modified ggNMOS

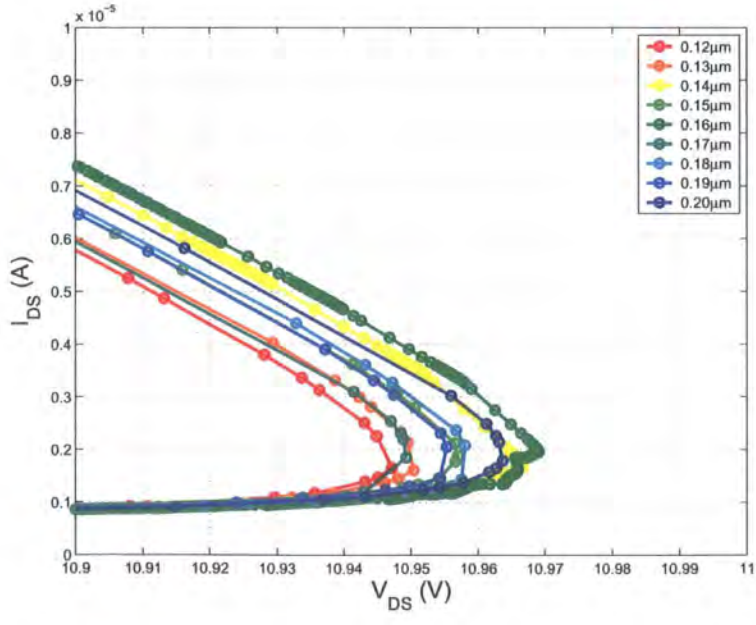

Figure 6.14: Detail from Figure 6.12 (triggering voltage) of the IV characteristics of a modified ggNMOS

6.2.2 ggNMOS Device with a Range of Drain Lengths

This is a rerun of the previous experiment in section 6.2.1 with a constant gate length and a variable drain length. Five devices were designed with a $0.25\mu m$ gate length and drain lengths of $0.75\mu m$, $1.06\mu m$, $1.5\mu m$, $2.12\mu m$ and $3\mu m$. The layout of the devices is similar to the transistors in Figure 6.5. The aim in this section is to find out if the resistivity of the extended drains is constant or if it is subject to conduction modulation, as has been reported in the literature [148], or velocity saturation. The results in this section will be compared to the results derived from the modified drain structures in the next section 6.2.3. Figure 6.15 shows the IV curves of the five devices in breakdown.

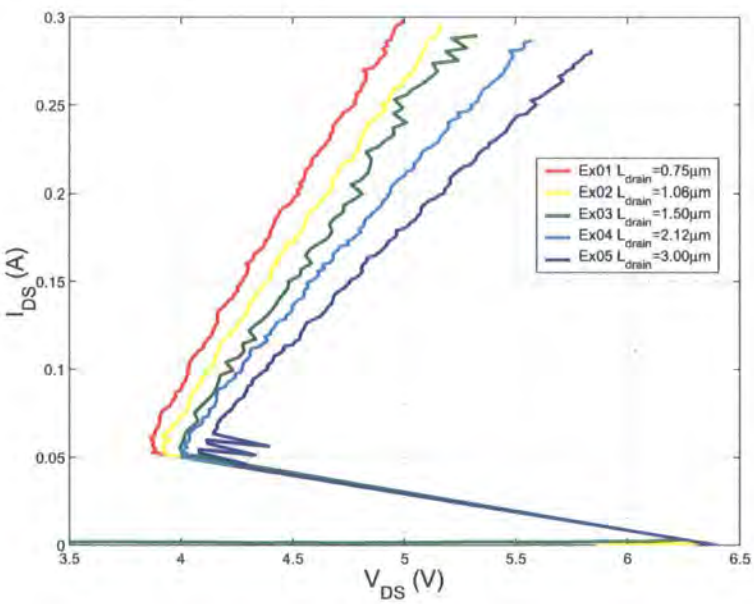

Figure 6.15: IV curves of the test ggNMOS's with variable drain lengths

Figure 6.15 shows a slight increase in resistance as drain length increases. Two *least squares best fit lines* were fitted to the IV curves for current values $75mA - 125mA$ and $200mA - 250mA$. The reciprocal of the gradient (effectively the resistance) of these lines vs drain length are displayed in Figures 6.16 and the x-axis intercept vs drain length in Figure 6.17.

The x-axis intercept does not vary with drain length which shows that for the range

of values used, the drain length has no effect on the operation of the ggNMOS other than to add a series resistance. The slope of the best fit lines, and hence device resistance, does increase as the drain length increases. Calculating a best fit line for the values in Figure 6.16 results, ignoring the anomalous values for $1.5\mu m$, in a gradient of $1.63\Omega/\mu m$ ($2.49\Omega/\mu m$ for measurements taken between 200mA-250mA) and an offset of 3.32Ω (3.58Ω for measurements taken between 200mA-250mA). The offset is due to other resistances in series with the extended drain such as the wire bonds, the source resistance and the resistance of the metal tracks. The ggNMOS drains are $22.5\mu m$ wide and the overall drain width is $44.9\mu m$ given that there are two transistors in parallel. This means that the measured surface resistivity is $55\Omega/\square$ ($73\Omega/\square$ for measurements taken between 200mA-250mA). The low current resistivity measure is reasonably close to the foundry specification value of $59\Omega/\square$. The fact that the ggNMOS resistance increases with current is probably due to velocity saturation effects and shows that conduction modulation is not happening at 250mA. The increases in resistance could also be due to IR heating which also reduces the carrier mobility.

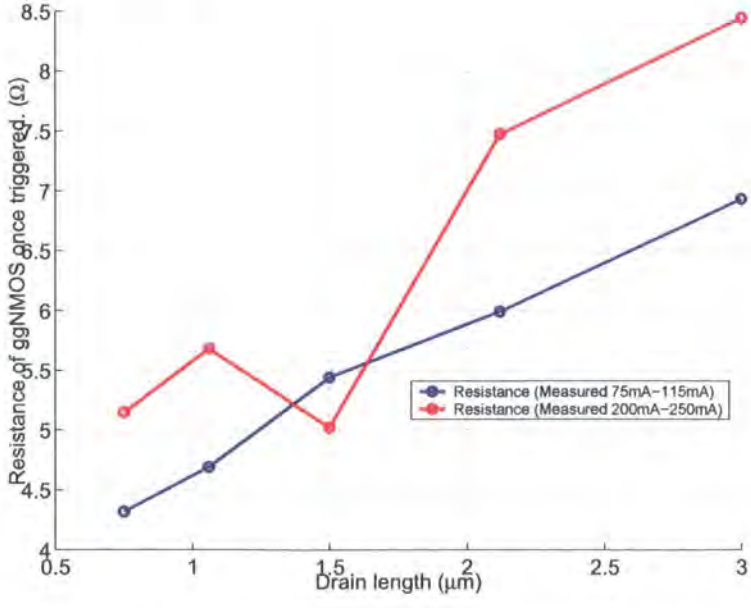

Figure 6.16: The resistance of the ggNMOS once the device is triggered over a range of drain lengths

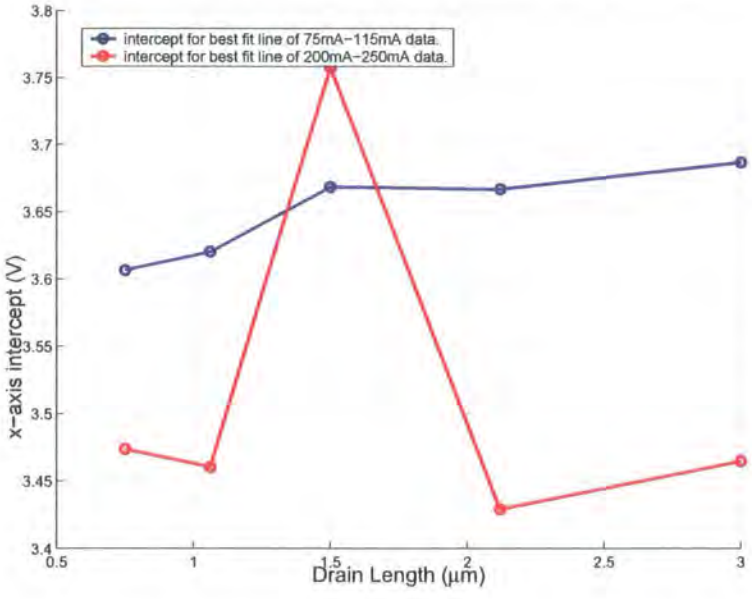

Figure 6.17: The offset voltage of the ggNMOS once the device is triggered over a range of drain lengths

6.2.3 ggNMOS with Holes in the Active Layer

The aim in this experiment is to create ggNMOS structures with a lower area usage and lower capacitance but still sufficient drain resistance to ensure efficient ggNMOS operation. Section 1.2.1 explained that the greater the failure voltage (V_{t2}), with respect to the triggering voltage (V_{t1}), the greater the current spreading across the width of the device and the greater the current sharing between devices that are wired in parallel. It is for this reason that the drain extension was added to the ggNMOS design. The downside of the drain extension is that it uses chip area and increases the parasitic capacitive loading on the IO. It would be preferable to increase the drain extension resistance without an increase in area or parasitic capacitance.

In an attempt to increase the drain resistance without increasing area, the standard foundry supplied ggNMOS transistor design was modified by separating the drain into fingers. The width of each finger was set at the Minimum Design Rule (MDR) dimension of $0.22\mu m$ and the size of the holes was set at $0.28\mu m$ (which is MDR for holes) $0.40\mu m, 0.55\mu m, 0.74\mu m$ and $1.00\mu m$. Figure 6.18 shows the layout of a

$0.28\mu m$ hole width ggNMOS design. The aim is to reduce the cross sectional area of the drain region and therefore increase the drain resistance. Of particular interest is whether the surface resistance (measured in $\Omega/\square$) of a fingered device is comparable to the values measured in section 6.2.2 and to the values specified by the foundry of $59\Omega/\square$. The IV curves of the ggNMOSs with holes in the active layer are shown in Figure 6.19.

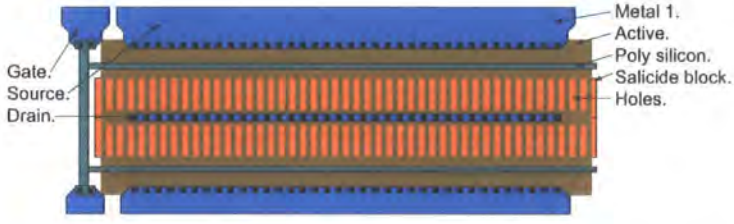

Figure 6.18: The layout of a ggNMOS with holes in the extended drain

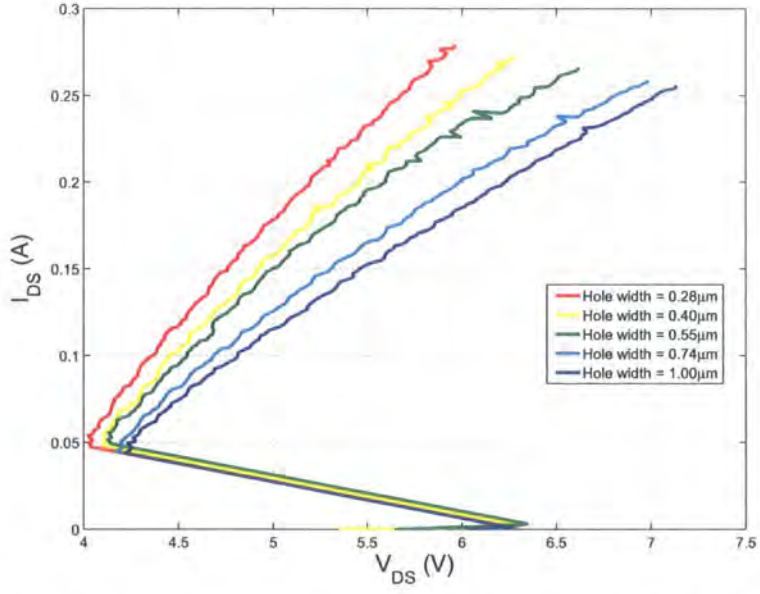

Figure 6.19: The IV characteristics of the ggNMOSs with holes in the extended drain

Best fit lines were fitted to the IV curves for current values of $75mA - 115mA$ and $200mA - 250mA$ giving a graph of Resistance vs Hole Size (Figure 6.20). The graph also includes the resistance of a ggNMOS without any holes in the extended drain; this result being obtained from the measurements in section 6.2.2. For a $75mA$ to $115mA$ current range, the graph has a slope of $49\Omega/\square$ ($72\Omega/\square$ for measurements taken between $200mA - 250mA$). Both measurements straddle the foundry value

of 59Ω/□. Again the drain resistance increases with test current which agrees with the results in section 6.2.2.

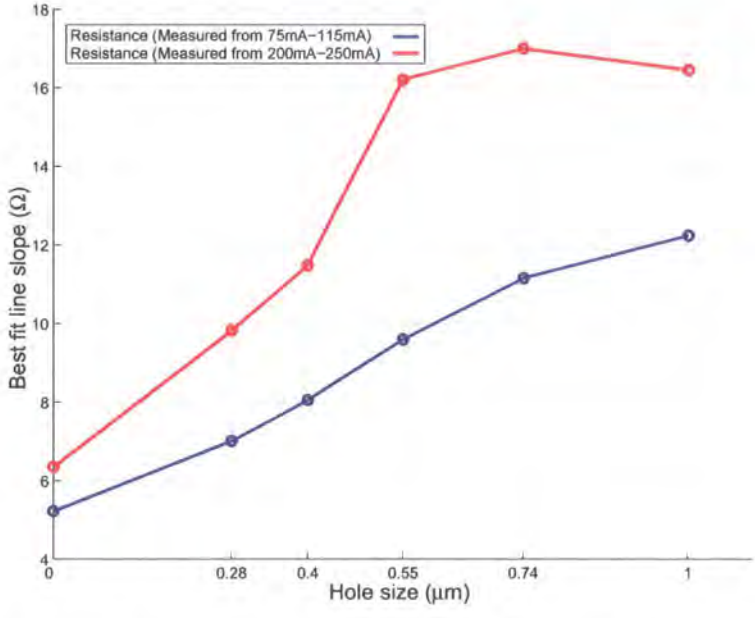

Figure 6.20: The drain resistance vs hole size of the ggNMOSs with holes in the extended drain

Figure 6.21 shows the x-axis intercept for these Best Fit Lines vs Hole Size. This graph shows that for low currents, the drain modifications have little or no effect on the operation of the ggNMOS.

This novel device design has been shown to be particularly effective in increasing the drain resistance. Therefore it should be possible to decrease the drain length without reducing the ESD performance. However this requires further investigation.

6.2.4 n-well Resistors with Holes in the Active Layer

These resistors have the same layout as the ggNMOSs in section 6.2.3 except that the poly-silicon gate has been removed. The aim is to compare the resistance of the drain of the ggNMOSs in section 6.2.3 separately from the transistor to find if there are any high current effects or interactions between the ggNMOS and the extended

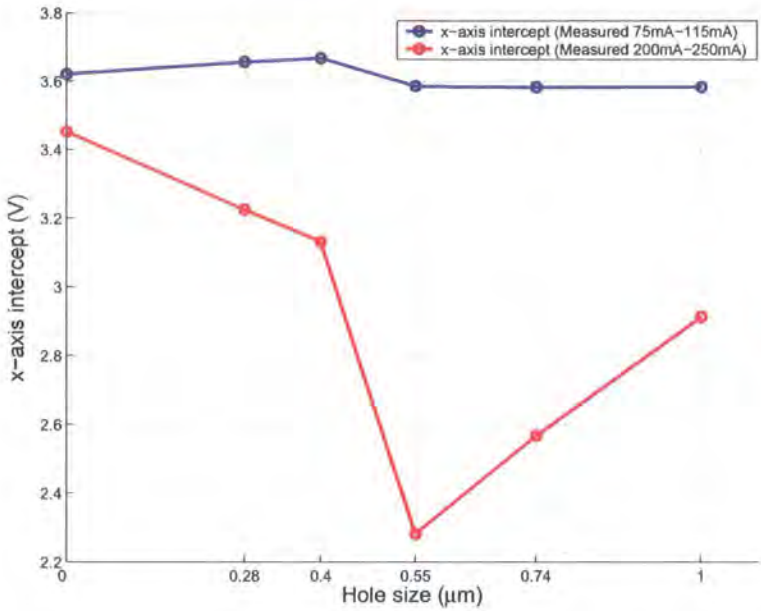

Figure 6.21: The voltage offset vs hole size of the ggNMOSs with holes in the extended drain

drain resistor.

These test structures were designed in such a way that they could be tested automatically using a Teradyne Tester at CSR. This tester is a development version of a Teradyne production chip tester used at the foundry to characterise the chip. The tester was programmed by a test engineer to pass $1mA - 7mA$ in $1mA$ steps through each test structure using the built-in multiplexer. The tester then measured the voltage at the top of each test structure using another multiplexer (Figure 6.4). This initial test was done to determine how much current could be passed through the chip without exceeding the power supply voltage of the chip. The resistance of each device was calculated using Equation 6.1.

$$Resistance = \frac{V_{device} - V_{ground}}{I_{test}} \tag{6.1}$$

The results were then divided by the calculated resistance at $1mA$ so that any change in resistance as the current increases can be detected. The resultant data

are shown in Figure 6.22. The graph shows that the resistance of the standard n-well resistors (red lines) varies by less than 0.2% over the current range and that the n-well resistors with holes (blue lines), the resistance varies by less than 2.5% for currents up to $6mA$. The reason why the $7mA$ reading had a high error was because the current supply voltage exceeded the chip supply voltage. The standard n-well resistors did not suffer this problem as their resistance is lower. This test shows that the resistance is constant and that the chip is able to drive the test structures at currents of up to $6mA$.

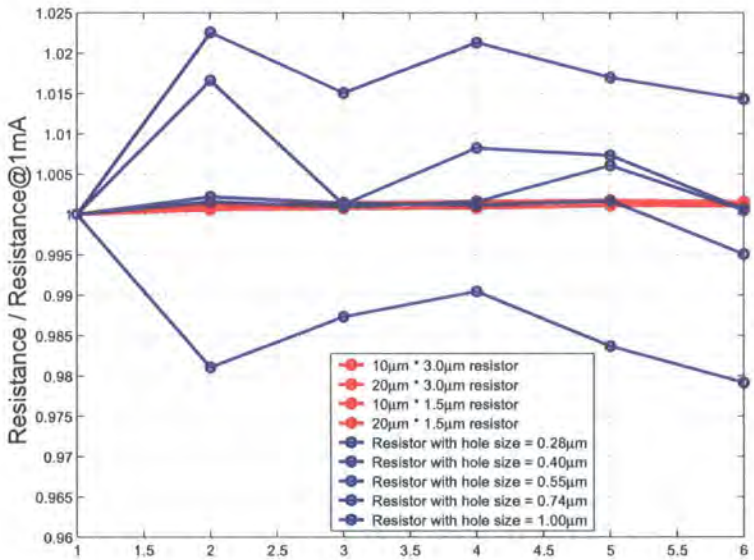

Figure 6.22: Change in measured Resistance vs Current

The tests were repeated on 25 chips at $4mA$ using the Teradyne tester. The resistance was calculated using Equation 6.1 and is presented in Figure 6.23 (the figure also shows the resistance data of the ggNMOS with holes in the drain resistor for comparison, see section 6.2.3). The green line in Figure 6.23 shows the difference between the blue and red line and is reasonably constant over the range of hole sizes indicating that the resistance of the two types of resistor is the same at high and low currents. The modified ggNMOS structure has a higher measured resistance than the *resistor with holes in it* because the voltage across the test structure is measured outside of the chip and therefore includes the resistance of the wire bond, bond pad

and tracks.

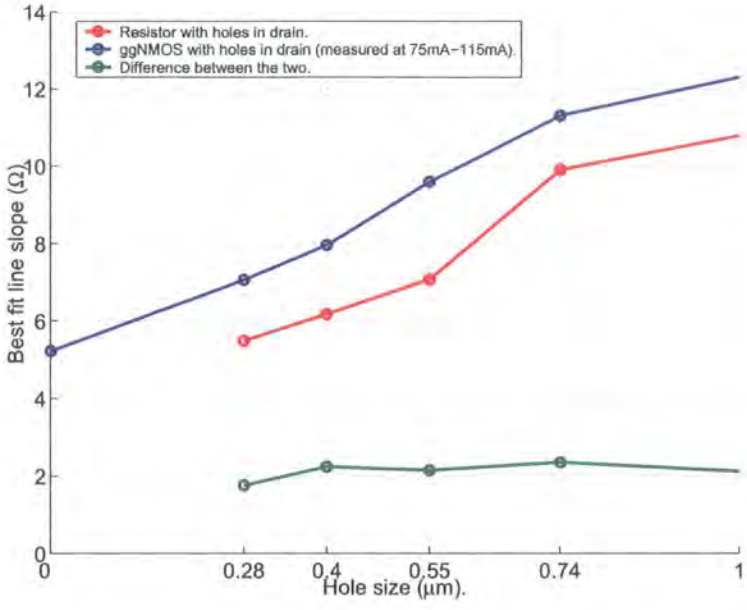

Figure 6.23: Resistance of ggNMOS and standard resistors for different active hole sizes

Fitting a least squares best fit line to the resistance values of the *resistors with holes in them* gives a resistivity of 53.45Ω/□ which is close to the foundry value of 59Ω/□.

6.2.5 Resistance of N+ Resistors

In this experiment, four resistors were designed in order to calculate the surface resistivity whilst any parasitic parallel resistance and contact resistances were calibrated out (Figure 6.25). The contact parasitic resistance error is due to the voltage sense contact not being directly on the resistor (Figure 6.24). The parasitic parallel resistance could be due to lateral diffusion of the resistor dopant or fringe effects of the optical mask.

Equation 6.2 was calculated for all four resistors and the values of surface resistivity ($\rho_{surface}$), parallel parasitic resistance (R_P) and contact resistance ($R_{contact}$) were adjusted to produce the minimum error (R_{error}) using Microsoft Excel. The contact resistance was calculated to be 21.1Ω and the surface resistivity to be 56.0Ω/□. The

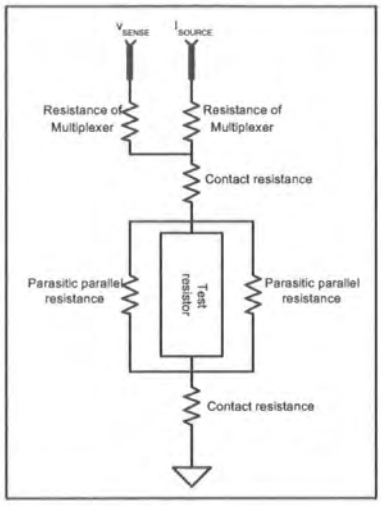

Figure 6.24: The location of the parasitics around the test resistors

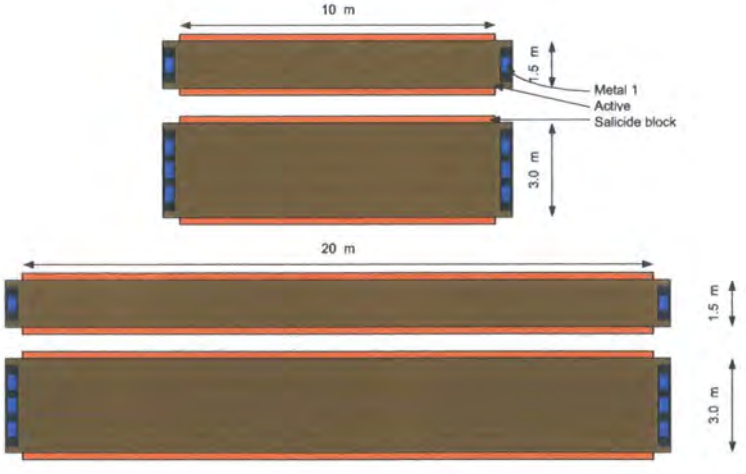

Figure 6.25: Layout of the test resistors

parallel parasitic resistance was calculated to be $> 1M\Omega$ though the solver was not able to come up with a definitive value. However, as it is large, the resistance can be ignored.

$$R_{error} = (R_{contact} + \frac{(\frac{L}{W} * \rho_{surface}) * (L * R_P)}{(\frac{L}{W} * \rho_{surface}) + (L * R_P)} - R_{meas})^2 \tag{6.2}$$

Equation 6.2 combines the estimated process error (R_P) and resistance error due to contact resistance ($R_{contact}$), with the projected resistance ($\frac{L}{W} * \rho_{surface}$) and compares it to the actual measured resistance (R_{meas}) to produce an overall term (R_{error}). Excel was used to find the error terms (R_P and $R_{contact}$) that produce the lowest overall error (R_{error}). These values can be employed to calculate the projected resistance for any size of resistor.

6.2.6 Conclusions

The results from section 6.2.1 have added considerably to our understanding of the ggNMOS as a ESD clamp, particularly since we were able to reproduce the unexpected result in simulation. These results showed that the avalanching process is much more dominant in the triggering mechanism than the original simulation results would suggest and therefore V_{t1} is almost exclusively dictated by the breakdown voltage of the drain-substrate diode. These results add further evidence from the simulations in Chapter 5 which suggested that the snapback operation was due to a number of parasitic structures working together. We must also remember that the ggNMOS is supposed to provide an alternative safe path through the chip for ESD currents. Therefore we have to be sure that this path, through the ESD clamp, is preferential to any other path through the circuit being protected. Given that the triggering voltage is only weakly dependent on the gate length, it makes it more difficult to protect a transistor in an output driver with an ESD clamp across it as

there would be a significant risk that the output driver would trigger preferentially to the ESD clamp.

The experiments detailed in section 6.2.2 have shown that the surface resistivity of the extended drain is broadly the same as the surface resistivity of a silicided blocked n-poly-silicon resistor for low currents ($\sim 100mA$). For high currents ($\sim 225mA$) the resistivity increases by $\sim$ 30% though it is unknown whether this is caused by velocity saturation or Joule heating effects. If caused by Joule heating effects it would indicate that the temperature of the extended drain had increased by $\sim 200°C$ using foundry temperature coefficient data. The experiments also failed to show any conduction modulation effects. In light of this result, the designer of a ggNMOS can predict how long the extended drain needs to be to achieve a failure voltage greater than the triggering voltage ($V_{t2} > V_{t1}$). This has been shown to ensure good width scaling across multiple ggNMOS clamps operating in parallel. Any prediction based on the low current surface resistivity of an n-poly silicide blocked resistor would produce a conservative design given that the resistivity of the extended drain has been shown to increase with current. Given that the ESD phenomena is a constant current effect (as $R_{ESD-source} >> R_{ESD-clamp}$), this increase in resistance will result in increased power dissipation in the drain resistor. However, this should not decrease the overall performance of the device (e.g. its ability to conduct ESD current to ground) as device performance is most likely limited by the power dissipation in the reverse biased drain-substrate pn junction where the electric field is highest. It may be possible to use this increase in drain resistance for high currents to decrease further the drain length and save area. This would require further experimentation with an array of devices being tested to the point of failure so that interaction between adjacent but separate devices could be determined.

We were able to show that restricting the width of the drain by blocking selected

areas of the doping in the drain extension produces entirely predictable performance; it was also shown that the surface resistivity of the extended drain remained constant regardless of the size of the holes in the active mask of the extended drain. This makes it relatively easy for the designer to predict high current IV performance and therefore he is able to use this design technique with a high level of confidence.

In this Chapter the correlation between $V_{t2} >> V_{t1}$ and maximum ESD tolerance has not been revisited as this has been studied many times. In addition, it would have required a large area for the test structures. The ultimate test of the validity of *ggNMOSs with holes in the active layer* as an area efficient ESD protection clamp would require testing of multiple designs in silicon each with a total gate width of $\sum 300\mu m$. Each design would have a different drain length and/or size of hole in active layer.

6.3 Summary of contribution to knowledge

In this chapter silicon measurements showed that if the gate length was increased, V_{t1} and V_{hold} would increase. A similar result was demonstrated using a finite element semiconductor simulator as reported in Chapter 5 though the variation in V_{t1} was significantly less. Also the leakage current of the silicon transistors was found to be significantly less than the leakage measured in the finite element semiconductor simulations. The finite element simulations were modified to reduce the leakage current and the variation in V_{t1} was found to be closer to that measured in silicon. The author believes that, when leakage current is low, the reduction in variation in V_{t1} for different gate lengths is due to triggering voltage being more dependant on the drain-substrate avalanching voltage (which does not vary with gate length) and less dependant on the parasitic bipolar transistor gain (which does vary with gate length).

This is a significant result as it shows that a ggNMOS with a long gate length may be as effective as a ggNMOS with a short gate length. The advantage of a ggNMOS with a long gate length is that the leakage current is lower, this is especially significant at high temperatures.

In this chapter it was also demonstrated that adding holes in the extended drain of a ggNMOS increased the drain resistance in a predictable manner. If the effective width of the extended drain was decreased by 50% the drain resistance doubled. This shows that this methodology can be implemented with ease and it also suggests that the resistance of the extended drain is not influenced by the avalanching drain-substrate junction as reported by Esmark et al. [25] and Liou at al. [148].

Chapter 7

Conclusion

In this conclusion the three key areas of investigation of this thesis will be revisited. The TLP measurements made on the BC01b in Chapter 3 will be reviewed and analysed in detail in light of the lessons learnt from the SPICE and finite element simulations in Chapters 4 and 5, the test chip results from Chapter 6 and the body of knowledge within the literature. Further detailed explanation is given as to why the circuit behaves as it does and suggestions given and explained as to how we can improve future ggNMOS design. Subsequently the usability of the finite element simulation software will be reviewed and conclusions drawn on how such software could be improved.

7.1 The BC01b RF output driver

To assist the reader the ESD and output driver circuit will be reviewed. Figure 7.1 illustrates the schematic of the output stage of the BC01b. The layout of the output stage is in Figures 7.2, 7.3 and 7.4 (poly, active, p^+, contacts, silicide block and relevant metal layers are visible). Figure 7.3 shows detail from 7.2: the largest ESD

protection clamp (ggNMOS1) and the diodes (D1/D2). Figure 7.4 is also detail from 7.2; it illustrates the layout of the secondary ESD protection (ggNMOS2) and the cascoded output driver (NMOS3).

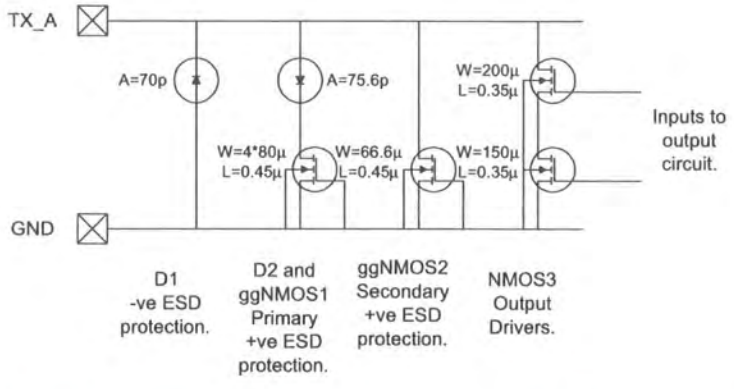

Figure 7.1: Output stage of the BC01b's TX_A identical to TX_B

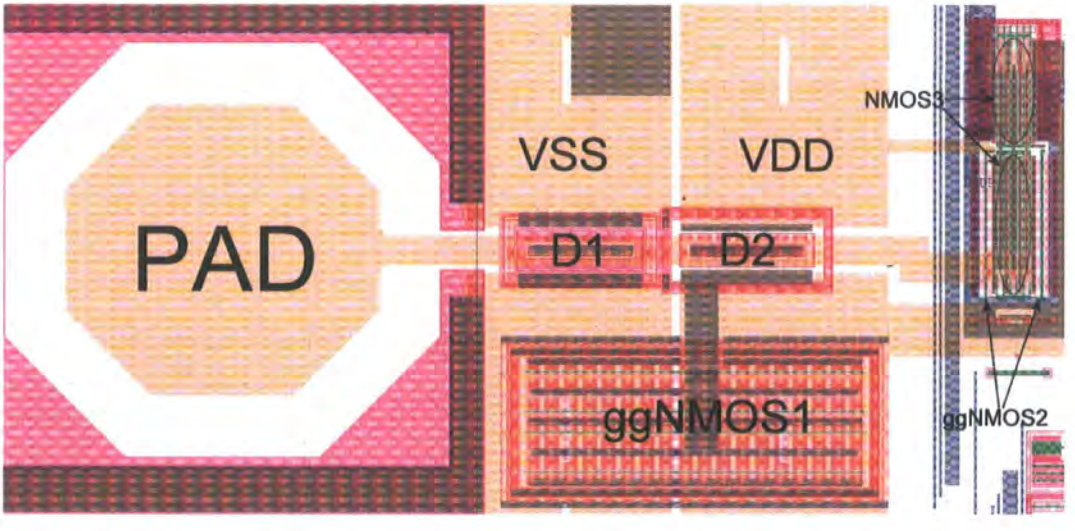

Figure 7.2: Layout of the output stage of the BC01b's TX_A identical to TX_B)

There are a number of features that are important to note. Figure 7.4 shows the layout of the cascoded NMOS output driver. The dimensions in this part of the design are primarily dictated or limited by yield and transistor RF performance considerations during normal operation. Both gate length and gate to contact spacing are at a minimum for this process node ($0.35\mu m$) and the whole transistor is fully silicided. The two transistors that make up the cascoded transistor are located in separate active regions $2.95\mu m$ apart. The layout ggNMOS2 clamps differs significantly from the output driver; the gate length is $0.45\mu m$ and a silicide block layer overlays and

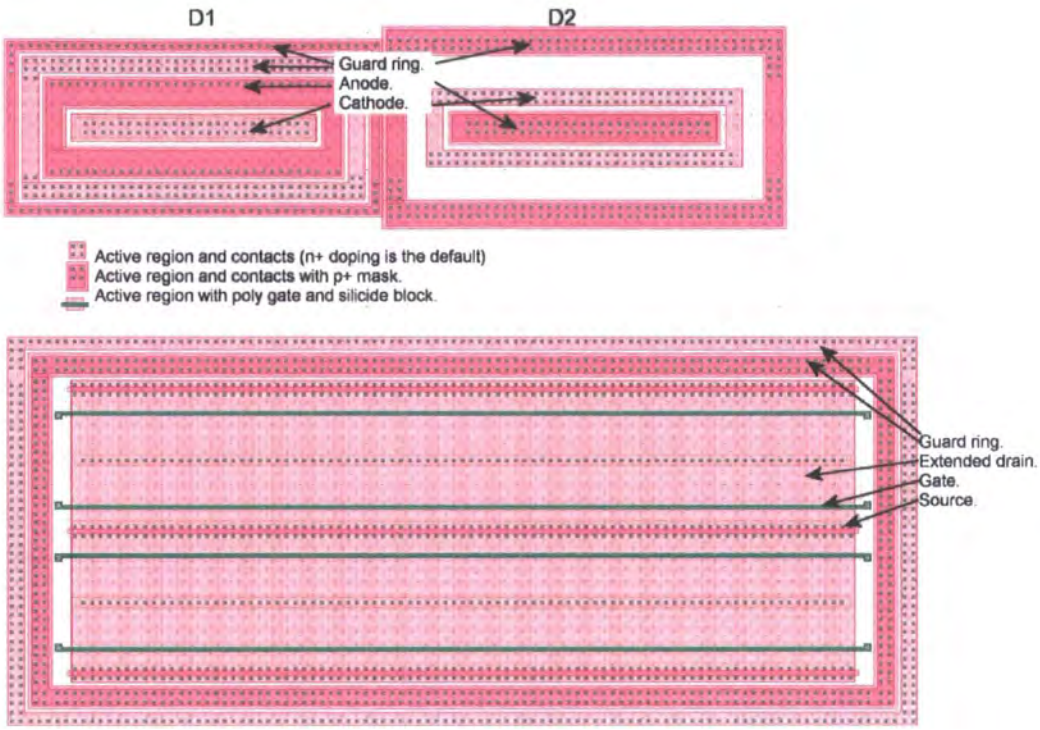

Figure 7.3: Detail from Figure 7.2 showing D1, D2 and ggNMOS1

extends each side of the gate by $1.2\mu m$.

The construction of ggNMOS1 is similar to ggNMOS2s (Figure 7.3), the difference is that the length of the silicide block between the gate and the drain contacts is $4\mu m$ and that the drain contacts do not extend along the whole width of the transistor. The diodes, D1 and D2, are of a single finger construction with dimensions of $25\mu m * 2.8\mu m$ and $27\mu m * 2.8\mu m$ respectively. The reader must also take note of the location and spacing between each of the components.

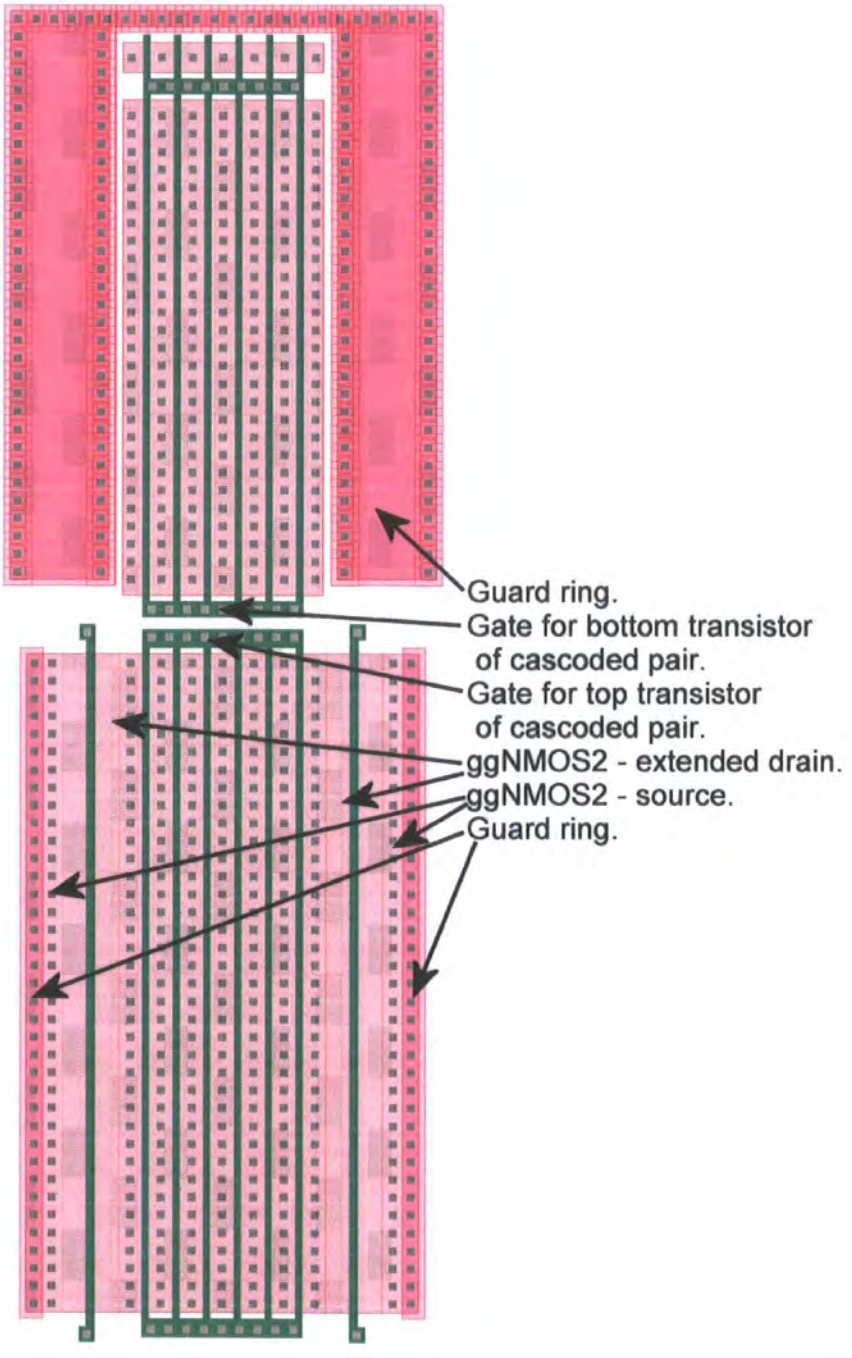

Figure 7.4: Detail from Figure 7.2 showing ggMOS2 ESD protection and ggNMOS3 output drivers

7.2 The effect of rise time on triggering performance

It has been shown many times that as the rise time is decreased, the triggering voltage decreases (See Tables 3.1 and 3.2 for measurements, simulations in Figures 5.22-5.25 and examples in the literature [123][119]). The measurements of the BC01b ggNMOS2 (with ggNMOS1 isolated by the FIB) displayed in Table 3.2, show that as the rise time decreases from $16ns$ to $2.5ns$, the triggering voltage is largely unchanged but, as it is decreased to $1ns$, V_{t1} falls by $\sim 450mV$. The same measurement on the ggNMOS1 with a diode D2 in series produces a similar result except that the decrease in V_{t1}, surprisingly, is much greater at $\sim 1.5V$. The measurements also show that, for a $16ns$ rise time, the ggNMOS2 has a triggering voltage that is $\sim 3.5V$ less than that of the ggNMOS1 and D2. It is therefore no surprise that for low energy TLP pulses, at around the triggering voltage, ggNMOS2 triggers in preference to ggNMOS1 (Figures 3.17-3.19).

Up until now the effect of the diode D2 has been neglected though, given the assumption that ggNMOS1 and ggNMOS2 should have the same triggering voltage, (the only difference between the two devices is the extended drain length and the transistor width, neither have yet been shown to have a significant effect on triggering voltage) it is surprising that for a $16ns$ rise time the difference in triggering voltage between the two types of transistor is $\sim 3.5V$. The figure of $\sim 3.5V$ is far in excess of the normal forward voltages of diodes suggesting that significant current is required to trigger a ggNMOS; TLP measurements of similar diodes (very similar construction but different process node) with an area of $138pm^2$, found that 2.1A is needed to achieve a forward voltage of $3.5V$. The equivalent current for a $75pm^2$ diode is $1.15A$. This current seems excessively high given that a 50Ω system charged to $12V$ can only provide $240mA$ of current during the TLP test. Probably

the method used for deriving the current is flawed either because IV characteristics for diodes do not correlate or because the triggering voltage for the ggNMOS1 is greater than that for ggNMOS2. ggNMOS1 has a longer silicide blocked drain than ggNMOS2 and the simulation results presented in Figures 5.34-5.35 have provided evidence that the triggering voltage will be marginally higher for DC simulations due to the extra resistance of the extended drain. This effect may be more pronounced at this process node.

Despite the fact that the ggNMOS1 structure in series with D2 has a higher triggering voltage than ggNMOS2 for fast rise times, Tables 3.1 and 3.2 show that their triggering voltage starts to converge as the rise times get faster. This is likely to be due to the capacitance across D2 effectively decreasing the forward voltage across the device thus decreasing the effective triggering voltage of the ggNMOS1 in series with D2.

It is surprising however that, for higher energy TLP pulses, the ggNMOS1 is triggered in preference to ggNMOS2 which suggests that the triggering voltage of ggNMOS1 in series with D2, falls below that of ggNMOS2. This can be inferred from the IV curves in Figures 3.17-3.19. On a simple level this can be attributed to the fast dV/dt that occurs in the rising edge of high energy TLP pulses causing a high current to cross D2 without building up a significant voltage. It must be pointed out in this respect that even if the voltage across the diode were negligible and the triggering voltage of the two devices were equal, we would still find that both devices would have an equal chance of triggering. However the graph clearly shows that ggNMOS1 is favored over ggNMOS2.

ggNMOS1 and ggNMOS2 are both the same distance from the pad and the major differences in their design is limited to device width and length of silicide blocked drain extension. It is possible that the length of the extended drain affects the triggering voltage by increasing the drain-substrate capacitance thus facilitating

the triggering mechanism for fast rise times and high dV/dt TLP pulses. If this is the case, then the extended drain will have been shown to promote uniform triggering by another mechanism. Currently, the resistance of the extended drain is relied upon to spread the current flow along the whole width of the device during snapback; the silicide block layer therefore is required on the extended drain to increase the resistance. The disadvantage of the silicide block layer is that it increases the transistor on-resistance and decreases the Q-factor of any RF output due to the increased ESR on the parasitic capacitance. This capacitively coupled triggering mechanism may require more area than the current spreading mechanism that relies on the drain resistance. In addition it may also result in a more acceptable parasitic load on any RF signal, decrease the transistor on-resistance and require no silicide block processing for manufacture.

Further work should be carried out to determine whether a fully silicided extended drain can be used to capacitively trigger the whole width of a ggNMOS. This solution may not work well in multi fingered devices as once one finger has triggered it will reduce the dV/dt across all the neighbouring fingers thus preventing them from triggering.

7.3 The effect of transistor proximity on triggering performance

In this section the IV characteristics of ggNMOS1 (Figure 3.17) will be compared to those of ggNMOS2 (Figures 3.20-3.22). The yellow line in Figure 3.17 is the IV curve of ggNMOS1 measured in isolation for currents up to $2.5A$. Once the device is triggered, the voltage across the device increases steadily from $\sim 6V$ to $\sim 13V$ as the TLP current is increased. This is in sharp contrast to the magenta line in Figures 3.20-3.22 which shows the IV characteristics of ggNMOS2 for currents

up to $0.7A$. In this case, once the device is triggered, V_{DS} increases steadily from $\sim 6V$ to $\sim 7V$ as the TLP current increases. Once the TLP current reaches $\sim 300mA - 350mA$, the voltage across the device decreases sharply to the voltage measured when the test current was only $\sim 150mA - 175mA$ suggesting that both transistors are now conducting. As the current is further increased, V_{DS} across the device resumes increasing steadily up to $\sim 8V$. At that point the TLP current reaches $\sim 700mA$ and leakage across the device increases sharply indicating device failure. The IV curve of the ggNMOS2 clearly shows that for low TLP energies, only one ggNMOS finger is triggered but, if the TLP energy is high enough, then both transistors are triggered. This effect has not been observed for ggNMOS1.

The reader should note from Figures 7.3 and 7.4 that the individual transistors forming ggNMOS1 are all located in close proximity whereas the transistors forming ggNMOS2 are $\sim 12\mu m$ apart. This result shows that placing ggNMOSs in close proximity will assist the spreading of the snapback effect. The increase in bulk voltage once a device is triggered will assist the triggering of neighbouring devices thus increasing the ESD current handling capability of the protection structures.

This triggering proximity effect is likely to trigger any adjacent transistor with sufficient V_{DS} into the snapback mode by increasing the potential of the substrate thus forward biasing the base-emitter bipolar junction. Such an effect has not only been demonstrated in this thesis but a similar effect has also been demonstrated by Ker et al. [45] and Texas Instruments [83] through the use of substrate pumping to promote triggering. The advantage of this effect has already been explored. It has to be pointed out that if a transistor that does not have an extended drain (thus is not tolerant of ESD current passing through it), is in the same well and in close proximity to the avalanching transistor, it is likely that it will also trigger resulting in catastrophic failure of the transistor. During an ESD event not only do we want the ESD clamps to conduct the charge but we also need to be sure that

the charge does not take any other unsafe discharge path, for example through an output driver transistor connected between the output and ground. We therefore need to be sure that any output drivers are placed some distance ($\sim 5\mu m$) away from any ggNMOS devices. In section 3.6 it was shown that the ESD damage was likely to be in the ggNMOS2 secondary ESD protection. The layout in Figure 7.4 shows that ggNMOS2 is in the same well and in close proximity to the top transistor of the NMOS3 output driver. It is likely that the increase in bulk potential caused by ggNMOS2, will have induced snapback in the top transistor of NMOS3 cascoded pair. Fortunately the current was blocked by the bottom transistor located in a separate well $\sim 3\mu m$ away thus preventing catastrophic current flow through the output driver.

To use this mechanism to its fullest extent the ggNMOS clamp should be made of a single finger thus effectively reducing the spacing between any adjacent fingers to zero. Single finger devices have already been shown to have a more consistent ESD performance over a range of t_R when compared to multi finger devices [128] though the author of this thesis has not seen any evidence to suggest that this has been investigated further. There are a few problems with single finger devices; firstly their width ($\sim 300\mu m$) makes their layout in practice difficult and secondly the voltage of the gate in the center of the device may vary considerably during the ESD event due to the resistivity and length of the gate thus this can cause problems of its own. A possible layout is shown in Figure 7.5.

Both of these methodologies (keeping transistors that you do want to trigger in close proximity, preferably have a single finger, and those that you do not far away as well as the extended drain to promote triggering) can also be used to make area efficient output drivers. Typically, an output driver can be used as a ggNMOS so long as the device width is sufficient though this does mean that the whole device needs to have an extended drain. This is not an efficient use of silicon area. It

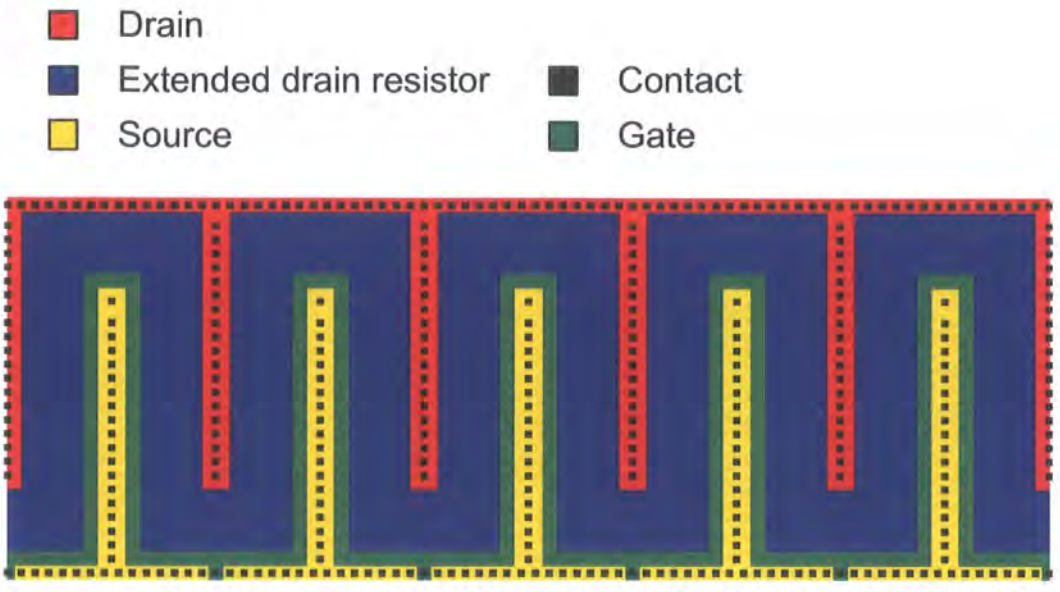

Figure 7.5: Suggested layout for single finger construction ggNMOS with multiple gate taps and efficient silicon area usage

may be possible that if the output driver was divided into two sections which are connected in parallel a more efficient output driver could be created. The first section of this hybrid output would have a drain extension and sufficient width to dissipate the ESD energy and the second section would have no drain extension but would provide the remainder of the output driver capability. The second section would play no part in the conduction of the ESD current therefore we have to be sure that it does not trigger during the ESD event. To ensure that this does not happen the two transistors would have to be separated by a strong guard ring to prevent the first transistor from triggering the second through the substrate. The lack of extended drain in the second transistor would further inhibit trigging.

In this thesis, a greater understanding of the snapback mechanism has been gained and this knowledge has been used to provide detailed analysis of transistor operation under ESD conditions. The use of effects that promote snapback (such as the extended drain-bulk capacitance, substrate pumping, weak substrate guarding and gate biasing) and effects that promote the spreading of snapback (triggering proximity effect and extended drain resistance) as well as techniques that impede

snapback triggering or spreading, can be used to design area efficient circuits with lower parasitic loading. These techniques can be applied deliberately to optimise what are essentially hybrid circuits that need to have good ESD performance as well as good RF output capability and provide plenty of scope for further research.

7.4 Use of finite element simulation to explore ESD effects in silicon

An in depth understanding of how the snapback process occurs allows us to optimise further the ggNMOS's design leading to more efficient and effective solutions. A considerable amount of insight of the snapback process was derived from the finite element simulations, though this was somewhat restricted by the implementation of the software used. The designers of both Taurus and Sequoia Finite Element Software were primarily interested in allowing the user to simulate transistors with relatively standard proportions under limited conditions. The significant advantage of simulations is that we can try out blue sky ideas that may not be possible to measure or manufacture in silicon. Simulating novel semiconductor structures was difficult and simulating portions of a structure, such as a drain-bulk diode in isolation, was even more so. Many of these problems were caused by preconceptions of the software designer on its use. More flexibility would be very helpful. In some cases experiments were attempted that are impossible to carry out in silicon, such as injecting carriers into a specific point in the structure to try to gain an insight into the operation of the device. Such simulations failed. It may be the case than an open source simulator is the only route to the flexibility desired. The problems encountered can be summarised twofold. In some cases portions of the source code ware modified to allow the user to produce arbitrary structures. This was a tricky process given that these sections of code were not designed to be modified by the

user, and was strongly discouraged by the software supplier. However, such an approach did ultimately work. The other problem was with mathematical convergence. The author believes it may be related to the modelling surrounding the contacts. It was not possible to ascertain how the contacts were modelled but it seemed that the problem was related to the fact that semiconductor material at either end of a pn junction, will have a different potential due to the doping causing shifts in the fermi levels. In the author's opinion the simulator has difficulty in calculating the correct potential for the contacts or the correct work function difference between the contacts and the semiconductor material that the contact is connected to. These kind of problems illustrate the need for an open source Finite Element Simulator that is considerably more flexible in its implementation.

7.5 Conclusion

In this thesis a detailed description has been given of how the RF output drivers of the BC01b operate under ESD conditions. The Transmission Line Pulse (TLP) tester has been shown to be a very effective tool for the analysis of the operation of a ESD protection circuit under conditions that are similar to an ESD event. The FIB enhances the usefulness of the TLP tester by isolating individual components. Further detailed understanding on the operation of the ggNMOS is gained through electrothermal simulation and measurements of test structures in a $0.18\mu m$ test chip. This information coupled with the current body of knowledge in the literature is used to come to a detailed understanding on how the ggNMOS ESD clamp operates under ESD conditions. This understanding is used to determine how the ggNMOS transistors operate and interact with each other in the BC01b RF output driver circuit and therefore show why the circuit behaves as it does. This knowledge is also used to suggest how the triggering homogeneity of a ggNMOS could be improved as well as how to prevent triggering in circuits where it must not happen. These

techniques are used to demonstrate the design of more efficient (both in area and parasitic capacitance) ggNMOS circuits.

Finally the use of finite element simulation software has been assessed and the author came to the conclusion that though such tools are very useful there is considerable scope for improvement in the implementation of the tool. It would be desirable to have a tool that is more flexible in its implementation such that arbitrary structures can be simulated and more transparent in its operation such that the user can understand how the simulator is operating.

Bibliography

[1] D. G. Pierce, "Electrostatic discharge (ESD) failure mechanisms," www.ESDA.org, September 2001, Tutorial from the ESDA 2001 conference.

[2] K. Verhaege and C. Russ, "Wafer cost reduction through design of high performance fully silicided ESD devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Sarnoff Corporation, 2000.

[3] B. Keppens, M. Mrgens, J. Armer, P. Jozwiak, G. Taylor, R. Mohn, C. Trinh, C. Russ, K. Verhaege, and F. de Ranter, "Active-area-segmentation (AAS) technique for compact, ESD robust, fully silicided NMOS design," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Sarnoff Corporation, 2003.

[4] S. Reyntjens and R. Puers, "A review of focused ion beam applications in microsystem technology," *Journal of Micromechanics and Micro Engineering.*, vol. 11, pp. 287–300, January 2001.

[5] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. Wiley-Interscience, 1981.

[6] JEDEC, *Electrostatic Discharge (ESD)Sensitivity Testing Human Body Model (HBM) JESD22-A114-B*, JEDEC Solid State Technology Association.

[7] C. Duvvury, "Quote: Many of the RF devices in market do not have any ESD protection incorporated," Workshop on Silicon Technology Scaling and ESD Reliability - Roadmap and Reality, EOS/ESD Symposium, 2004.

[8] A. Amerasekera, "Quote: Lots of RF chips are being commercialized with no ESD protection on the RF pins," Tutorial on RF ESD protection, EOS/ESD Symposium, 1999.

[9] B. Unger, "ESD basics tutorial." www.ESDA.org, September 2001, Tutorial from the ESDA 2001 conference.

[10] D. Pierce and D. Durgin, "An overview of electrical overstress effects on semiconductor devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-3. Booz-Allen & Hamilton Inc., 1981.

[11] C. Duvvury, "ESD on-chip protection in advanced technologies," www.ESDA.org, September 2002, Tutorial from the ESDA 2002 conference.

[12] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits*, 2nd ed. John Wiley & Sons, 1996.

[13] Electrostatic Discharge Association. [Online]. Available: http://www.esda.org

[14] T. Green, "A review of EOS/ESD filed failures in military equipment," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-10, 1988.

[15] R. G. Wagner, J. Soden, and C. F. Hawkins, "Extent and cost of EOS/ESD damage in an IC manufacturing process," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-15, 1993.

[16] R. Merrill and E. Issaq, "ESD design methodology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-15, 1993.

[17] Semiconductor Industiry Association. (2002, Feb.) Summary of the international technology roadmap for semiconductors. [Online]. Available: http://www.sia-online.org/downloads/itrs_2001.pdf

[18] S. Wolf, *Silicon Processing for the VLSI Era*, 4th ed. Lattice Press, 2002.

[19] G. Groeseneken, "Invited paper: Hot carrier degradation and ESD in submicron CMOS technologies: How do they interact?" in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. IMEC, Belgium, 2000.

[20] K.-L. Chen, "Effects of interconnect process and snapback voltage on the ESD failure thresholds on NMOS transistors." in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-10. Texas Instruments, Inc., 1988.

[21] A. Ochoa Jr., F. Sexton, T.F.Wrobel, C. Wash, and R. Sokel, "Snap back: A stable regenerative breakdown mode of MOS devices," *IEEE Transactions on Nuclear Science*, vol. NS-30 N. 6, pp. 4127–4130, December 1983.

[22] P. Gaitonde, S. Gaul, T. Crandell, and S. Earles, "Predictive simulation to improve reliability of a snapback-based NMOS clamp," *Integrated Reliability Workshop Final Report, 2005 IEEE International*, pp. 3 pp.–, 17-20 Oct. 2005.

[23] S. M. Sze and G. Gibbons, "Avalanche breakdown voltages of abrupt and linearly graded p-n junctions in Ge, Si, GaAs, and GaP," in *Applied Physics Letters.* American Institute of Physics, 1966.

[24] M.-D. Ker and K.-H. Lin, "Double snapback characteristics in high-voltage nMOSFETs and the impact to on-chip ESD protection design," *Electron Device Letters, IEEE*, vol. 25, no. 9, pp. 640–642, Sept. 2004.

[25] K. Esmark, W. Stadler, M. Wendel, H. Gobner, H. Gunggenmos, and W. Fichtner, "Advanced 2D-3D ESD device simulation a powerful tool already

used in a pre-Si phase," *Micro Electronics Reliability*, vol. 41, pp. 1761–1770, Feburary 2001.

[26] M. Litzenberger, C. Furbock, D. Pogany, E. Gornik, K. Esmark, and H. Gossner, "Investigation of 3D Phenomena in the Triggering of gg-nMOS Electrostatic Discharge Protection Devices," *Solid-State Device Research Conference, 2000. Proceeding of the 30th European*, pp. 520–523, 11-13 September 2000.

[27] M. Litzenberger, R. Pichler, D. Pogany, E. Gornik, K. Esmark, and H. Gossner, "Influence of Layout Parameters on Triggering Behaviour in 0.35um and 0.18um Process gg-nMOS ESD Protection Devices," *Solid-State Device Research Conference, 2001. Proceeding of the 31st European*, pp. 335–338, 11-13 September 2001.

[28] T. J. Maloney, "Designing power supply clamps for electrostatic discharge proteciton of integrated circuits." in *Microelectronics Reliability.* Pergamon, 1998.

[29] J. W. Miller, "Applications and process dependent ESD design strategy," www.ESDA.org, September 2001, Tutorial from the ESDA 2001 conference.

[30] T. Wang and M.-D. Ker, "On-chip ESD protection design by using polysilicon diodes in CMOS technology for smart card applications," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Sunplus Technology; National Chiao-Tung University, 2000.

[31] E. Worley, A. Salem, and Y. Sittampalam, "High current characteristics of devices in a $0.18\mu m$ CMOS technology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Conexant Systems, 2000.

[32] R. Velghe, P. de Vreede, and P.H.Woerlee, "Diode network used as ESD protection in RF applications." Philips, Tech. Rep., 2002.

[33] R. Velghe, P. de Vreede, and P. Woerlee, "Diode network used as ESD protection in RF applications," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Philips Semiconductors Nijmegen; Philip Research Laboratories, Eindhoven, 2001.

[34] P. Leroux, M. Steyaert, K. Leuven, V. Vassilev, H. Maes, and K. Leuven, "A 6mW, 1.5dB NF CMOS LNA for GPS with 3KV HBM ESD-protection," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. IMEC, Belgium, 2002.

[35] S. Hyvonen and E. Rosenbaum, "Diode-based tuned ESD protection for 5.25-GHz CMOS LNAs," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-27. Department of Electrical and Computer Engineering University of Illinois at Urbana-Champaign, 2005.

[36] V. Vassilev, G. Groeseneken, S. Jenei, H. Maes, R. Venegas, and M. Steyaert, "Modelling and extraction of RF performance parameters of CMOS electrostatic discharge protection devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. IMEC, KUL/ESAT, 2002.

[37] C. Richier, P. Salome, G. Mabboux, I. Zaza, A. Juge, and P. Mortini, "Investigation on different ESD protection strategies devoted to 3.3V RF applications (2GHz) in a 0.18μm CMOS process," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. STMicroelectronics, 2000.

[38] E. Worley and A. Bakulin, "Optimization of input protection diode for high speed applications," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Conexant Systems, 2002.

[39] T. Smedes, J. van Zwol, P. Jong, and A. Heringa, "The impact of substrate resistivity on ESD protection devices," in *Electrical Overstress / Electro-*

static Discharge Symposium Proceedings, vol. EOS-24. Philips Semiconductors/Philips Research Laboratories, 2002.

[40] K. Verhaege, "ESD Protection Design for RF and Advanced Technology Applications," www.ESDA.org, September 2002, Tutorial from the ESDA 2002 conference.

[41] T. J. Maloney and S. Dabral, "Novel Clamp Circuits for IC Power Supply Protection." in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-17. Intel., 1995.

[42] M.-D. Ker and W.-L. Wu, "ESD Protection Design with the Low-Leakage-Current Diode String for RF Circuits in BiCMOS SiGe Process," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-27. Nanoelectronics and Gigascale Systems Laboratory Institute of Electronics, National Chiao-Tung University, Hsinchu, Taiwan, 2005.

[43] S. Dabral, R. Aslett, and T. Maloney, "Designing on-chip power supply coupling diodes for ESD protection and noise immunity," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-15. Intel, 1993.

[44] T. Maloney and S. Dabral, "Novel clamp circuits for IC power supply protection," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-15. Intel, 1993.

[45] M.-D. Ker, T.-Y. Chen, and C.-Y. Chang, "ESD protection design for CMOS RF integrated circuits," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. National Chiao-Tung University; Industrial Technology Research Institute, 2001.

[46] R. van Dalen, G. Hurkx, E. Hijzen, P. Weijs, and A. den Dekker, "Using thin emitters to control BV ceo effects in punch-through diodes for ESD protection," in *Electrical Overstress / Electrostatic Discharge Symposium Pro-*

ceedings, vol. EOS-23. M.A.A. in't Zandt, Philips Research Leuven, Philip Semiconductors, 2001.

[47] C.C.Russ, M. Mergens, J. Armer, P. Jozwiak, G. Kolluri, and L. Avery, "GGSCRs: GGNMOS triggered silicon controlled rectifiers for ESD protection in deep sub-micron CMOS processes," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Sarnoff Corporation; K.G. Verhaege, Sarnoff Europe, 2001.

[48] M. Mergens, C. Russ, J. Armer, P. Jozwiak, R. Mohn, and K. Verhaege, "High holding current SCRs (HHI-SCR) for ESD protection and latch-up immune IC operation," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Sarnoff Corporation, Sarnoff Europe, 2002.

[49] S. Joshi, E. Rosenbaum, and R. Ida, "A study of vertical SiGe thyristor design and optimisation," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. University of Illinois at Urbana-Champaign, Motorola, Inc., 2003.

[50] Y. Morishita, "New ESD protection circuits based on PNP triggering SCR for advanced CMOS device applications," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. NEC Corporation, 2002.

[51] B. Caillard, F. Azais, P. Nouet, S. Dournelle, and P. Salome, "STMSCR: A new multi-finger SCR-based protection structure against ESD," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. University Montpellier II/CNRS, STMicroelectronics, 2003.

[52] K. Kunz, C. Duvvury, and H. Schichijo, "5V tolerant fail-safe ESD solutions for a $0.18\mu m$ logic CMOS process," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Texas Instruments, Inc., 2001.

[53] C. Duvvury, R. Rountree, and O. Adams, "Internal chip ESD phenomena beyond the protection circuit." in *Reliability Physics Symposium 1988. 26th Annual Proceedings., International.* Texas Instruments., 1998.

[54] H. G. Feng, K. Gong, and A. Z. Wang, "A novel on-chip electrostatic discharge protection design for RF ICs," *Micro Electronics*, vol. 32, pp. 189–195, October 2000.

[55] S. Voldman, A. Botula, D. Hui, and P. Juliano, "Silicon germanium heterojunction bipolar transistor ESD power clamps and the Johnson limit," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. IBM Communications Research development Center; IBM Corporation; University of Illinois at Urbana-Campaign, 2001.

[56] Y. Ma and G. Li, "A novel on-chip ESD protection circuit for GaAs HTB RF power amplifiers," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. University of California, 2002.

[57] M. Yintat and G. Li, "InGaP/GaAs HBT DC-20 GHz distributed amplifier with compact ESD protection circuits," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-26. Integrated Nanosystems Research Facility, Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697, United States, 2004.

[58] C. Duvvury, C. Diaz, and T. Haddock, "Achieving uniform nMOS device power distribution for sub-micron ESD reliability," *Electron Devices Meeting, 1992. Technical Digest., International*, pp. 131–134, 13-16 Dec 1992.

[59] M. Mergens, K. Verhaege, C. Russ, P. Jozwiak, G. Kolluri, and L. Avery, "Multi-finger turn-on circuits and design techniques for enhanced ESD performance and width scaling," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Sarnoff Corporation, 2001.

[60] M. Ker, C.-H. Chuang, and H.-C. Jiang, "ESD protection design for mixed-voltage I/O buffer by using stacked-NMOS triggered SCR device," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. National Chiao-Tung University; Industrial Technology Research Institute, 2001.

[61] H. Koizumi, Y. Komine, Y. Ojtomo, and M. Shimaya, "ESD protection in fully-depleted CMOS/SIMOX with a tungsten-clad source/drain," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. NTT Telecommunications Energy Laboratories, 2000.

[62] J. Smith and G. Boselli, "A MOSFET power supply clamp with feedback enhanced triggering for ESD protection in advanced CMOS technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Texas Instruments, Inc., 2003.

[63] J. Li, E. Rosenbaum, and R. Gauthier, "Compact, Timed-Shutoff, MOSFET-Based Power Clamp for On-Chip ESD Protection," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-26. Department of Electrical and Computer Engineering University of Illinois at Urbana-Champaign, 2004.

[64] J. Li, R. Gauthier, S. Mitra, C. Putnam, K. Chatty, R. Halbach, and C. Seguin, "Design and Characterization of a Multi-RC-Triggered MOSFET-based Power Clamp for On-Chip ESD Protection," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-28. IBM Semiconductor Research and Development Center, IBM Systems and Technology Group, Essex Junction, VT 05452 USA, 2006.

[65] T. Maloney, S. Poon, and L. Clark, "Methods for designing low-leakage power supply clamps," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Intel Corporation, 2003.

[66] S. Poon and T. Maloney, "New considerations for MOSFET," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Intel Corporation, 2002.

[67] C.A.Torres, J. Miller, M. Stockinger, M.D.Akers, M. Khazhinsky, and J. Weldon, "Modular, portable, and easily simulated ESD protection networks for advanced CMOS technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Motorola Inc., 2001.

[68] M. Stockinger, J. W. Miller, M. G.Khazhinsky, C. Torres, J. Weldon, B. Preble, M.J.Bayer, M. Akers, and V. Kamat, "Boosted and distributed rail clamp networks for ESD protection in advanced CMOS," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Motorola, Inc., 2003.

[69] C. Salling, J. Hu, J. Wu, C. Duvvury, R. Cline, and R. Pok, "Development of substrate-pumped nMOS/Lnpn protection for a $0.13\mu m$ technology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Texas Instruments, Inc., 2001.

[70] S. Hyvonen, S. Joschi, and E. Rosenbaum, "Comprehensive ESD Protection for RF Inputs," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. University of Illinois at Urbana-Champaign, 2003.

[71] V.Vassiliev, G. Groeseneken, H. Maes, S. Thijs, P. Segura, P.Wambacq, M. Natarjan, P.Leroux, and M. Steyaert, "Co-design methodology to provide high ESD protection levels in the advanced RF circuits," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. IMEC and KUL, ESAT, 2003.

[72] S. Thijs, D. Linten, N. Mahadeva Iyer, W. Jeamsaksiri, A. Mercha, J. Ramos, X. Sun, G. Carchon, P. Soussan, T. Nakaie, M. Sawada, T. Hasebe,

P. Wambacq, S. Decoutere, and G. Groeseneken, "Class 3 HBM and class M4 MM ESD protected 5.5 GHz LNA in 90 nm RFCMOS using above-IC inductor." in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-27. IMEC, Belgium, 2005.

[73] S. Sherry, P.-Y. Tan, E.-C. Chua, S.-C. T. Carol, I. Manna, S. Redkar, and S. Ansari, "Copper interconnect microanalysis and electromigration reliability performance due to the impact of TLP ESD," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Chartered Semiconductor Mfg. Ltd., 2002.

[74] T. Smedes and Y. Li, "ESD phenomena in interconnect structures," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Philip Semiconductors, 2003.

[75] P. Y. Tan, I. Manna, Y. C. Tan, K. F. Lo, and P. H. Li, "A study of high current characteristics of devices in a $0.13\mu m$ CMOS technology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Chartered Semiconductors Mfg. Ltd., 2002.

[76] D. G. Pierce, "Electro-thermomigration as an electrical overstress failure mechanism," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-7. Sandia National Laboratories, 1985.

[77] C. Ito, K. Banerjee, and R.W.Dutton, "Analysis and optimization of distributed ESD protection circuits for high-speed mixed-signal and RF applications," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Stanford University, 2001.

[78] M. D. Ker and B. J. Kuo, "Optimization of broadband RF performance and ESD robustness by π-model distributed ESD protection scheme," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-26.

Nanoelectronics and Gigascale Systems Laboratory, Institute of Electronics, National Chiao-Tung University, Hsinchu 30050, Taiwan, 2004.

[79] K. Esmark, W. Stadler, M. Wendel, H. Grossner, X. Guggenmos, and W. Fichtner, "Advanced 2D/3D device simulation - a powerful tool already used in a pre-Si phase," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Infineon Technology AG; ETH Zurich, 2000.

[80] M. B. Heer, S. Dubec, V. Pogany, D. Gornik, E. Dissegna, M. Cerati, L. Zullino, L. Andreini, and A. A. Tazzoli, "Analysis of the Triggering Behaviour of Low Voltage BCD Single and Multi-Finger gc-NMOS ESD Protection Devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-28. Institute for solid state electronics, Vienna University of Technology, Floragasse 7, A-1040 Vienna, Austria, 2006.

[81] C. Musshoff, H. Wolf, H. Giesew, P. Eggew, and X. Guggenmos, "Risetime effects of HBM and square pulses on the failure thresholds of GGNMOS transistors," *Reliability of Electron Devices, Failure Physics and Analysis, 1996. Proceedings of the 7th European Symposium on*, pp. 1743–1746, 1996.

[82] M. Litzenberger, R. Pichler, S. Bychikhin, D. Pogany, E. Gornik, K. Esmark, and H. Gossner, "Effect of pulse risetime on trigger homogeneity in single finger grounded gate nMOSFET electrostatic discharge protection devices," *Reliability of Electron Devices, Failure Physics and Analysis, 2001. Volume 41, Number 9*, pp. 1385–1390, 2001.

[83] C. Duvvury, S. Ramaswamy, A. Amerasekera, R. Cline, B. Andresen, and V. Gupta, "Substrate pump NMOS for ESD protection applications," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Texas Instruments, 2000.

[84] K. Chatty, D. Alvarez, R. Gauthier, C. Russ, M. Abou-Khalil, and B. J. Kwon, "Process and design optimization of a protection scheme based on NMOSFETs with ESD implant in 65nm and 45nm CMOS technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-29. IBM Semiconductor Research and Development Center, IBM Systems and Technology Group, Essex Junction, VT 05452 USA, 2007.

[85] T. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicised nMOS output transistors by insuring uniform current flow," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-11. Semiconductor Process and Design Center, Texas Instruments, 1989.

[86] A. Z. Wang, C. Tsay, A. Lele, and P. Deane, "A study of NMOS behavior under ESD stress simulation and characterization," *Micro Electronics Reliability*, vol. 38, pp. 1183–1186, 1998.

[87] J. Miller, M.Khashinsky, and J. Weldon, "Engineering the cascoded NMOS output buffer for maximum V_{t1}," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Motorola, Inc., 2000.

[88] R. Gauthier, M. Muhammad, C. Putnam, W. Stadler, K. Esmark, P. Riess, and A. Salman, "Evaluation of diode-based and NMOS/lnpn-based ESD protection strategies in a triple gate oxide thickness $0.13\mu m$ CMOS logic technology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. IBM Microelectronics Semiconductor Research and Deveopment Center; Infineon Technologies AG, George Mason University, 2001.

[89] V. Gupta, A. Amerasekera, S. Ramaswamy, and A. Tsao, "ESD-related process effect in mixed-voltage sub $0.5\mu m$ technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-20. Texas Instruments, Inc., 1998.

[90] R. M. Steinhoff, J.-B. Huang, P. Hower, and J. Brodsky, "Current filament movement and silicon melting in an ESD-robust DENMOS transistor," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Texas Instruments, Inc., 2003.

[91] S. Bychikhin, V. Dubec, M. Litzenberger, D. Pogany, E. Gornik, G. Gross, K. Esmark, M. Stecher, W. Stadler, H. Gieser, and H. Wolf, "Investigation of ESD protection elements under high current stress in CDM-like time domain using backside laser interferometry," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Vienna University of Technology/Infineon Technologies/ Frauenhofer Institute, 2002.

[92] C.-S. Kim, H.-B. Park, Y.-G. Kang, M.-G. Lee, S.-W. Lee, C.-H. Jeon, H.-G. Kim, Y.-J. Yoo, and H.-S. Yoon, "A novel MOS transistor for high performance ESD protection devices in a $0.18\mu m$ CMOS technology utilizing silicide process," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Hyundai Electronics industries Co., Ltd, 2000.

[93] H. Wolf, H. Gieser, and W. Wilkening, "Analyzing the switching behavior of ESD-protection transistors by very fast transmission line pulsing," *Journal of Electrostatics*, vol. 49, pp. 111–127, 2000.

[94] M.-D. Ker, S.-H. Chen, and C.-H. Chuang, "ESD failure mechanisms of analog I/O cells in 0.18um CMOS technology," *Device and Materials Reliability, IEEE Transactions on*, vol. 6, no. 1, pp. 102–111, March 2006.

[95] J. Z. Chen, A. Amerasekera, and C. Duvvury, "Design methodology for optimizing gate driven ESD protection circuits in submicron in CMOS processes," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-19. Texas Instruments, Inc., 1997.

[96] K. Bock, C. Russ, G. Badenes, G. Groeseneken, and L. Deferm, "Influence of well profile and gate length on the ESD performance of a fully silicided $0.25\mu m$ CMOS technology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-19. IMEC, Belgium., 1997.

[97] G. Krieger, "The dynamics of electrostatic discharge prior to bipolar action related snap-back," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-11. VLSI Technology Inc., 1989.

[98] Y. Fong and C. Hu, "The effect of high electric field transients on thin gate oxide MOSFETs," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-9. Department of Electrical Engineering and Computer Sciences, University of California, Berkly., 1987.

[99] G. Boselli, C. Duvvury, and V. Reddy, "Efficient pnp characteristics of pMOS transistor in sub-$0.13\mu m$ ESD protection," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Texas Instruments, Inc., 2002.

[100] EIA and JEDEC, *Electrostatic Discharge (ESD)Sensitivity Testing Machine Model (MM) EIA/JESD22-A115-A*, Electronic Industries Association.

[101] Department of Defense, *Mil Spec 883 Method 1315*, Department of Defense.

[102] S. Voldman, A. V. Laecke, J. Rascoe, L. Lanzerotti, D. Sheridan, B. Ronan, and S. Ames, "Test methods, test techniques and failure criteria for evaluation of ESD degradation of analog and radio frequency (RF) technology," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. IBM Microelectronics Division, Princeton University, 2002.

[103] S. Hyvonen, S. Joshi, and E. Rosenbaum, "Combined TLP/RF Testing System for Detection of ESD Failures in RF Circuits," in *Electrical Overstress /*

Electrostatic Discharge Symposium Proceedings, vol. EOS-25. University of Illinois at Urbana-Champaign, 2003.

[104] Electrostatic Discharge Association, *ESD ANSI EOS/ESD-S5.1-1993*, Electrostatic Discharge Association.

[105] H. Hyatt, "Critical Considerations for ESD Testing." in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-6. Experimental Physics Corp, 1984.

[106] Electronic Industries Association of Japan. [Online]. Available: http://www.jeita.or.jp/eiaj/

[107] M. Chanie, "Device testing, why two different models?" www.ESDA.org, September 2001, Tutorial from the ESDA 2001 conference.

[108] D. Pommerenke and M. Aidam, "ESD: waveform calculation, field and current of human and simulator ESD," in *Microelectronics Reliability.* Pergamon, 1996.

[109] M. Streibl, K. Esmark, A. Sieck, W. Stadler, M. Wendel, J. Szatowski, and H. Gossner, "Harnessing the base-pushout effect for ESD protection in bipolar and BiCMOS technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Infineon Technologies, 2002.

[110] J. Barth, J. Richner, K. Verhaegge, M. Kelly, and L. Henry, "Correlation considerations II: Real HBM to HBM testers," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Barth Elecronics, Sarnoff Europe, Delphi Dalco Electronics,Systems, Ion Systems, 2002.

[111] R. Consiglio and I. Morgan, "A Method of Calibration for Human Body Model ESD Testers to Establish Correlatable Results," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-10. Advanced Micro Devices, 1988.

[112] M. Strauss, D. Lin, and T. Welsher, "Variations in Failure Modes and Cumulative Effects Produced by Commercial Human-Body Model ESD Simulators," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-9. AT&T Bell Labs, 1987.

[113] M. Jon and T. Weisher, "The metallurgical study of ESD damage in 256k dram devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-9. AT&T Bell Labs, 1987.

[114] L. van Roozendaal, A. Amerasekera, P. Bos, W. Baelde, F. Bontekoe, P. Kersten, E. Korma, P. Rommers, P. Krys, U. Weber, and P. Ashby, "Standard ESD testing of intergrated circuits," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-12. MOS technology research, Philips Research Laboratories, P.O.Box 80.000, 5600JA Eindhoven, The Netherlands, 1990.

[115] J. Barth, D. Dale, K. Hall, H. Hyatt, D. McCarthy, J. Nuebel, and D. Smith, "Measurements of ESD HBM events, simulator radiation and other characteristics toward creating a more repeatable simulation or; simulators should simulate." Barth Electronics, Inc. 1589 Foothill Drive, Boulder City, NV 89005, Tech. Rep., 1996.

[116] H. Hyatt, "ESD: Standards, threats and systems hardness fallacies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-24. Hyger Physics, Inc., 2002.

[117] P. Bossard, R. Chemelli, and B. Unger, "ESD damage from triboelectrically charged IC pins," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-2, 1980.

[118] L. Henry, M. Kelly, T. Diep, and J. Barth, "The importance of standardization CDM ESD test head parameters to obtain data correlation," in *Electri-*

cal Overstress / Electrostatic Discharge Symposium Proceedings, vol. EOS-22. ESD/TLP Consulting; Delphi Delco Electronics Systems; Texas Instruments; Barth Electronics, Inc., 2000.

[119] J. Barth and J. Richner, "Correlation considerations: Real HBM to TLP and HBM testers," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. BarthElectronics, Inc., 2001.

[120] T.J.Maloney and N.Khurana, "Transmission line pulsing techniques for circuit modelling of ESD phenomena," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-10. Intel Corporation., 1988.

[121] H. Gieser and M. Haunschild, "Very-fast transmission line pulsing of integrated structures and the charged device model," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-18. Fraunhofer-Institut Festkorpertechnologie (IFT/TMS ATIS Applied Test of Integrated Circuits), Hansastr, 27d, D-80686 Munich, Germany, 1996.

[122] K. Yokoi and T. Wantanbe, "Invited paper: A study of wafer level ESD testing," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. NEC Corporation, 2000.

[123] J. Barth, J. Richner, K. Verhaege, and L.G. Henry ESD/TLP Consultant, "TLP calibration, correlation, standards, and new techniques," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Barth Electronics Inc; Sarnoff Corporation, 2000.

[124] M. Lee, C. Liu, C.-C. Lin, J.-T. Chou, H. Tang, Y. Chang, and K. Fu, "Comparison and correlation of ESD HBM (Human Body Model) obtained between TLPG, wafer-level, and package-level tests," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. United Microelectronics Corp., 2000.

[125] H. Hyatt, J. Harris, A. Alanzo, and P. Bellew, "TLP measurements for verification of ESD protection device response," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. Littlefuse and Hyger Physics Inc, 2000.

[126] W. Stadler, X. Guggenmos, P. Egger, H. Gieser, and C. Musshoff, "Does the ESD-failure current obtained by transmission-line pulsing always correlate to human body model tests?" *Microelectronics Reliability*, vol. 38, pp. 1773–1780, 1998.

[127] S. G.Beebe, "Methodology for layout design and optimization of ESD protection transistors," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-18. Advanced Micro Devices., 1996.

[128] C. Musshoff, H. Wolf, H. Giesew, P. Eggew, and X. Guggenmos, "Risetime effects of HBM and square pulses on the failure thresholds of GGNMOS transistors." *Reliability of Electron Devices, Failure Physics and Analysis, 1996. Proceedings of the 7th European Symposium on*, pp. 1743–1746, 1996.

[129] R. Ashton, "Transmission line pulse measurements: Parametric analyzer for ESD on-chip protection," www.ESDA.org, September 2003, Tutorial from the ESDA 2003 conference.

[130] E. Grund, "Snapback Device Studies Using Multilevel TLP and Multi-impedance TLP Testers," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-27. Oryx Instruments Corp., 47341 Bayside Parkway, Fremont, CA 94538 USA, 2005.

[131] T. Daenen, S. Thijs, N. Mahadeva Iyer, V. Vassilev, V. De Heyn, and G. Groeseneken, "Multilevel Transmission Line Pulse (MTLP) Tester," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-27. IMEC, Belgium, 2005.

[132] J. E. Barth, K. Verhaege, L. G. Henry, and J. Richner, "TLP calibration, correlation, standards, and new techniques," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 24, no. 2, pp. 99–108, April 2001.

[133] K. Verhaege, P. J. Roussel, G. Groeseneken, H. E. Maes, H. Gieser, C. Russ, P. Egger, X. Guggnemoss, and F. G. Kuper, "Analysis of HBM ESD testers and specifications using a 4th order lumped element model," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-15. IMEC, Belgium., 1993.

[134] J. Lee, M. Hoque, J. Liu, G. Croft, W. Young, and J. Bernier, "A method of determining a transmission line pulse shape that produces equivalent results to human body model testing methods," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-22. University of Central Florida and Intersil Corporation, 2000.

[135] A. Guihaume, B. Foucher, J. Chante, P. Galy, S. Bardy, and F. Blanc, "Human body model test of a low voltage threshold SCR device: Simulation and comparison with transmission line pulse test," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. EADS CCR, CEGELY, Pole Universitaire Leonardo De Vinci, Philips Semiconductors, 2001.

[136] J. van Zwol, W. Kemper, and P. Bruin, "Transmission line pulsed photo emission microscopy as an ESD troubleshooting method," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-25. Philips Semiconductors, 2003.

[137] D. Pogany, M. Litzenberger, P. Kamvar, E. Gornik, C. Furbock, G. Gross, K. Esmark, H. Gossner, and M. Stecher, "Study of trigger instabilities in smart power technology ESD protection devices using laser interferometric thermal

mapping technique," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. Vienna University of Technology, Infineon Technologies, 2001.

[138] SAAT, "HBM ESD test results for the BC01b," available from William Pugsley.

[139] Cambridge Silicon Radio. BC01b specification. [Online]. Available: http://www.csr.com

[140] Bluetooth SIG. Bluetooth specification. [Online]. Available: http://www.bluetooth.org/spec/

[141] SEQUOIA Design Systems, Inc. Suppliers of finite element electrothermal semiconductor simulation software. [Online]. Available: http://www.sequoiadesignsystems.com/

[142] Synopsys, Inc. Suppliers of simulation software. [Online]. Available: http://www.synopsys.com

[143] *Taurus Manual*, 1st ed., www.Synopsys.com.

[144] V. Vassilev, H. Maes, M. Lorenzini, G, Groeseneken, and M. Steyaert, "Analysis and improved compact modelling of the breakdown behavior of sub-0.25 micron ESD protection ggNMOS devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-23. IMEC, KUL, ESAT, 2001.

[145] V. Vassilev, M. Lorenzini, P. Jansen, G. Groeseneken, S. Thijs, N. Mahadeva Iyer, M. Steyaert, and H. Maes, "Advanced modelling and parameter extraction of the MOSFET ESD breakdown triggering in the 90nm CMOS node technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-26. IMEC and KUL, ESAT, 2004.

[146] H. Hayashi, T. Kuroda, K. Kato, K. Fukuda, S. Baba, and Y. Fukuda, "ESD Protection Design Using a Mixed-Mode Simulation for Advanced Devices," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-26. Oki Electric Industry Co., Ltd. 550-1 Higashiasakawa-cho, Hachiouji-shi, Tokyo, PD 193-8550 Japan, 2004.

[147] M. G. Khazhinsky, J. W. Miller, M. Stockinger, and J. C. Weldon, "ESD protection of single output buffers for Maximum Failure Voltage in advanced CMOS technologies," in *Electrical Overstress / Electrostatic Discharge Symposium Proceedings*, vol. EOS-26. Freescale Semiconductor, Inc., 3501 Ed Bluestein Blvd., M/D K10, Austin, TX 78721, United States, 2004.

[148] J. Liou and G. X. Gao, "Design and modeling of on-chip electrostatic discharge (ESD) protection structures." in *Proc. 24th International Conference on Microelectronics (MEIL 2004)*, NIS, Serbia and Montenegero., 2004, pp. 619–624.

[149] *Sequoia Manuial*, 1st ed., SEQUOIA Design Systems, Inc.

Appendix A

Taurus

Taurus runs on either UNIX or Linux based systems and is run using a script file as an input. All default simulation parameters are specified in a configuration file and can be overridden by the input script file. This allows complete flexibility in the choice of physical models used and model parameters. A summary of the models is available from Synopsys [143].

Below is an example of a Taurus input script with an explanation of what each instruction does. The purpose of this script is to create the simulation structure shown in Figure A.1. The following script will then simulate the electrical operation of this transistor. The first 4 lines of this script (below) are not interpreted by Taurus as they are preceded by '#' which indicates that it is a comment.

```
1    # Synopsys Taurus-Device Example 0.13um NMOSFET
2    # Structure generation
3
4    # Enable device mode
```

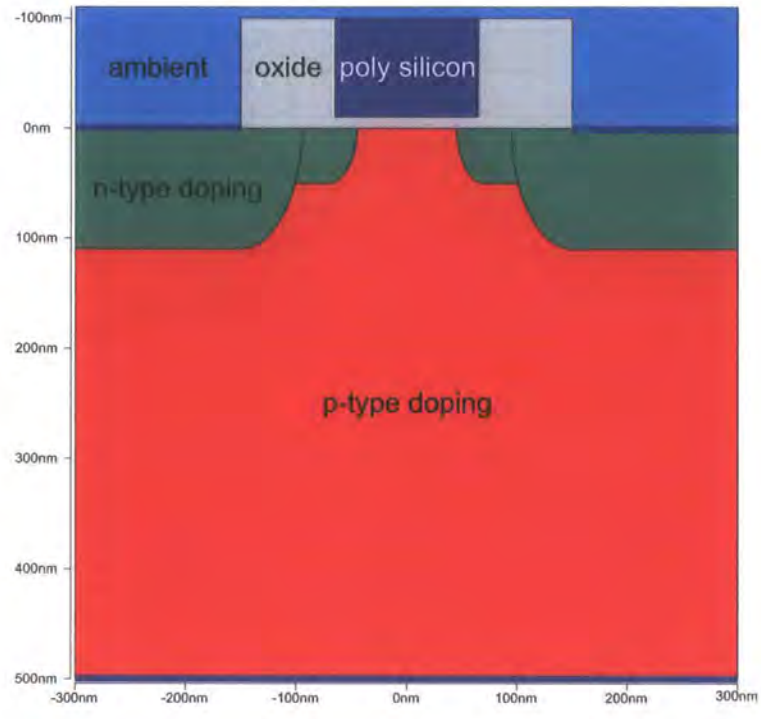

Figure A.1: An un-simulated transistor structure

The first instruction to the simulator is to tell it whether it is simulating a device operation or device fabrication. The software is also able to simulate the fabrication of a transistor including the process variations to produce structures that can subsequently be electrically simulated to derive their device operation characteristics. All instructions up to line 95 define the structure's dimensions and doping profiles. The subsequent instructions (lines 96 to 131) are used to create a simulation grid defining the individual cell size for calculations.

```
5    Taurus {device}
6
```

This first instruction defines the overall size of material to be used ($-300nm < x < 300nm$ and $-110nm < y < 500nm$). The x-axis increases from left to right and the y-axis increases from the top to bottom. Three region types are defined which will be of either silicon, silicon oxide or electrode (effectively metal)

material. Finally grid lines are added at important points in the structure.

```
7   # Define the device size, list the regions, and specify fixed mesh lines.
8   DefineDevice (
9       minX=-300nm, maxX=300nm,
10      minY=-110nm, maxY=500nm,
11
12      Region (name=silicon1, material=silicon),
13      Region (name=oxide1, material=oxide),
14      Region (name=gate, material=electrode),
15
16      x=-150nm, x=-65nm, x=0nm, x=65nm, x=150nm,
17      y=-100nm, y=-3nm, y=-1.5nm, y=0nm
18  )
19
```

In this next instruction the whole region below the x-axis is defined as silicon material (Figure A.1 green and red regions).

```
20  # Define the silicon substrate region
21  DefineBoundary (
22      region=silicon1,
23      Polygon2D (
24          Point (x=-300nm, y=0nm), Point (x=-150nm, y=0nm),
                                Point (x= 150nm, y=0nm),
25          Point (x= 300nm, y=0nm), Point (x= 300nm, y=500nm),
                                Point (x=-300nm, y=500nm)
26      )
```

```
27  )
28
```

The gate oxide and the spacers are defined by a single region of silicon oxide. It is used to form two 85nm long spacers either side of the gate oxide as well as the $130nm$ long $3nm$ thick gate oxide, (Figure A.1 grey region).

```
29  # Define the oxide region
30  DefineBoundary (
31      region=oxide1,
32      Polygon2D (
33          Point (x=-150nm, y=-100nm), Point (x= -65nm, y=-100nm),
34          Point (x= -65nm, y=-3nm), Point (x=65nm, y=-3nm),
35          Point (x=65nm, y=-100nm), Point (x= 150nm, y=-100nm),
36          Point (x= 150nm, y= 0nm), Point (x=-150nm, y= 0nm)
37      )
38  )
```

The poly silicon region above the gate oxide and in between the spacers is defined as an electrode region to form the gate contact (Figure A.1 dark blue region).

```
39
40  # Define the electrode gate region
41  DefineBoundary (
42      region=gate,
43      Polygon2D (
44          Point (x=-65nm, y=-100nm), Point (x= 65nm, y=-100nm),
```

```
45          Point (x= 65nm, y=-3nm), Point (x=-65nm, y=-3nm)
46      )
47  )
```

The DefineContact instruction assigns nodes at the top left, top right and bottom to be the source, drain and substrate contacts respectively. These points are displayed as blue lines in Figure A.1. All these nodes are set at the same potential and are connected to a stimulus as defined by the simulation input file.

```
48
49
50  # Flat Contacts
51  DefineContact (name=source,X (min=-500nm, max=-149nm),
                                Y (min= -1nm, max=1nm))
52  DefineContact (name=drain, X (min= 149nm, max= 500nm),
                                Y (min= -1nm, max=1nm))
53  DefineContact (name=substrate, X (min=-500nm, max= 500nm),
                                Y (min=499nm, max=501nm))
54
```

The instruction at line 57 uniformly p-type dopes the whole of the silicon region (Figure A.1 green and red regions).

```
55
56  # Substrate Doping: P-type Uniform
57  Profile (name=Ptype, region=silicon1, Uniform (value=6e17))
58
```

The n-type dopants are added in two stages. The first implant is done after the gate oxide has been defined and therefore the spacing between the source and the drain implants is $\sim 130nm$, i.e. the gate length. The second set of implants is added after the spacers are defined which means that the spacing between the second set of implants is $\sim 300nm$.

The LDD source doping will extend just past the start of the gate oxide and it will have a peak doping value of $1 * 10^{20} ions/cm^3$ at the surface which will decrease following a gaussian profile to $6 * 10^{17} ions/cm^3$ at a depth of 50nm. The lateral spread of the dopant underneath the gate oxide is defined by the *LateralRatio=0.5* entry so that the dopants will diffuse $25nm$ under the gate oxide. The level of doping in this region is set to ensure correct switching operation of the transistor. The *n-type* doping is denoted by the green region in Figure A.1.

```
59  # Source extension doping: N-type Gaussian
60  Profile (
61      name=Ntype, region=silicon1,
62      Gauss (
63          peakValue=1e20, depthValue=6e17, depth=50nm,
                                lateralERFC=true, LateralRatio=.5,
64          Edge (Point (x=-500nm, y=0nm), Point(x=-65nm, y=0nm) )
65      )
66  )
```

More dopant is added to the source side of the transistor though just past the spacers and it will have a peak value of $2 * 10^{20} ions/cm^3$ which will decrease following a

gaussian profile to $6 * 10^{17} ions/cm^3$ at a depth of 110nm. The level of doping in this region is set to ensure that an ohmic contact exists between the silicon and the metal contacts.

```
67
68  # Source doping: N-type Gaussian
69  Profile (
70      name=Ntype, region=silicon1,
71      Gauss (
72          peakValue=2e20, depthValue=6e17, depth=110nm,
                              lateralERFC=true, lateralRatio=.5,
73          Edge (Point (x=-500nm, y=0nm), Point (x=-150nm,y=0nm))
74      )
75  )
76
```

The next two instructions doped the drain side of the transistor in the same way as the source was doped.

```
77  # Drain extension doping: N-type Gaussian
78  Profile (
79      name=Ntype, region=silicon1,
80      Gauss (
81          peakValue=1e20, depthValue=6e17, depth=50nm,
                              lateralERFC=true, lateralRatio=.5,
82          Edge ( Point (x=65nm, y=0nm), Point (x=500nm, y=0nm))
83      )
84  )
```

```
85
86  # Drain doping: N-type Gaussian
87  Profile (
88      name=Ntype, region=silicon1,
89      Gauss (
90          peakValue=2e20, depthValue=6e17, depth=110nm,
                              lateralERFC=true, lateralRatio=.5,
91          Edge ( Point (x=150nm, y=0nm), Point (x=500nm ,y=0nm))
92      )
93  )
94
95
```

This instruction adds gridlines to create a coarse $100nm$ grid in the device.

```
96  # Initial coarse regrid
97  Regrid (gridProgram=pm, meshSpacingX=100nm, meshSpacingY=100nm)
98
```

The grid is further refined such that regions with a high doping gradient have a higher mesh density.

```
99  # Regrid on doping
100 Regrid (
101     gridProgram=pm, meshSpacing=16nm, region=silicon1,
102     Criterion (name=NetDoping, delta=.5, type=asinh)
103 )
```

```
104
```

The grid directly underneath the gate oxide is further refined to a maximum mesh dimension of 8Å.

```
105 # Regrid in channel
106 Regrid (
107     gridProgram=pm, region=silicon1, meshspacingy=8A,
108     minX=-65nm, maxX=65nm, maxY=6nm
109 )
110
```

Further refinement of the grid based on the electrostatic gradients in the chip can be implemented; but before this is done the electrostatic potential needs to be calculated. The next section finds a zero carrier solution for the transistor with all inputs set to $0V$. A zero carrier solution is one where the generation, recombination and movement of charge do not form part of the calculation thus it requires less computational effort to achieve a solution. As there is no net charge flow between the regions in the transistor or new generation/recombination, these calculations are not necessary; this is therefore purely an electrostatic solution.

```
111 # Zero-carrier solve at equilibrium
112 Symbolic (carriers=0)
113 Solve {}
114
```

The initial solution allows further refinement of the mesh where there are regions with a high electric field.

```
115 # Regrid on potential
116 Regrid (
117 gridProgram=pm, meshSpacing=16nm,
118 Criterion (name=ElectricPotential, delta=.1, type=linear)
119 )
120
```

The regions with high doping gradients or high electric fields will have a greater variation in material properties once the transistor is being simulated. Therefore the mesh size needs to be adjusted to compensate. An alternative solution is to have a fine mesh over the whole of the transistor but this is an inefficient use of the computer resources and it is therefore better to be selective.

This next instruction ensures that all the triangles in the mesh are acute (all angles $< 90^0$).

```
121 # Regrid to desired grading factor and maximum element angle
122 Regrid (
123     gridProgram=pm, region=silicon1,
124     gradingFactor=2.01, MaximumAngle (value=90)
125 )
126
```

The transistor is solved, again with no carriers, and the structure is saved to file.

```
127 # Redo zero-carrier solve
128 Solve {}
129
130 # Save structure
131 Save (meshfile=mos_struc_ex.tdf)
```

The previous file creates the structure to be simulated and the following file simulates snapback in the ggNMOS.

```
1   # Synopsys Taurus-Device Example 0.13um N-Channel MOSFET
2   # Breakdown due to impact ionization
3
```

Again the simulator needs to be told whether it is simulating a manufacturing process or device operation.

```
4   # Enable device mode
5   Taurus {device}
6
```

This instruction loads the structure file created by the previous file.

```
7   # Load device structure
8   DefineDevice (meshfile=mos_struc_ex.tdf)
9
```

```
10  # Specify common physics models to use
11  Physics (
12      Global (
13          Poissons (
14              FermiStatisticsActive=true
15          )
16      )
17      Silicon (
18          ElectronContinuity (
19              Mobility (
20                  LowFieldMobility (
21                      conModelActive=true, conModel=AnalyticModel,
22                      surfModelActive=true, surfModel=LombardiSurfaceModel
23                  )
24                  highFieldMobility=true
25              )
26          )
27      )
28  )
29  # Set gate workfunction
30  Contact (name=gate, workfunction=4.17)
```

The default set up of *Taurus* (which models are selected and the variables for the models) is defined by a file in the *Taurus* directory and is suitable for most simulations. In this case impact ionization modelling is activated for both electrons and holes using the ChynowethFieldII model as it plays a crucial part in the operation of a ggNMOS during snapback.

```
31  # Turn on Impact Ionization
32  Physics (
33      Silicon (
34          ElectronContinuity (
35                              impactIonizActive=true,
36                              ElectronImpactIonization (ChynowethFieldIIActive=true)
37          )
38          HoleContinuity (
39                              impactIonizActive=true,
40                              HoleImpactIonization (ChynowethFieldIIActive=true)
41          )
42      )
43  )
44
```

All the electrodes are set to $0V$ except for the drain which is set to $2V$ to speed up the simulation as it is expected that nothing interesting will happen when $V_{DS} < 2V$. It was also found that at times the simulation would fail if instructed to simulate the transistor starting at $V_{DS} = 0V$.

```
45  # Set bias on contacts. Set drain to 2V.
46  Voltage( electrode=source, value=0.0 )
47  Voltage( electrode=gate, value=0.0 )
48  Voltage( electrode=drain, value=2.0 )
49  Voltage( electrode=substrate, value=0.0 )
50
```

An initial solution is calculated without any of the carriers followed by a two carrier (electrons and holes) solution. The initial solution acts as a stepping stone to a full solution.

```
51  # Specify zero-carrier solution
52  Symbolic (carriers=0)
53
54  # Do Initial Solve
55  Solve {init}
56
57  # Specify two-carrier solution and re-solve
58  Symbolic (carriers=2)
59  Solve {}
60
```

The *Continuation* instruction applies a voltage source, with a variable resistance, to the structure. It then attempts to solve the structure such that the drain voltage has moved *'forward'* along the load line of the resistor. The value of the resistor and the applied voltage (to the resistor and the transistor in series) is determined by the previous solution. If the solution proves impossible to achieve the simulator decreases the simulation step size and tries to solve it again. It is known that during snapback the voltage across the device decreases after it has been triggered, this is effectively negative resistance. The simulator is able to follow the IV curve by applying a negative resistance in series with it.

This instruction continues until either the voltage or current go outside of their respective minimums and maximums or the simulator fails to solve the structure 12 times in succession.

The IV curve of the structure is saved in a logfile.

```
61  # Use continuation to trace snapback
62  Continuation (
63      logfile=mos_brkdn_ex.data,
64      electrode=drain, cvStep=0.4,
65      cvMin=0, cvMax=5, ciMin=0, ciMax=1e-2
66  )
67
```

The structure after the final simulation, including the impact ionization values, is saved for future reference.

```
68  # Save TDF file with impact ionization rates
69  Save (
70      meshfile=mos_brkdn_ex.tdf,
71      Add (HoleImpactIonization, ElectronImpactIonization)
72  )
```

Synopsys supply a detailed manual [143] covering most aspects of the software. It covered the physics background for all of the models and all of the modelling commands in detail.

Appendix B

TLP Stimulus

The TLP voltage source given in Equation B.1, is made up of the sum of Equations B.2, B.3 and B.4 where t is time in seconds, t_R is the rise time in seconds, t_L is the pulse length in seconds and A is the pulse amplitude in volts. The sum of the three Equations results in the waveform illustrated in Figure B.1. Inequalities at the end of each Equation ensure that it is active at the right time.

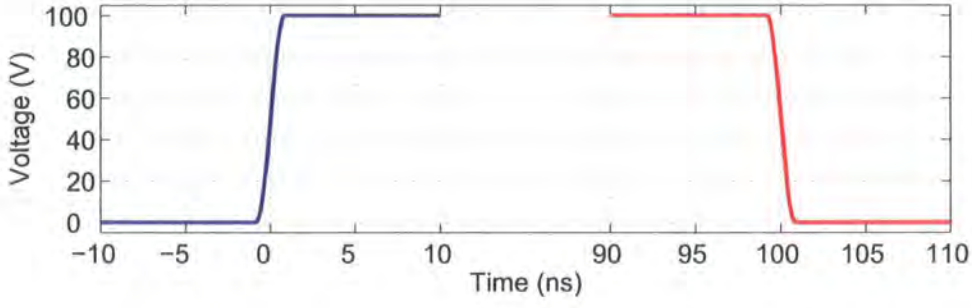

Figure B.1: The voltage characteristics of a TLP waveform with t_R and t_F of $1ns$ and a duration of $100ns$

$$V_{DS-TOTAL} = V_{DS-1} + V_{DS-2} + V_{DS-3} \tag{B.1}$$

$$V_{DS-1} = A * \frac{-cos\frac{1.85*(t+0.845)}{t_R} + 1}{2} * (-0.845 <= t) * (t < 0.845 * t_R))) \quad \text{(B.2)}$$

$$V_{DS-2} = A * ((0.845 * t_R) <= t) * (t < (t_L - 0.845 * t_R)) \quad \text{(B.3)}$$

$$V_{DS-3} = A * \frac{cos\frac{(t_L - 0.845 * t_R - t)*1.85}{t_R} + 1}{2} * (t_L - 0.845 * t_R <= t) * (t < (t_L + 0.845 * t_R)) \quad \text{(B.4)}$$