ESD Protection Design Methodology in Deep Sub-micron CMOS Technologies

by

Nitin Mohan Anil Kumar

Project Report

Course E&CE 730 (Topic 9)

VLSI Quality, Reliability and Yield Engineering

Winter 2003

Course Instructor: Professor Manoj Sachdev
Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario, Canada

Table of Contents

Table of Figures

1. Introduction

Electrostatic discharge (ESD) is a charge re-balancing process between two adjacent objects, which involves a rapid discharge of accumulated static electricity. The ESD is a major cause of failure during the manufacturing, testing, handling and assembly of integrated circuits (ICs). ESD/EOS (electrical overstress) is responsible for nearly 40% of the failures of IC customer returns [1]. Human beings, wafer-processing equipments, testing and automation equipments that come in contact with the ICs, can generate the static charge responsible for the ESD events. Endorsing the appropriate handling conditions and on-chip protection circuits can minimize the chances and after-effects of the ESD events on the ICs. The on-chip ESD protection circuit design is non-trivial and it requires an accurate modeling of the actual ESD events. The protection circuits must survive the ESD events and also provide protection to the core-circuitry by clamping the excessive voltage and sinking the excessive current. Also the protection circuit must be fast enough to respond to the ESD event and protect the core-circuitry before the ESD pulse destroys them. Metal-oxide-semiconductor (MOS) devices are particularly vulnerable to the ESD events due to the breakdown of the thin gate oxide layer and the current flow near the surface of the substrate [2][3]. Presently, majority of ICs are fabricated in advanced complementary MOS (CMOS) technologies due mostly to the low static power consumption, high noise margins, convenient scaling and high integration. Therefore, it is more challenging to design the ESD protection circuits in the modern ICs.

Due to aggressive scaling of the modern CMOS technologies, the protection capability of the devices degrades with the advent of new generations of the technology, i.e., the ESD protection circuits designed for one generation of the technology may not be suitable for the other generation of the technology. Therefore, it is required to understand the mechanisms involved in the ESD related device failure and make the ESD protection design methodology more systematic and transferable to the newer generations of the technology. As the technology enters in the nano-scale regime, significant research and design efforts are needed to design in the ESD protection circuits due to the shallower junctions, thinner gate oxides, higher complexity of the doping profiles, narrower widths of the metal lines and vias, and higher levels of interconnects. Use of lightly doped drains (LDDs) and silicide in newer technologies also exacerbates ESD performance of the devices [4][5]. If the LDD diffusions are shallower than the source/drain diffusions then for the same current level, the current density is higher in the LDD regions. This causes localized heating and therefore higher chance of an ESD related damage. On the other hand, silicided source/drain diffusions lead to current localization by concentrating the current flow on the surface of the device and reducing the ballasting resistance that distributes the current

[6][7]. The increasing requirement of high-speed I/Os further aggravates the design constraints on the ESD protection circuits. The empirical, trial and error method to design the ESD protection schemes is based on the fabrication of several test protection structures, application of the gradually increasing voltage pulses and measurement of the functionality of the protection structures. This method is time consuming and destructive in nature, i.e., it eventually damages the device under test. Moreover, it does not help in the evolution of the protection circuits for the future technologies. A better design methodology includes accurate modeling of electrical characteristics of the key devices under the extreme voltage and current conditions similar to those of an actual ESD event and extraction of the critical circuit parameters that affect the protection capability of the circuit. The model can be used to optimize the design and also it can predict the protection performance of the design for the future technologies.

Figure 1.1 shows the typical failure threshold currents of ESD protection devices for industry standard process conditions and technology nodes.

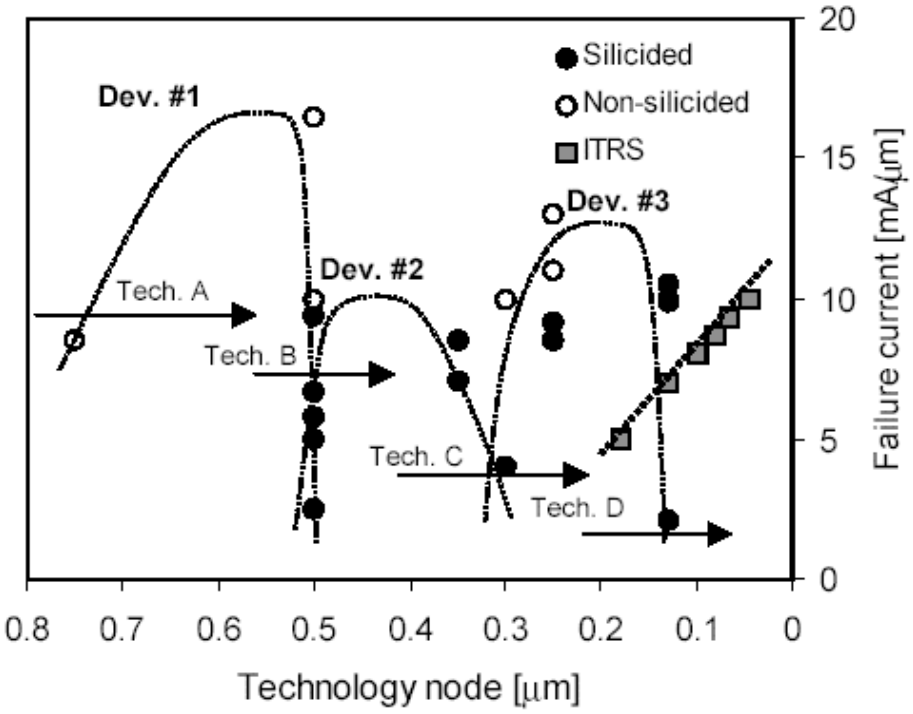


Figure 1.1 Reported typical failure threshold current of ESD protection devices (NMOS transistors) for various process conditions and technology nodes (Compiled by the TCAD group, Stanford University - data from Philips, TI, IBM and IMEC)

In this case, the failure current depicts the second breakdown triggering current (explained later) that causes the thermal runaway and damages the device permanently. As shown in the figure 1.1, the robustness of ESD protection device increases as the protection scheme matures. However, with the advent of a new technology, the protection level reduces drastically since the previous scheme does not work for the new technology. It can also be observed that the non-silicided devices exhibit better ESD protection than their silicided counterparts because of the reasons explained earlier. The International Technology Roadmap for Semiconductor (ITRS) predicts that the ESD protection of the devices should improve with the advancement of the process technology [8]. However, as the process technology enters the nano-scale regime, devices become more vulnerable to the ESD stress due to the reasons mentioned earlier. Therefore, innovative ESD protection schemes are needed not only to restore the protection performance to the previous levels but also to further improve it.

2. ESD protection devices

Ideally, a protection device should be capable of protecting the core circuitry without damaging itself in the presence of ESD events. In addition, it should not interfere with the normal operation of the core circuitry. Therefore, an ideal protection device should be able to shunt large amount of ESD current with a negligible ohmic voltage drop [9]. If the on-resistance of the protection device is not negligible, the generated voltage may damage the thin gate oxide of MOSFETs in the core circuitry. To avoid unintentional triggering of the protection circuitry, the sustaining voltage of the protection devices should be higher than the supply voltage (V_{dd}) with a safety margin. In the presence of an ESD event, the protection devices should trigger instantaneously so that the protection devices consume the destructive energy of the event before it damages the core circuitry. The following devices are used to implement most of the ESD protection circuits:

2.1. Resistor

A resistor is typically used to limit the current, drop the voltage, slow down the transients and isolate the ESD protection networks. Diffused resistors are generally preferred over the thin film resistors due to their large current handling capability. The high current I-V behavior of a diffused resistor is shown in figure 2.1 [10][11]. For high voltages, the current saturates due to velocity saturation of the electrons. A further increase in the voltage causes a permanent thermal failure.

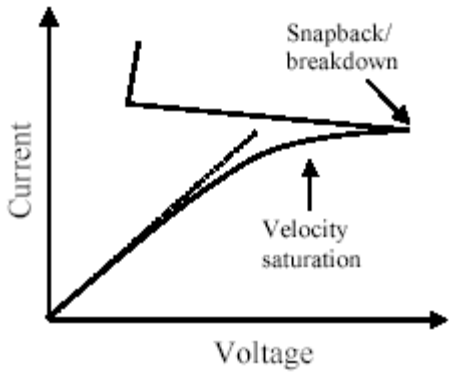


Figure 2.1 Typical I-V behavior of a diffused resistor [10]

2.2. Diode

A diode is the simplest voltage-clamping device. It has excellent current handling capability and low turn-on voltage in the forward bias. However, the diode exhibits poor voltage clamping and higher on-resistance in the reverse breakdown region. For high power supply (V_{dd}) implementations, multiple diodes can be added in series but this arrangement increases the on-resistance. A typical ESD protection circuit using diodes is shown in figure 2.2. For normal operation, the diodes do not conduct due to the reverse bias condition. As soon as an ESD event occurs and the I/O pad voltage goes beyond $V_{dd} + V_{ON\text{-}DIODE}$ or $V_{ss} - V_{ON\text{-}DIODE}$, one of the diode conducts and clamps the pad voltage either at $V_{dd} + V_{ON\text{-}DIODE}$ or $V_{ss} - V_{ON\text{-}DIODE}$. If a large negative ESD pulse appears at the I/O pad with respect to the V_{dd} pin or a large positive ESD pulse appears at the I/O pad with respect to the V_{ss} pin, the corresponding diode goes into reverse breakdown and clamps the pad voltage at the diode reverse breakdown voltage with respect to the corresponding power supply pin.

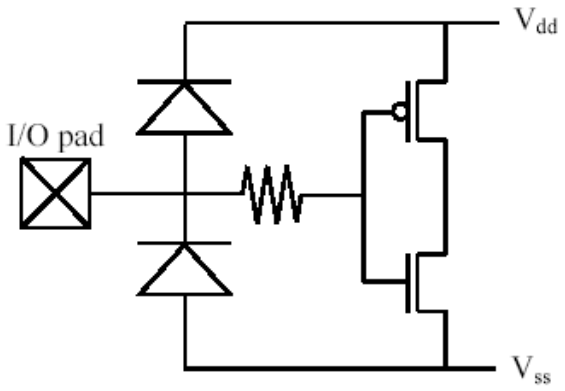


Figure 2.2 A typical ESD protection scheme using diodes

2.3. NMOS transistor

In CMOS technology, an NMOS transistor contains a parasitic n-p-n (source-substrate-drain) bipolar transistor as shown in figure 2.3 (a). Interestingly, this n-p-n transistor is a bulk device (unlike MOSFET, which is a surface device) and therefore it can handle large currents if turned on, which is called snapback conduction. Out of several configurations, gate-grounded NMOS (ggNMOS) configuration is most basic one and is shown in the figure 2.3 (a). The typical I-V behavior is shown in figure 2.3 (b).

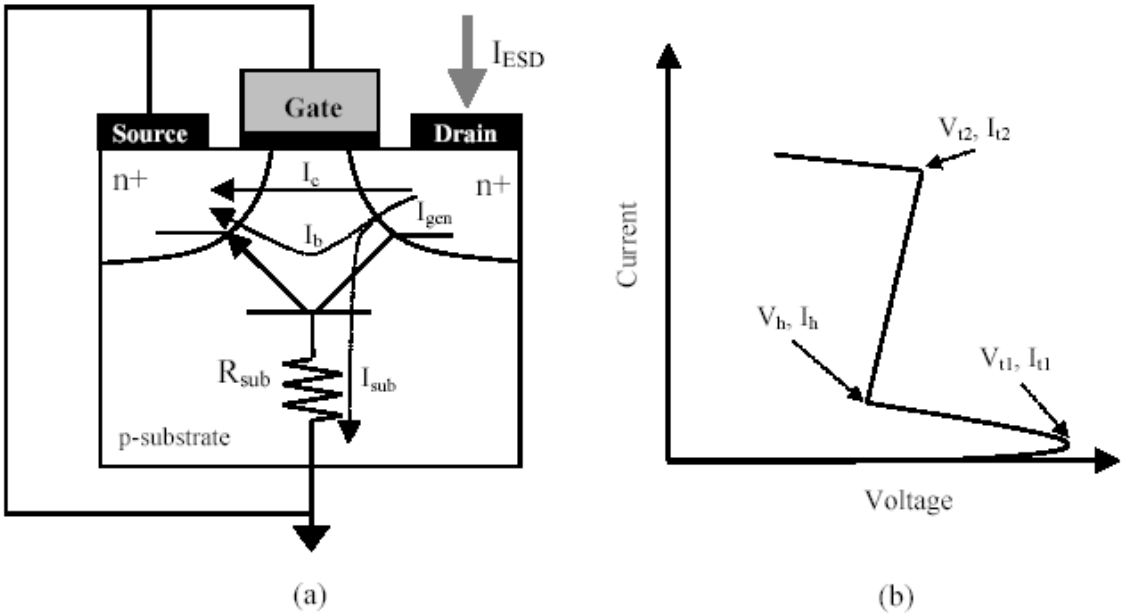


Figure 2.3 (a) Typical operation of the gate-grounded NMOS (ggNMOS) and (b) I-V behavior

Under the normal operation, the NMOS transistor is OFF (V_{gs} = 0) and there exists zero current from the drain-to-source (except the negligible leakage and subthreshold current). An ESD event at the drain node builds a high electric field across the reverse biased drain-substrate junction, which triggers the impact ionization or avalanche multiplication process. The avalanche generation current I_{gen} can be given in terms of the collector current (drain-to-source current) as follows:

$$I_{gen} = (M-1)\, I_c$$

where M is an empirically determined multiplication factor. The voltage drop across the substrate resistance (R_{sub}) increases the base-emitter voltage (V_{be}) of the n-p-n transistor and eventually

turns on the transistor (when V_{be} = 0.7V). This condition is shown in the figure 2.3 (b) by V_{t1}, I_{t1} and the suffix t1 stands for the time it takes to reach the trigger point, which is usually on the order of 1ns but is very dependent upon the ESD pulse height and rise time. In this situation, the transistor operates in a self-biasing mode i.e. the current is mostly dependent on the source to substrate electron injection and very less on the avalanche multiplication. The regenerative condition for snapback to occur is given by [12]:

$$\beta (M - 1) \geq 1$$

where β is the current gain of the parasitic lateral bipolar n-p-n transistor. Since the high electric field across the drain-substrate junction is no longer needed to maintain the current, the drain voltage quickly drops to the BV_{CEO} of the n-p-n bipolar transistor. The condition is shown in the figure 2.3 (b) by V_h, I_h and it is analogous to the hold voltage of a silicon controlled rectifier (SCR) device. In the snapback mode, the current increases with a slope $1/R_{sb}$, where R_{sb} is the dynamic snapback resistance, which is equal to the resistance of the source and drain diffusions and contacts and is usually on the order of a few ohms. The snapback phenomenon is non-destructive provided the current does not increase to the level that it triggers thermal runaway (or second breakdown), which damages the device permanently. The thermal runaway triggers at the time t2 and (V_{t2}, I_{t2}) depicts this condition in the figure 2.3 (b). At this moment, a localized hot spot forms in the high joule-heating (**J . E**) region. The high temperature increases the localized resistivity due to the mobility degradation, which further increases the temperature. As the temperature increases, the intrinsic carrier concentration increases exponentially and eventually exceeds the background dopant concentration [13]. Therefore the resistivity reaches a maximum point and then decreases, which increases the current and thus the temperature. This is a positive feedback phenomenon that leads to very high current and localized temperature until the material melts and causes a void [14].

The above simple theory overestimates the destructive nature of the second breakdown. In fact, the second breakdown refers to the negative resistance region when the voltage is decreased due to a reduction in the resistivity of the silicon in the hot spot. The resistivity around the hot spot still remains high and therefore the current does not always increase to the point of melting the material and causing a void. Thus, the second breakdown is not synonymous with the device failure.

The ESD protection device should be designed in such a way that the operating voltages in the high current regime remain smaller than the thin gate oxide breakdown voltage (BV_{ox}). However, the voltages should be larger than the supply voltage (V_{dd}) with a safety margin to avoid any unintentional triggering of the ESD protection device due to noise or voltage overshoot. Based on the above discussion, the design window of the ESD protection device can be obtained as shown in figure 2.4 [21].

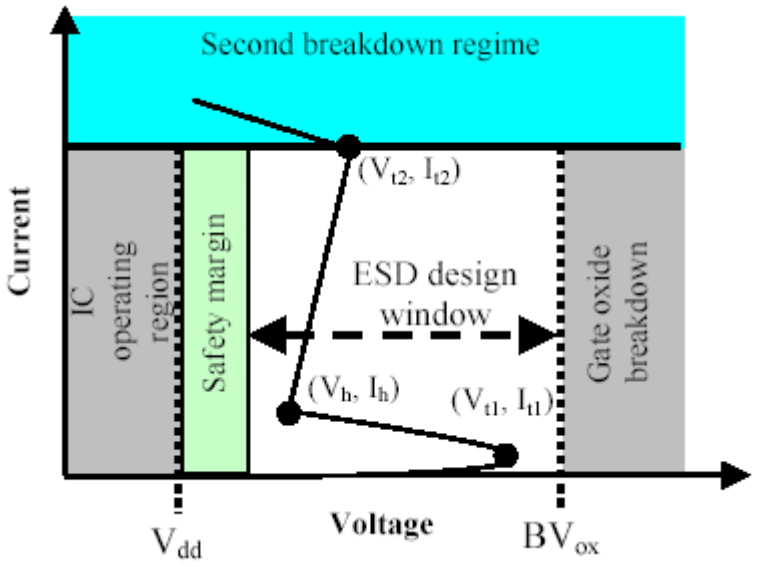
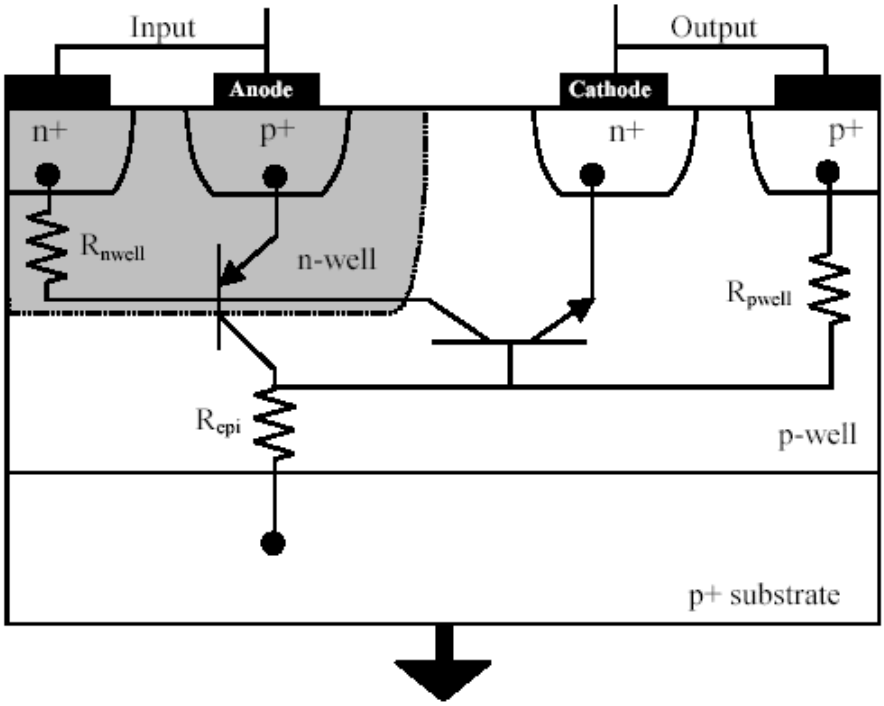


Figure 2.4 Design window of the ESD protection device

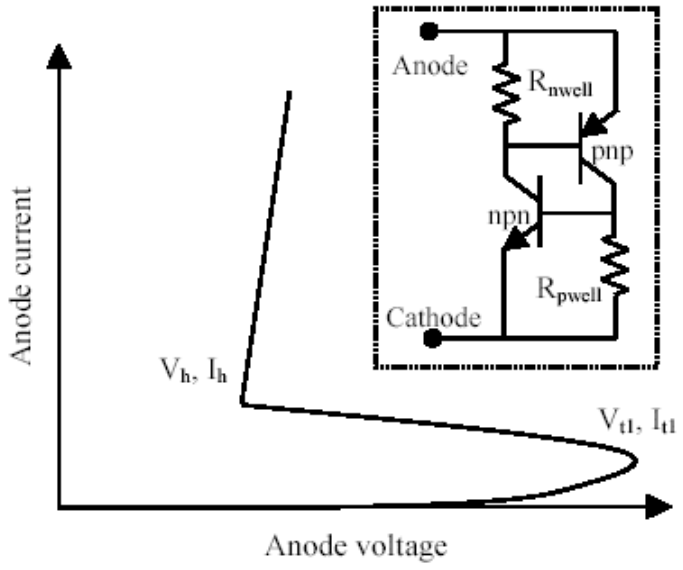
2.4. Silicon controlled rectifier (SCR)

The SCR has a large current handling capability due to the combination of two bipolar transistors functioning in self-regenerative condition. Figure 2.5 shows the cross section of a lateral SCR in CMOS technology and its high current I-V behavior. Under the normal operation, the n-well/p-well junction is reverse biased, no current flows and both the bipolar transistors remain OFF. As the voltage across the input and output increases, it eventually breaks down the n-well/p-well junction and the current increases by the avalanche multiplication process. This current develops a voltage drop across the p-well resistor (R_{pwell}) and as the base-to-emitter voltage of n-p-n transistor, $V_{be\text{-}npn}$ = 0.7V, the n-p-n transistor turns on. The collector current of the n-p-n transistor further drops the base voltage of the p-n-p transistor with respect to its emitter, which eventually turns on the p-n-p transistor ($V_{be\text{-}pnp}$ = -0.7V). In this situation the SCR does not require a high voltage across input and output to maintain the current and the voltage across the SCR reduces. The minimum voltage required to maintain this conduction state is called

the holding voltage (V_h) and the corresponding current is called the holding current (I_h). This condition is shown in the figure 2.5 (b) by (V_h, I_h) and the SCR is called latched in this state.



(a)



(b)

Figure 2.5 (a) Cross section of lateral SCR in CMOS and (b) its high current I-V behavior

The necessary condition for SCR to remain in this regenerative state is [16]:

$$\beta_{npn}\,\beta_{pnp} \geq 1$$

However, in the advanced CMOS processes, the breakdown voltage of the n-well/p-well is about 20V and it depends on the doping level and profile. This voltage is too large to be useful in many ESD protection circuits. Inserting an additional device between the n-well and p-well reduces the breakdown of the n-well/p-well junction. However, it is difficult to obtain a breakdown voltage less than 10V in typical advanced CMOS technologies. Often the ESD protection is achieved in two or more stages. The first stage consists of SCRs and the subsequent stages may include NMOS transistors or diodes. The NMOS/diodes respond quickly to an ESD event and protect the core circuit by shunting some of the current. A resistor isolates the two stages and develops a voltage drop across them due to the current flowing into the second stage. This voltage turns on the SCRs, which further shunt a large amount of destructive current.

3. Experimental techniques

The most common experimental techniques to characterize and analyze the ESD protection strength of the devices are transmission line pulsing (TLP) and photon emission microscopy (EMMI) [17].

3.1. Transmission line pulsing (TLP)

Typically, the ESD pulses occur with the pulse widths of few hundred nanoseconds. Therefore, it is important that similar voltage pulses should be generated to characterize the ESD protection strength of the devices since the DC measurements do not represent the actual transient behavior of the devices in the high current region due to excessive heating in the device. This can be achieved by using a transmission line in the arrangement shown in figure 3.1. The transmission line is charged to a voltage by a variable voltage source and then discharged through the device under test (DUT) via a switch [18]. The pulse width of the discharge pulse can be given by:

$$T_d = 2L/v$$

where L is the length and v is the phase velocity of the transmission line. The pulse width and amplitude of the pulses can be controlled by changing the length of the transmission line and

amplitude of the charging voltage source respectively. This TLP arrangement can be automated by using a PC and a probe station.

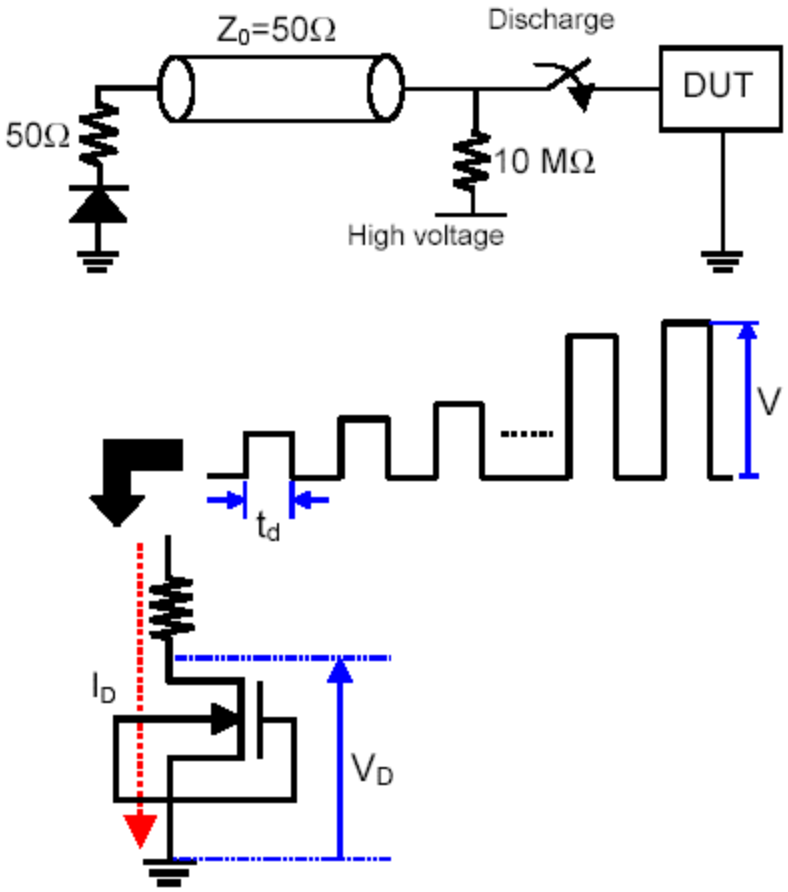


Figure 3.1 Transmission line pulsing method to characterize the ESD protection devices

3.2. Emission Microscopy (EMMI)

Photon emission microscopy is a popular technique to analyze failure and reliability of the semiconductor devices. Sensing the photon emission from different parts of the device gives an indication of failure and high current density regions [19]. A typical EMMI setup is shown in figure 3.2. Photons are emitted from the material when electrons in the conduction band recombine with the holes in the valence band. However, the photons may also be generated by intra-band transition, thermal radiation and tunneling currents in the oxide. In MOS transistors, the intra-band transition is one of the predominant mechanisms to generate the photons under high electric field and current conditions. The generated photons pass through the transparent

dielectric layers and scattered from the patterned metal layers. The collection and analysis of the photons emerging from the top of the wafer (through dielectric and metal layers) is referred to as frontside light emission microscopy. On the other hand, the corresponding analysis on the photons emitting from the back of the wafer (through the silicon substrate) is referred to as backside light emission microscopy. The EMMI technique can pinpoint the location of the failure on the silicon wafer. It should be noted that before the EMMI technique is used to determine the ESD protection strength of a device, the voltage pulses with reasonable pulse widths should stress the device since the real life ESD events are also transient.

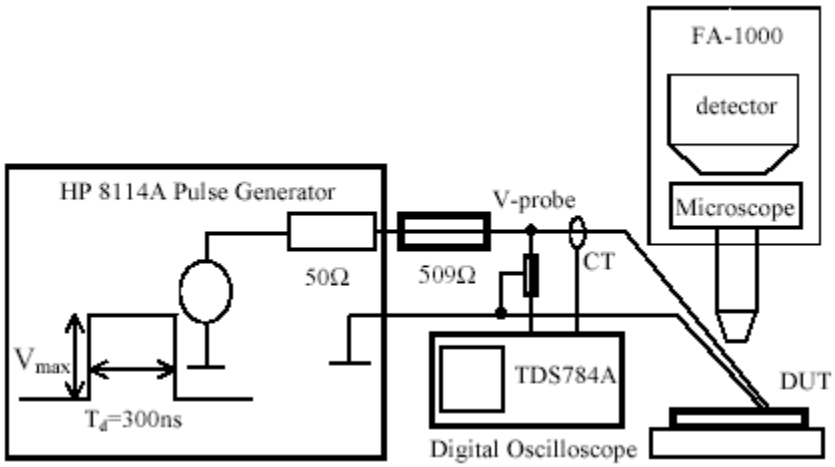


Figure 3.2 EMMI setup to determine the ESD protection strength of devices

4. ESD protection circuit design

An ESD protection circuit should provide a low impedance path from an I/O pad to a power supply pin (V_{dd} or ground) during an ESD event but it should have a high impedance path between the two pins under the normal operating conditions so that it does not interfere with the normal operation of the core circuitry. In addition, the protection circuit should also be able to clamp the input voltages below the breakdown voltage of the thin gate oxide in the MOSFETs. Typically, the oxide breakdown depends on the voltage amplitude as well as the pulse duration. The pulse duration to breakdown the oxide layer decreases with an increase in the voltage amplitude [2]. Therefore, the protection circuits should respond to an ESD event instantaneously so that it can clamp the incoming voltage to safe levels before it destroys the gate oxide.

There is a fundamental difference between designing the protection circuits for the input and output pins. Typically, the input and output pins are connected to gates and drains of the MOSFETs respectively. Designing the protection circuits for the output pins is more constrained because the excess resistance and capacitance is more prominent in this case and it may exceed the output signal specifications. Typically, the output buffer is utilized for the ESD protection and its sizing is limited by the chip specifications. As a result, the ESD protection of the output pins is usually achieved by the careful layout of the output buffers.

To protect the input pins, the most basic circuit employs reverse biased diodes between the input pin and the power supplies as explained in section 2.2 (shown in the figure 2.2). Since the diodes are in reverse bias under the normal conditions, their parasitic capacitance is minimal. Adding a diffused resistor in series can enhance the ESD protection strength of this circuit manifold. The diffused resistor forms a distributed diode along with a distributed resistance, which reduces the gate voltage at the input MOSFETs. The diffused resistor is more effective than a poly-silicon resistor in distributing the large current during an ESD event. This series resistor slows down the transient of the ESD event. However, the series resistor is in the signal path and it slows down the core circuitry too. Therefore, the value of the series resistance should be carefully traded-off between the ESD robustness and the performance of the IC.

Although the diode protection circuit is simple to implement, it is not suitable in the advanced deep-submicron technologies due to the reasons explained below:

- The reverse bias dynamic resistance of a diode can be too large to clamp the voltages to the safe levels for typical values of the current unless the diode area is very large. For example, a diode of 250μm^2 area with a typical impedance of 5000Ω-μm^2 has a resistance of 20Ω. For a typical ESD stress current of 1A, this diode will develop a voltage of 20V across it, which is well beyond the breakdown voltage of the gate oxide.
- Increasing the area of the diode can decrease the dynamic resistance but it also increases the parasitic capacitance of the diode, which slows down the IC.
- Decreasing the depletion widths of the diode junction can also decrease the dynamic resistance but it also increases the parasitic capacitance of the diode.
- The diode breakdown voltage can be larger than the gate oxide breakdown voltage in today's smaller technologies.
- Finally, the diode may not breakdown quickly enough to protect a circuit from a fast rising ESD pulse.

Due to the above reasons, other ESD protection circuits have been explored. A CMOS based implementation is shown in the figure 4.1. It is sometimes also referred as the grounded-

gate NMOS (ggNMOS) that is described in the section 2.3. The output buffer also acts as an ESD protection circuit. On the input side, the transistors of the protection circuit are OFF under the normal conditions. During a positive ESD pulse with respect to V_{SS} at the I/O pads, the drain-substrate junction becomes forward biased and conducts a large current.

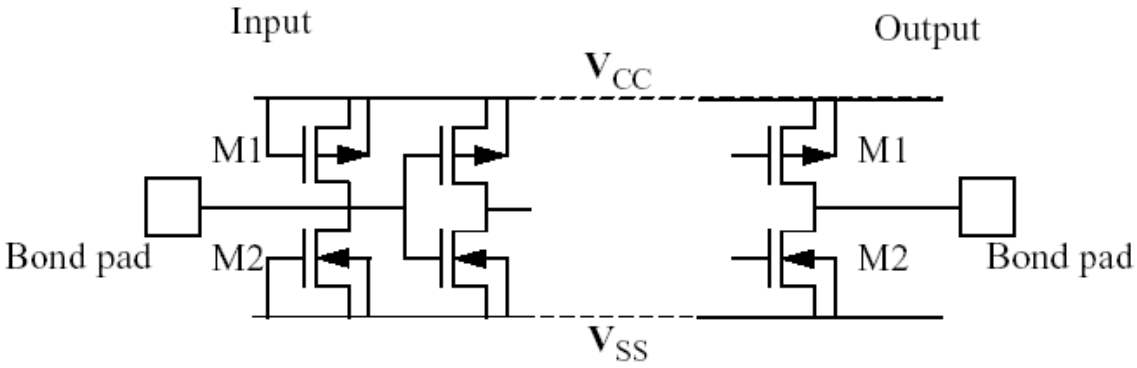


Figure 4.1 CMOS input and output ESD protection circuit

Similarly, for a negative pulse with respect to V_{SS} turns on the parasitic n-p-n bipolar transistor in the NMOS device (snapback mode) as described in the section 2.3. The PMOS transistor works in the same manner for the ESD stress with respect to the V_{CC} pin. In deep submicron devices, the conduction may occur due to punch through instead of the bipolar action, which reduces the value of the V_{t1} (refer to figure 2.3). However, the snapback mechanism still dominates at higher currents and thus the holding voltage (V_h) remains the same. If the power supplies are connected i.e. the IC is powered up, V_{SS} and V_{CC} form ac grounds. Therefore, during a transient stress, the NMOS and PMOS transistors appear parallel from the I/O pads to the power supplies. In this scenario, both the devices may conduct during an ESD event (one by drain-substrate conduction and the other by the bipolar action). However, it is reported that the NMOS device conducts before the PMOS device regardless of the polarity of the ESD pulse [20]. This implies that during a positive pulse, the parasitic n-p-n transistor in the NMOS conducts before the drain-substrate junction in the PMOS transistor turns on. On the other hand during a negative pulse, the drain-substrate junction in the NMOS transistor turns on before the parasitic p-n-p transistor in the PMOS conducts. This can be explained by the fact that the current gain of the parasitic n-p-n bipolar transistor in the NMOS is much larger than that of the parasitic p-n-p bipolar transistor in the PMOS due to the lower hole-diffusivity in the substrate. This means that the NMOS transistor goes into the snapback mode at a lower current. However, the PMOS transistors are needed for the ESD protection when the power supply pins are not connected.

Typically, the voltage that triggers the snapback in a MOSFET (V_{t1}) is 2-3V higher than the gate oxide breakdown voltage depending on the channel length of the transistor. However, the transistor can be pushed into the snapback mode at a lower temperature using a technique commonly referred as gate-bouncing. This is achieved by adding a resistor between the gate of the NMOS to the negative power supply (or ground) as shown in the figure 4.2. Similar arrangement can be done for the PMOS as well.

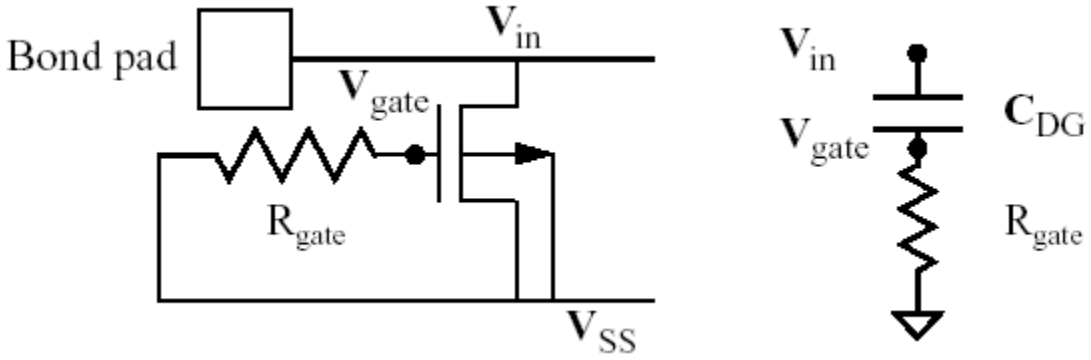


Figure 4.2 Gate-bouncing technique to trigger the NMOS at a lower voltage

During a positive ESD pulse, the voltage transient at the bond pad (V_{in}) is coupled to the gate (V_{gate}) by the drain-gate overlap capacitance. This transient voltage (V_{gate}) turns on the transistor temporarily that helps the snapback to occur at a lower voltage (V_{t1}). This transient voltage (V_{gate}) eventually decays to V_{SS} by a time constant of R_{gate} C_{DG} (as shown in the figure 4.2). Often a field oxide MOSFET is used with its gate connected to the drain. The threshold voltage of the field oxide device is much higher than the normal operating voltages of the circuit and it remains OFF under the normal conditions. During an ESD event, the field oxide transistor turns on and eventually goes into snapback clamping the voltage to a safe level. An additional advantage of using the field oxide device is that its gate oxide has a higher breakdown voltage and therefore it is more robust to the ESD stress. A combination of thin-oxide and field-oxide devices can also be implemented to achieve dynamic gate coupling as shown in figure 4.3 [21][22]. The source terminals of both the devices are tied to ground. The drain and gate pins of thin-oxide device are connected to the gate and drain terminals of field-oxide device respectively. The input is coupled to the gate of the thin-oxide device by the gate-drain overlap capacitances of both the transistors. As the input voltages reaches the threshold voltage of the field-oxide device, it turns on and discharges the thin-oxide gate to ground. It is important to turn off the thin-oxide

transistor once it enters the snapback mode because the normal MOSFET conduction takes place on the surface of the substrate and high currents confined at the surface may lead to premature thermal breakdown. In this technique, the amount of gate coupling can be controlled by the ratio of the gate widths of the two devices.

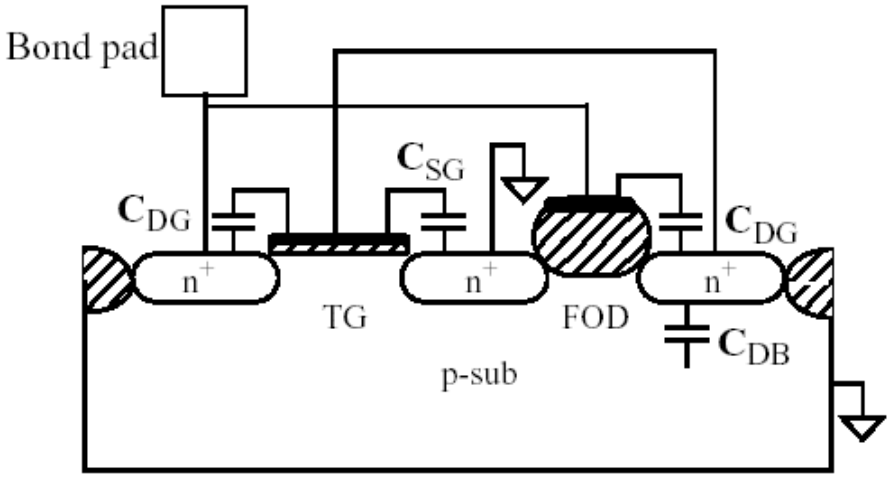


Figure 4.3 Dynamic gate coupling technique

Most of the modern ESD protection circuits are implemented in two or more stages. The first stage consists of the primary elements that are slow in response but conduct a large amount of current and are more robust to an ESD event. On the other hand, the second stage contains secondary elements that respond quickly to an ESD event but they exhibit limited current carrying capability. Following the above guidelines, an ESD protection scheme using ggNMOS devices is shown in figure 4.4.

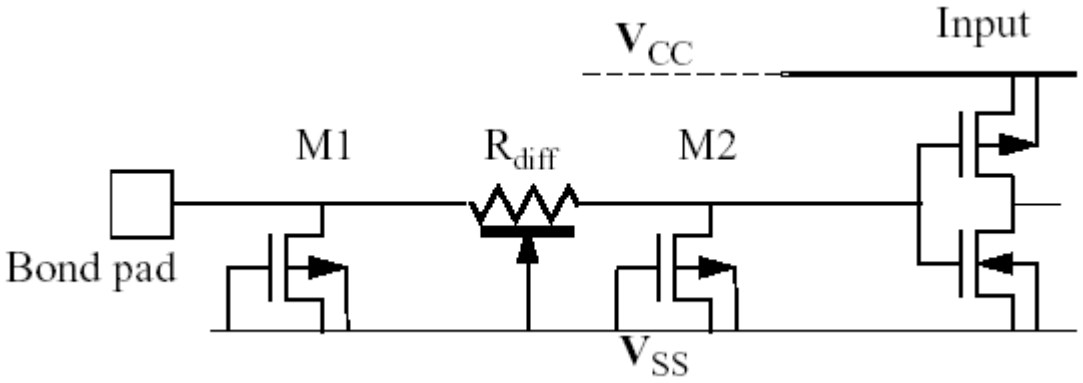


Figure 4.4 Two-stage ESD protection circuit using ggNMOS and diffused resistor

This scheme consists of a long-channel NMOS transistor (M1), a diffused resistor (R_{diff}) and a short-channel NMOS transistor (M2). The major portion of the turn-on time of a parasitic bipolar transistor is the delay transit time [21], which is given by L^2 / 2D, where L is the channel length and D is the diffusion coefficient. Therefore, smaller device (M2) will respond much faster than the larger device (M1). The diffused resistor develops a voltage drop that protects the M2 from large voltages and turns on the M1. Similarly, the PMOS devices can also be added to the circuit to protect it from an ESD event when the power supplies are not connected.

Another popular configuration for ESD protection circuits in CMOS is substrate-triggered NMOS (stNMOS) as shown in figure 4.5 [23].

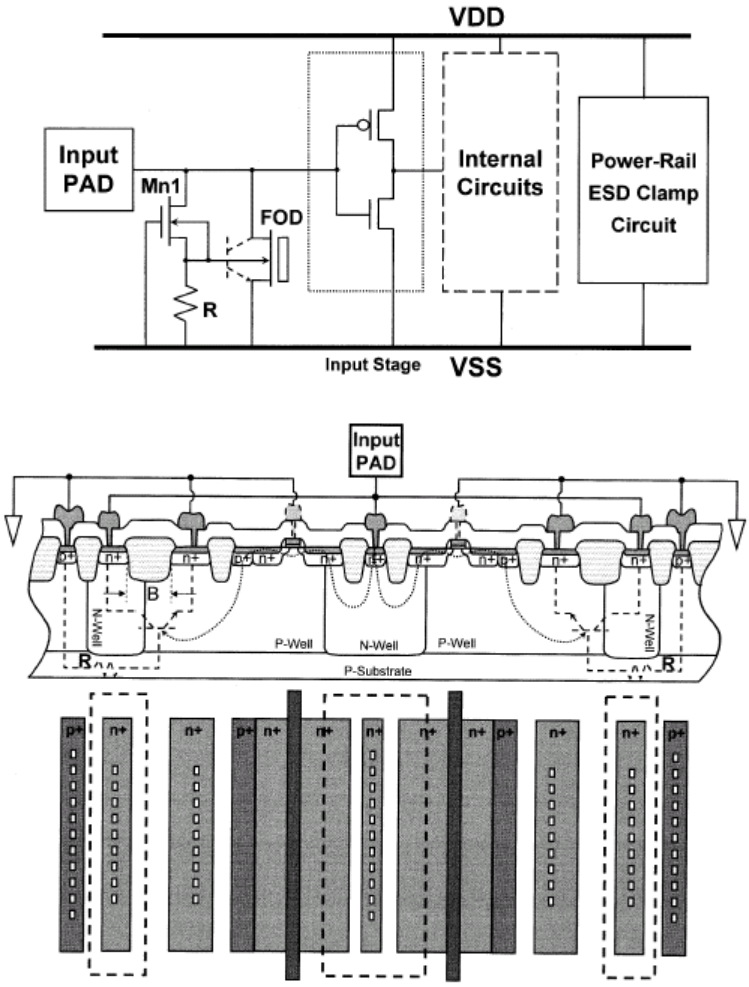


Figure 4.5 Substrate-triggered NMOS technique for ESD protection: circuit schematic, cross section and layout [23]

During an ESD event the short-channel transistor Mn1 turns-on quickly and resulting current develops a voltage drop across the resistor R. This voltage turns-on the parasitic bipolar transistor buried in FOD device by forward biasing the base-emitter junction. The bipolar conduction sinks a large amount of current without any permanent damage to the device and the internal circuits. The cross section of the circuit shows that the bipolar transistors are realized using n/p wells and the resistors are realized using the substrate resistance. This scheme can be implemented even for a regular n-well process since the substrate biasing is transient (occurs during an ESD event only). In the normal operating conditions, the grounded-gate transistor Mn1 is OFF and there is zero substrate bias at the FOD.

5. Non-uniform bipolar conduction

Typically, ESD protection schemes implement the devices in multi-finger fashion so that a wider device can be fabricated in a relatively small silicon area. To turn-on majority of fingers, ballast resistance is added at the drain of each finger. Most of the schemes concentrate on turning-on majority of the fingers. It is generally assumed that whole width of a finger carries ESD current if the finger turns-on. However, in the advanced deep-submicron CMOS technologies a non-uniform current conduction is reported even for a single finger [24]. It is observed that only part of the width is utilized for current conduction. A single finger of an NMOS device can be visualized as several narrow (small width) n-p-n bipolar transistor segments connected in parallel as shown in figure 5.1 [25].

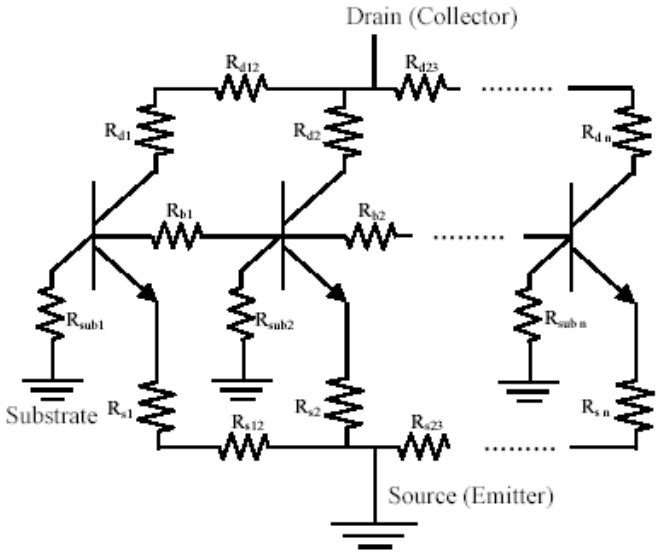


Figure 5.1 Circuit equivalent of a single finger NMOS transistor [25]

Intrinsic characteristic of each bipolar segment is different due to the statistical random distribution of the dopant fluctuations. Therefore as soon as some of the segments conduct, the common drain voltage is clamped to the holding voltage (V_h) of the conducting segments and the remaining segments may not turn-on. Increasing the width of the finger increases the variations among the segments and hence the non-uniformity in the conduction. Figure 5.2 shows that increasing the width in both silicided and non-silicided devices, the second breakdown current per unit width (I_{t2}) reduces. The non-uniformity is more severe in the silicided device. Figure 5.3 shows the variations of I_{t2} with the finger width for silicided and non-silicided devices.

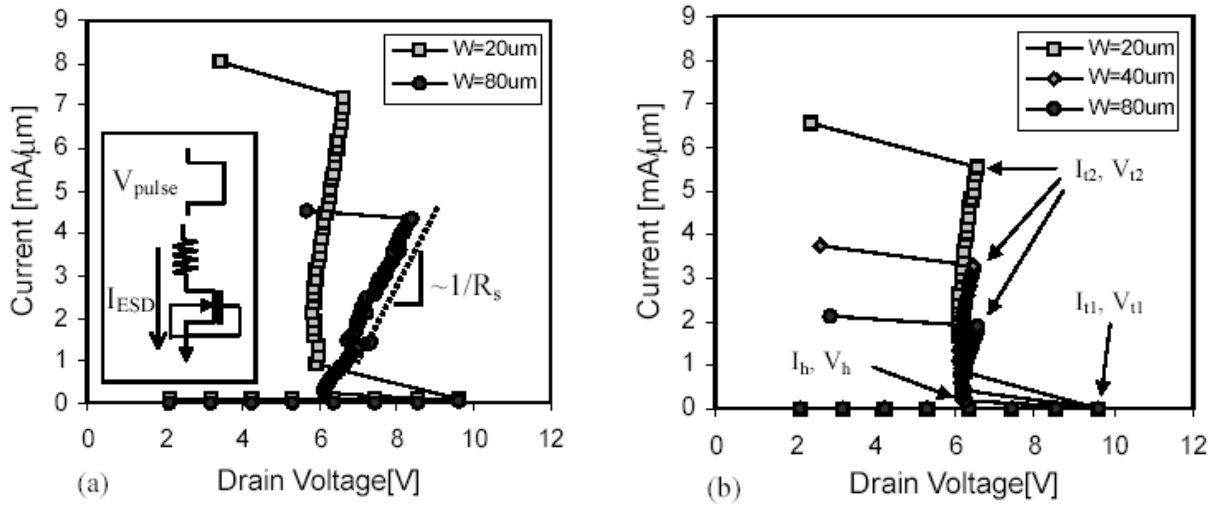


Figure 5.2 I-V characteristics of (a) non-silicided and (b) silicided devices of different widths [25]

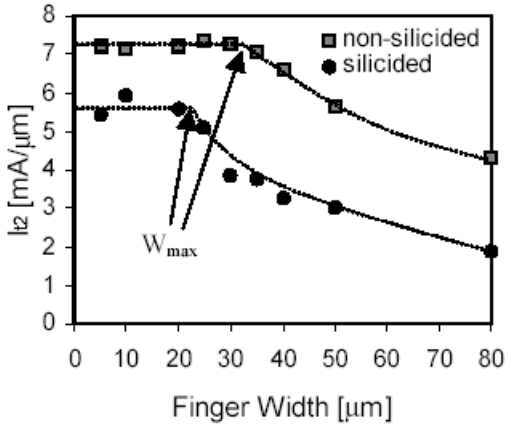


Figure 5.3 I_{t2} variations of non-silicided and silicided devices with the finger widths [25]

For low values of W, an increase in finger width does not affect the I_{t2} (i.e. the total second breakdown current $I_{T2} = I_{t2}$ x W increases). However, I_{t2} decreases for larger values of W. The I_{t2} roll-off point is shown in figure 5.3 by $W=W_{max}$. That is increasing the finger width beyond W_{max} does not increase the total second breakdown current (I_{T2}). It should be noted in figure 5.3 that the W_{max} of the non-silicided devices is higher than that of the silicided devices. It confirms the higher non-uniformity in the silicided devices.

5.1. Substrate bias effect

Substrate bias increases the current conduction uniformity of an NMOS finger. A positive substrate bias increases the base-emitter forward biasing in an NMOS finger and transistor goes into bipolar action with a very small Avalanche multiplication current from the drain. In other words, a positive substrate bias facilitates the faster bipolar conduction at much smaller drain currents. Figure 5.4 shows the increase in I_{t2} with substrate bias [25].

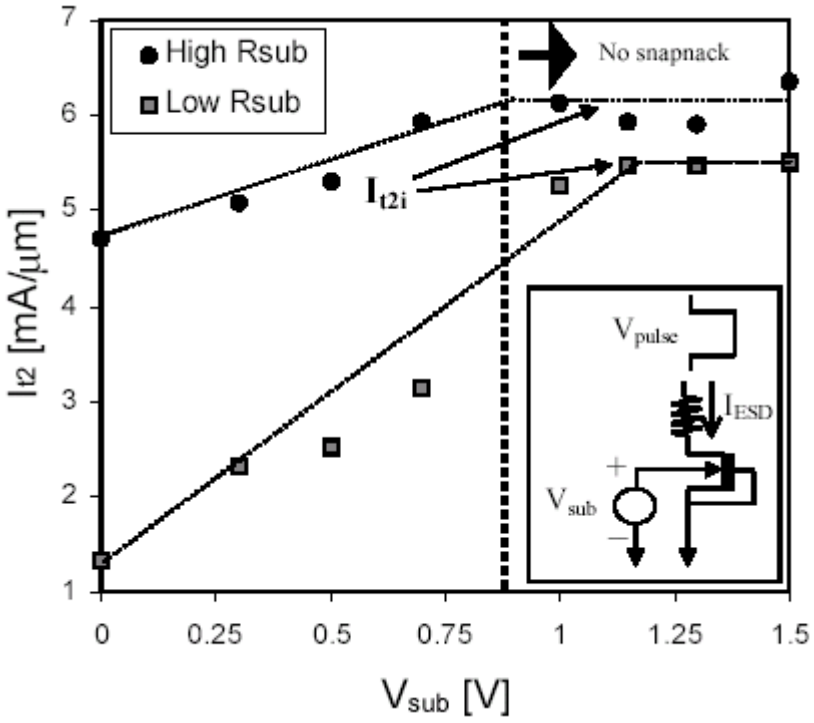


Figure 5.4 I_{t2} increases with positive substrate voltage and eventually saturates [25]

However, I_{t2} saturates at higher values of substrate voltage. This value of I_{t2} is called intrinsic second breakdown current (I_{t2i}) for the technology under consideration. Further increasing the substrate voltage does not give any improvement in the i.e. the local base-emitter junction voltage

cannot be increased by any external voltage. The external substrate voltage can reduce the non-uniformity in the base-emitter junction voltage along the width of the transistor. Figure 5.5 shows that the substrate bias improves the effective width of the NMOS current conduction under ESD.

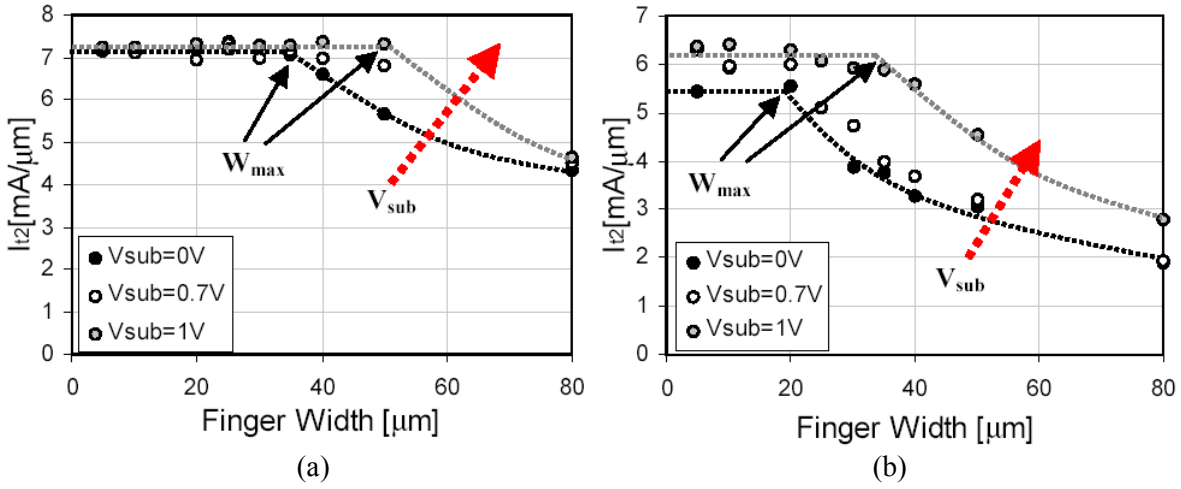


Figure 5.5 I_{t2} and W_{max} improvement with positive substrate bias for (a) non-silicided and (b) silicided devices [25]

It should be noticed that for the non-silicided devices with finger widths less than W_{max}, I_{t2} is almost independent of V_{sub}. This is probably due to the fact that for the finger widths less than W_{max}, the current conduction is already uniform and V_{sub} does not give any improvement. For widths higher than the current conduction is non-uniform and V_{sub} improves I_{t2} by reducing the non-uniformity. On the other hand in the silicided devices, majority of the current is localized to the surface (silicide layer). Therefore on the application of V_{sub}, I_{t2} improves because the current conducts deeper into the substrate that increases the total available volume for power dissipation and hence better ESD robustness.

5.2. Gate bias effect

Gate-coupling technique is popular in multi-finger structures to achieve better uniformity in turning on the fingers. However in the advanced salicided deep submicron technologies, gate coupling should be used carefully because excessive coupling may cause current localization near the surface (gate-oxide/substrate interface). Due to low thermal conductivity of the oxide layer this surface current results in high temperature regions (hot spots) causing a reduction in the ESD hardness. It is observed that the gate-bias affects the narrow and wide width fingers differently. For narrow finger width, I_{t2} degrades with increasing gate bias. On the other hand, gate bias

improves the I_{t2} of wider fingers as shown in figure 5.6 [26][27]. This can be explained by the fact that the current conduction in wide fingers is non-uniform and gate bias reduces this non-uniformity by facilitating the bipolar action. In the narrow fingers, current conduction is already uniform and gate bias only degrades I_{t2} by causing current localization at the surface. Figure 5.7 shows the variations in the value and location of the peak temperature with gate bias. As the gate bias increases the current conduction takes place closer to the Si/SiO_2 interface. This brings the peak temperature location near the interface. The temperature also increases slightly due to low thermal conductivity of SiO_2.

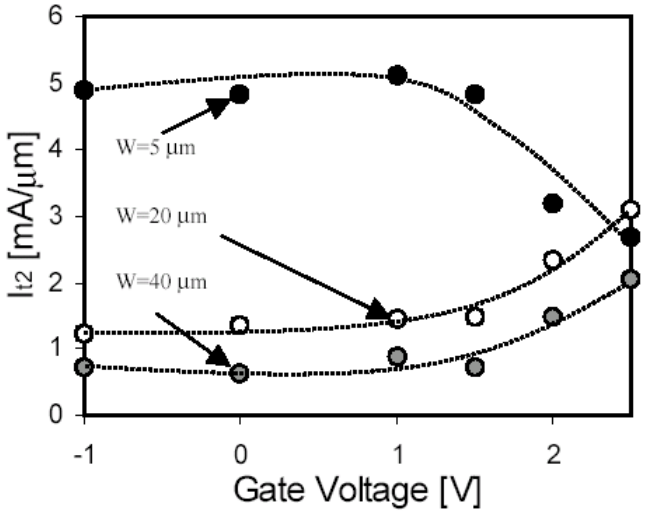


Figure 5.6 Influence of finger width on the gate voltage dependence of ESD robustness [27]

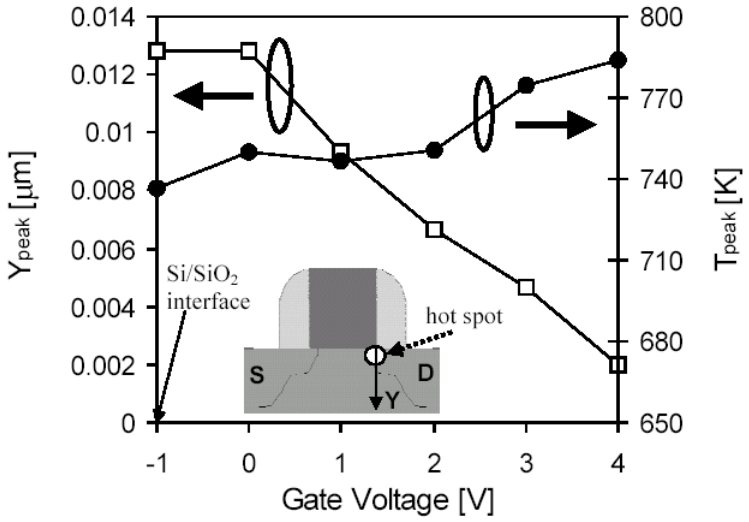


Figure 5.7 Variations of the value and the location of peak temperature with gate bias [27]

5.3. Effect of gate-to-contact spacing

It is widely observed and reported that an increase in the gate-to-drain contact spacing improves the ESD robustness of the non-silicided devices by increasing the ballast resistance. On the other hand increasing the gate-to-contact spacing hardly shows any significant improvement in the silicided devices even in the advanced 0.18 µm CMOS technology [28]. However, it is reported very recently that in salicided 0.13 µm CMOS technology (from TI) increasing the gate-to-drain contact spacing (GDCS) and the gate-to-source contact spacing (GSCS) significantly improves the ESD hardness of the device [29][30]. This can be explained by the fact that technology scaling exhibits higher process variations due to shorter gate length and shallower junctions. These process variations cause severe non-uniformity in the current conduction. Increasing the GDCS/GSCS reduces the non-uniformity and therefore increases the ESD hardness of the device as shown in figure 5.8.

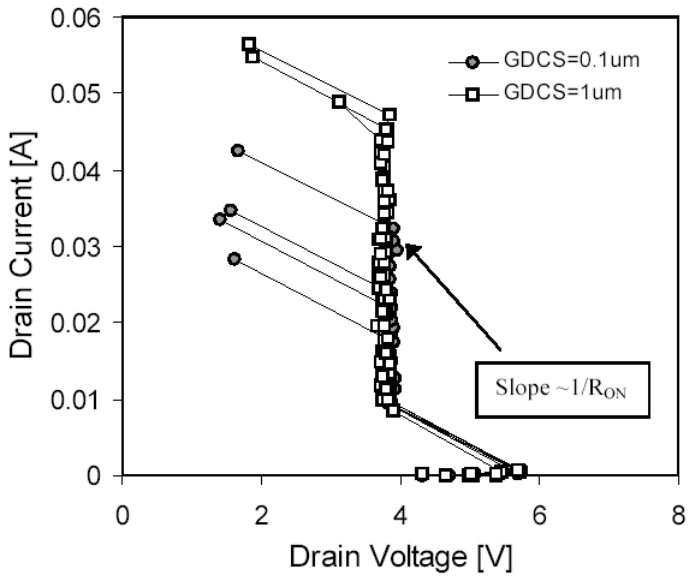


Figure 5.8 Improvement in the second breakdown current with an increase in GDCS [30]

It should be noticed in figure 5.8 that the on-resistances (R_{ON}) for the two values of GDCS are almost the same and the improvement in the ESD hardness is believed to be due to better uniformity in the current conduction across the finger width. It is also observed that the gate-to-contact spacing dependence of I_{t2} vanishes on the application of a substrate bias as shown in figure 5.9. This also confirms that the non-uniformity in the current conduction is responsible for the gate-to-contact spacing dependence of I_{t2}. Increasing the gate-to-contact spacing also increases the volume available to dissipate power that further improves the ESD robustness. In

0.13 µm CMOS technology devices are isolated by shallow trench isolation (STI) with an overlap length of n+ source/drain contact from STI. It is observed that increasing the S/D overlap length improves the total second breakdown current (I_{T2}) as shown in figure 5.10. Increasing the S/D overlap length increases the thermal volume that enhances I_{T2}. For the S/D contacts near STI, heat is enclosed in a smaller region due to thermal isolation caused by the STI. This further degrades the ESD hardness of the devices with smaller S/D overlap lengths.

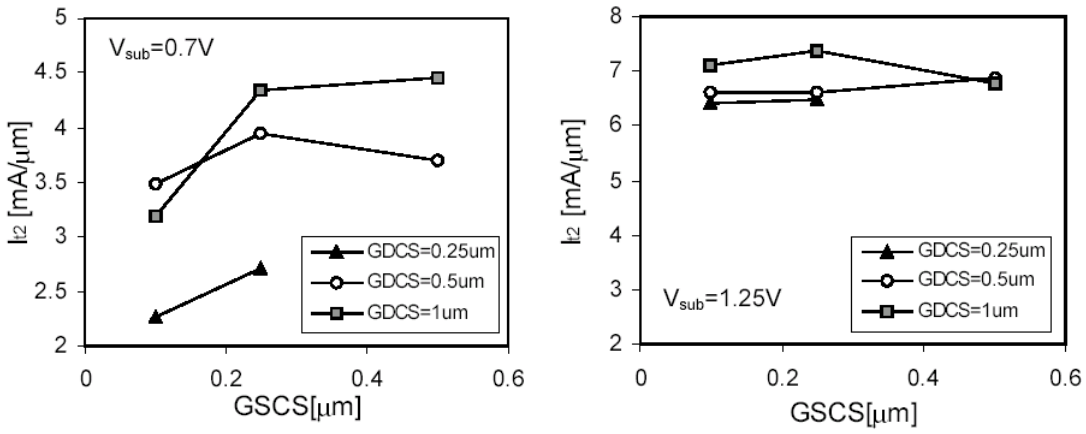


Figure 5.9 Influence of the substrate bias on the GSCS dependence of I_{t2} [30]

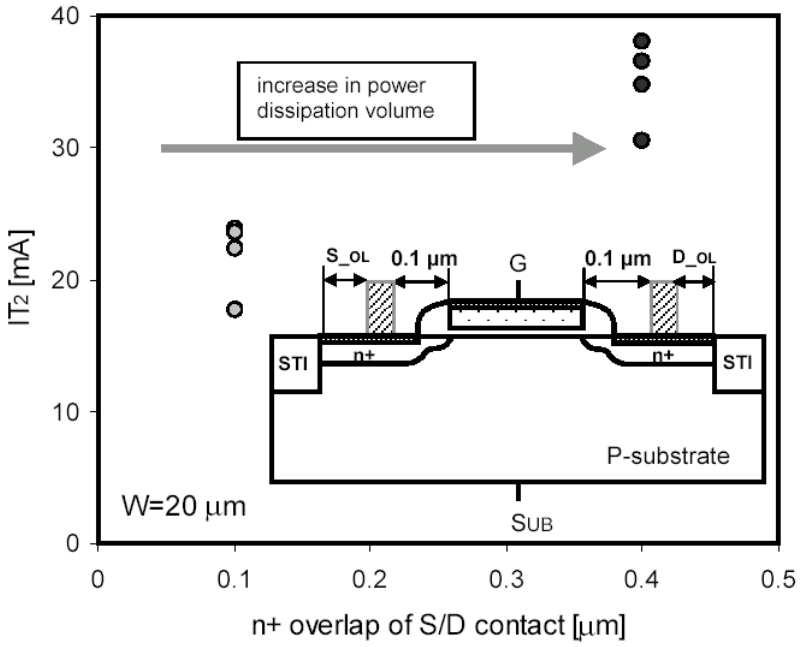


Figure 5.10 Total second breakdown current increases with the n+ S/D overlap [30]

5.4. Effect of gate length

The gate length is decreasing with every new generation of the technology that decreases the equivalent base width of the parasitic bipolar transistor and hence improves the current gain β. The current gain (β) depends on the gate length (L) by the following relation [31]:

$$\beta = \coth(L)$$

To the first order, it simplifies to $1/L^2$. It suggests that as the gate length decreases, the current carrying capability of the bipolar transistor improves i.e. the bipolar transistor carries larger ESD current for smaller substrate current and hence lower power dissipation. Therefore the ESD performance is expected to improve with the technology scaling. However, this trend has been consistent only for the non-silicided devices. In the advanced salicided CMOS technologies, the I_{t2} is reported to degrade with the gate length as shown in figure 5.11 [32].

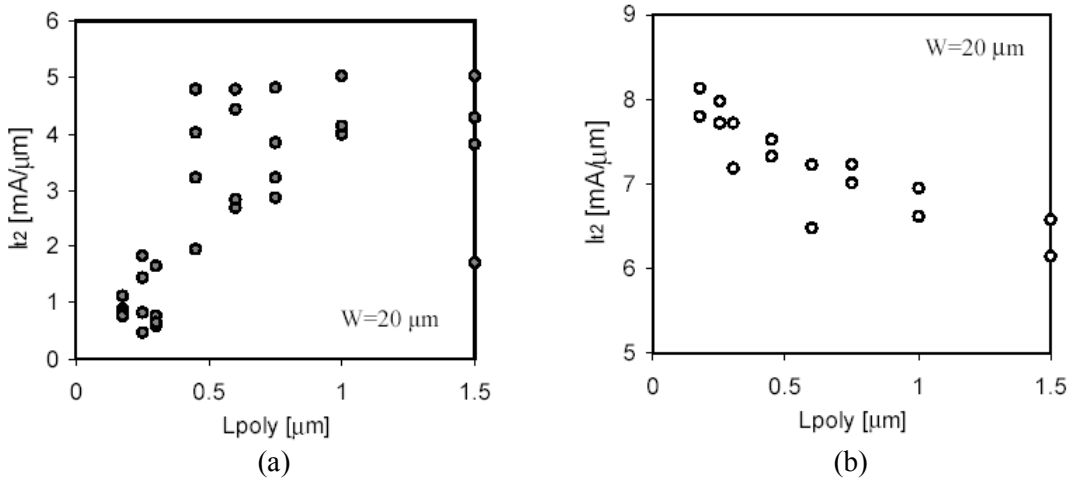


Figure 5.11 Gate length dependence of I_{t2} for (a) silicided and (b) non-silicided devices [32]

It is observed that the current gain improves with the technology scaling for both silicided and non-silicided devices confirming the bipolar current gain model. It is believed that the severe non-uniformity in current conduction across the finger width is responsible for this peculiarity. Smaller dimensions and surface conduction in advanced salicided technologies exhibit higher process variations that deteriorate the non-uniformity. This reverse gate length dependence vanishes for appropriate value of substrate bias as shown in figure 5.12 [32]. This confirms the major contribution of the non-uniformity in the reverse gate length dependence. Reduction in gate

length also reduces the available volume for power dissipation that further worsens the ESD performance of the devices. Figure 5.13 shows the temperature distribution across the channel in a short channel device. As expected, smaller gate length exhibits higher localization of the peak temperature i.e. the peak temperature is limited to a very small region that causes excessive localized heating and hence permanent damage in the device at lower ESD levels. In the relatively larger devices (larger L), the temperature distribution is wider and hence these devices are more robust to the ESD events.

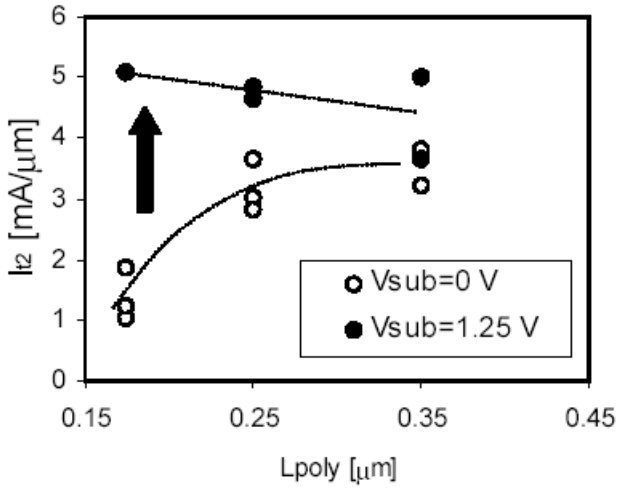


Figure 5.12 Influence of substrate voltage on reverse gate length dependence [32]

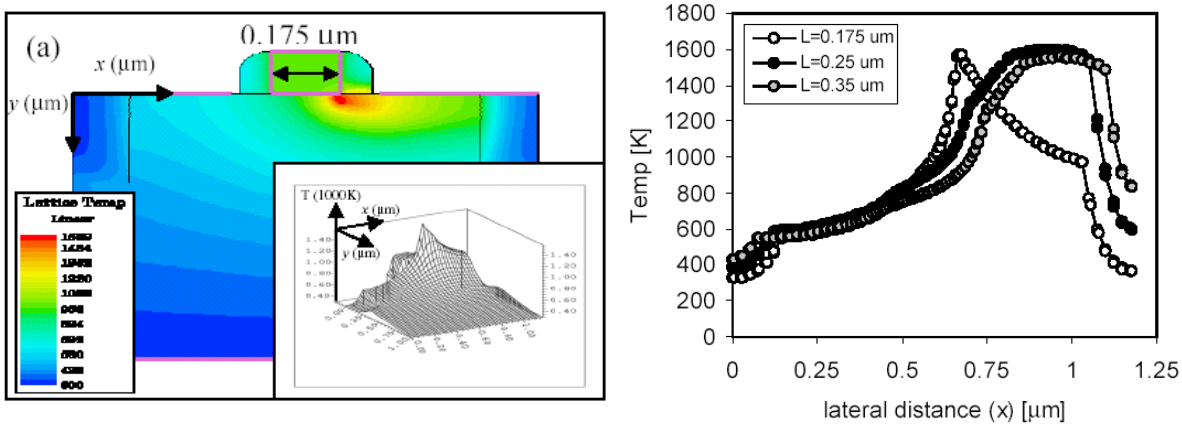


Figure 5.13 The temperature distribution along the channel (x) at y=0.05 µm [32]

6. Simulation Methods and Applications

Traditionally ESD circuits' designs are heavily dependent on measurement. These measurements are essentially destructive since purpose is to find the ESD withstand voltage and current limit until device fails permanently. Given the high voltage ESD pulse and high lattice temperature of the device near thermal failure, the conventional device models such as BSIM used for logic design cannot be used for ESD simulation. In this chapter we would discuss the evolution of device models under ESD stress, 2-D device simulator like Medici, and eventually full-blown mixed-mode ESD circuit simulation. Specific application of these simulations such as extraction of MOSFET I-V and P_f–t_f parameters would also be presented.

6.1. Lattice Temperature and Temperature Dependent Models

In ESD model, of semiconductor device, failure is defined as the time at which temperature at the hottest point of the device reaches a critical value, T_c. This critical temperature is defined as the melting temperature of silicon, more accurately, the temperature at which the intrinsic carrier concentration exceeds the doping level i.e. the onset of second breakdown. At the beginning, temperature gradient in the box changes in all directions until thermal equilibrium is reached. The typical heat flow equation at time t is given by [33]:

$$\rho C_p \frac{\partial T}{\partial t} = H + \nabla(k(T)\nabla T) \qquad 6.1$$

Where ρ is the density, k is the thermal conductivity, C_p is the specific heat capacity all assumed to be independent of temperature in this mode. In the steady state temperature distribution becomes constant since heat source H is constant.

Unlike devices at nominal voltages, devices at ESD voltages conduct very high current and this gives rise to higher level of heating, in addition due to uneven thermal conductivity of different layers in the device the temperature doesn't remain constant across the width and length of the device and the substrate. So the classic thermal box model for heat flow is no longer valid for electro-thermal simulations. If we couple the heat flow equation with Poisson's equation, the electron/hole current density equations, and the electron/hole continuity equations it results in following equation for heat generation term H (W/cm^3) of equation 6.1[34]

$$H = J_n \,.\, E + J_p \,.\, E + H_U \qquad 6.2$$

where E is the electric field, Jn and Jp are the electron and hole current densities, respectively, and H_U is the recombination contribution and is expressed as

$$H_U = [R - G].E_g \qquad 6.3$$

where R is the recombination term and G is the impact-ionization generation rate, and E_g is the band-gap energy. These parameters are the function of lattice temperature. Due to spatial variation of lattice temperature, Poisson and current-density equations need to be modified. Modified Poisson's equation is given as [34]

$$\bar{\nabla}.\varepsilon\bar{\nabla}(\psi-\theta) = -q(p - n + N_D^+ - N_A^-) - \rho_F \quad 6.4$$

where ψ is the electrostatic potential, q is the electron charge, p and n are the hole and electron concentrations, respectively, ε is the permittivity, N_D^+ and N_A^- are the ionized impurity concentrations, ρ_F is the fixed-charge density, and θ is the band structure parameter.

Putting additional thermal-diffusion terms in the current density equations we get [35]

$$J_n = qn\mu_n E + k\mu_n(T\bar{\nabla}n + n\bar{\nabla}T) \quad 6.5$$

$$J_p = qp\mu_p E - k\mu_p(T\bar{\nabla}p + p\bar{\nabla}T) \quad 6.6$$

where μ_n, and μ_p are the electron and hole mobilities, respectively.

As discussed the lattice temperature is no longer constant across the device, so the models must be based on the local temperature. The model used for this is Lombardi surface mobility model which accounts for parallel and perpendicular field and also incorporates temperature dependence. It is a semi-empirical model given as [36]

$$\mu_{sr} = \frac{DN}{E_\perp^2} \quad 6.7$$

$$\mu_{ac} = \frac{BN}{E_\perp} + \frac{CN.N_{total}^{EN}}{T\sqrt[3]{E_\perp}} \quad 6.8$$

$$\mu_b \neq function(T, E_\perp) \quad 6.9$$

where N_{total} is total doping concentration at particular point in the device, T is the local temperature, $E_\perp$ is the local perpendicular electric field, and BN, CN, DN and EN are coefficients which have different values for electrons and holes. These mobility terms when added in parallel give the overall mobility at each point.

Putting thermal electrodes along the edge of the device creates thermal boundary conditions and it acts as infinite current sink by enforcing a constant temperature at the contact. It is assumed that there is no heat flow across the non-contacted edges. Lumped linear thermal resistance and capacitance are placed to simulate the conduction of heat away from the part of the device.

The modeling here is based on the assumption that local electron and hole temperatures are equal to the local lattice temperature. As we know high electric field generated hot carriers

with characteristic temperature are higher than the lattice temperature. This high temperature is responsible for phenomena like impact-ionization current generation, mobility degradation etc. This requires the model based on local field instead of the model based on local temperature. Thermally electrons and holes are considered separate entity from the lattice. Although the heat capacity of electrons and holes has been found smaller than silicon lattice, it is actually the relative thermal conductivity of the lattice that matters. Thermal conductivity determines the thermal flux. For carriers flux is defined as product of the carrier thermal conductivity, K_c, and the gradient of the carrier temperature, similarly thermal flux for lattice is defined as the product of lattice thermal conductivity, K, and the gradient of lattice temperature. K_c is defined as [37],

$$k_c = \frac{3}{2} nk^2 T_c \mu_c / q = \frac{3}{2} nkD_c \qquad 6.10$$

where μ_c and D_c are carrier mobility and carrier diffusion constant, respectively.

In addition to carrier diffusion current, carriers also contribute towards heat conduction by virtue of heat current due to electric current. This component is given by the following equation [37],

$$S_{n,j} = \frac{3}{2} \frac{kT}{q} J \qquad 6.11$$

where J is current density.

The electron temperature can be found by [38]

$$qEv_n^{sat} = \frac{3}{2} k(T_e - T)/\tau \qquad 6.12$$

Where E is the electric field, T is the lattice temperature and τ is the energy relaxation time of electrons in silicon. From equation 6.12 we get the average electron temperature is approximately 5300K, this gives thermal conductivity of ~54 mW/cm-K. This value is far smaller than silicon lattice thermal conductivity of ~0.31 W/cm-K at 1000K [34]. The contribution in heat flux due to carrier diffusion is 1%. As per equation 6.11 contribution in heat flux due to current conduction is 5%. This analysis suggests that the equilibrium assumption between lattice and carriers leads to an approximately 6% underestimation of thermal dissipation in devices under ESD stress. In light of other unpredictability in simulation 6% error is reasonable.

6.2. Curve Tracing

By incorporation of thermal-diffusion equation and temperature dependent mobility and impact ionization model it is possible to simulate MOSFET snapback characteristics Fig. 6.1 under ESD stress. However this snapback curve is complex in the sense that it has flat and steep regions. Moreover there are regions where slopes change its sign. In the flat/steep regions

current/voltage changes little with the voltage/current. In regions where slope changes sign we have multi-valued voltage solutions. It has been found that a voltage boundary condition on the electrode being swept is stable in the region where current doesn't change much and vice versa. So in order to simulate the snapback characteristics with traditional method, the MOSFET boundary condition needs to be changed according to the region of the curve.

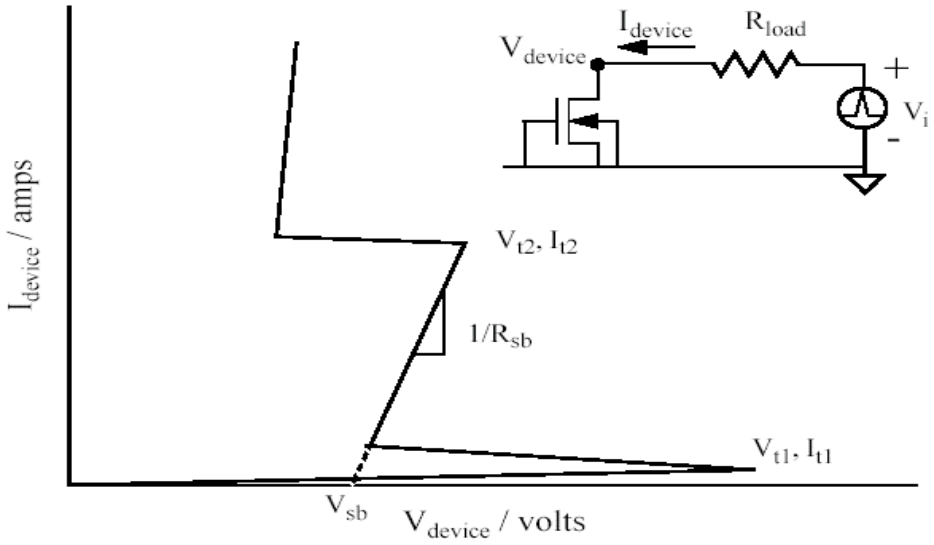


Figure. 6.1 The MOSFET snapback characteristics qualitative diagram

Using voltage boundary condition during the initial reverse bias of drain-substrate junction, then, switching to the current boundary condition when the current changes rapidly, can simulate the MOS snapback characteristics. After the junction breakdown again changing to the voltage boundary condition and stepping the voltage negatively in negative slope region. Then, again changing it to the current boundary condition in the snapback mode. This process is time consuming and requires prior knowledge of the curve characteristics, since simulator must know when to change the boundary condition. One way to get over this problem is to use high load resistor at the drain of transistor to remove negative slope multi-valued solutions region. But the condition on this resistor is that it should be greater than the differential resistance at any point of I-V characteristics. This again requires a priori knowledge of the MOSFET characteristics. The general solution to the curve-tracing problem is to dynamically change the boundary condition from voltage to current by using a voltage or current source with external load that ensures the convergence throughout the trace. Here we don't need prior knowledge of curve instead it relies on easily available quantities like voltage current and slope at each solution point. These quantities are directly available in device simulators from Jacobian matrix.

6.3. Mixed Mode Simulation

Many device simulators provide the option of connecting resistor and capacitor with the electrodes of the defined 2-D device and external ground which allows the simulation of the effect of parasitics in the device. Recently, interfacing of one or more numerically simulated device with spice like circuit simulator has been made possible. This paves way for SPICE like circuit simulation even at ESD voltage and current level. The full circuit can be solved either in coupled manner where the semiconductor equations describing the devices and the Kirchoff's equation describing the circuit are solved together or in decoupled manner where interface is created between device and circuit simulators and each does iteration in succession [39,40].

Mixed mode simulation facilitates the study of ESD circuits in simulation as opposed to going through destructive measurement. This is also useful for transient modeling of ESD tests such as HBM, MM, CDM, and TLP. Gate-bounce, substrate-bias, multiple fingers effect can be studied by merely adding lumped resistors in the circuit. Figures 6.2, 6.3 represent the gate-bounce, ballast resistor, and multiple finger simulation circuit. Slight layout variation is introduced for the simulation of non-uniform triggering due to random process variation in multiple finger ESD simulation. The simulation has limitation since it doesn't account for the thermal coupling among fingers.

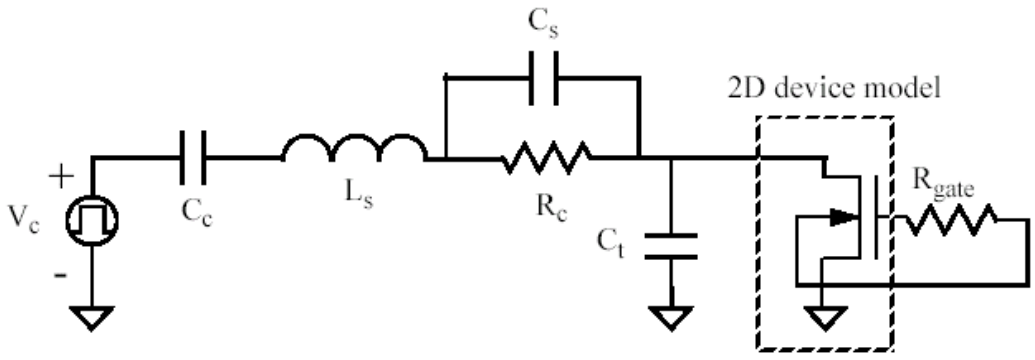


Figure. 6.2 Gate-bounce simulation setup

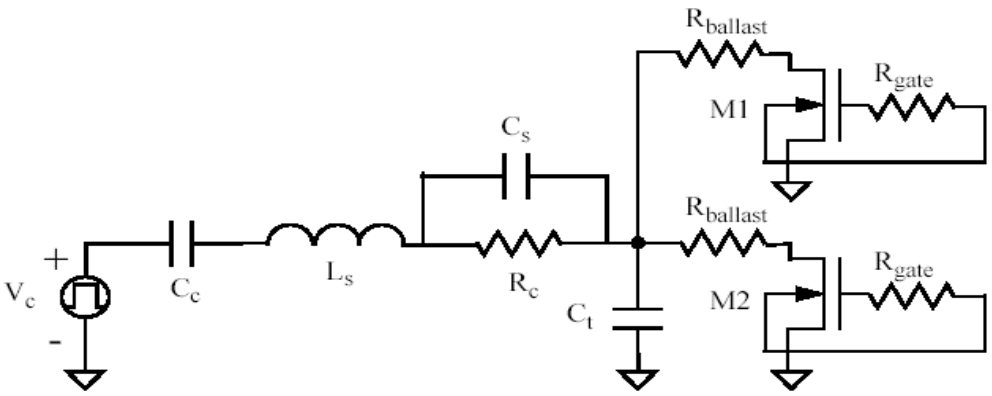


Figure. 6.3 Simulation setup for ballast-resistor along with gate-bounce in multi-finger structure

6.4. Extraction of MOSFET I-V parameters

Generating snapback I-V curve of the device is important to study its ESD behaviour. The first trigger point provides the maximum allowed voltage/current at drain before ESD circuit turns on. The snapback characteristics, determines what the input voltage would be when the given current is flowing through the device. The second breakdown provides the maximum withstand current i.e. ESD withstand level of the circuit. After this voltage/current level the device would undergo thermal runaway and permanent failure. All of these circuit parameters can be extracted from simulations to help expedite the design process without the cost of actually fabricating and then doing measurement till device fails permanently. Different types of curve can be generated from simulation such as: the curve for single transient pulse, curve by series of TLP simulations with increasing voltage level, and dc sweeping curve. Although TLP generated curve in itself gives a lot of information, its matching with dc curve is important. If the input pulse rise time is of the order of ~2 nanoseconds, it suggests there is no coupling of the other electrodes like gate, source or substrate on drain and they are properly grounded. However, when coupling is warranted as in the case of blast resistors at gate just to see how trigger points move TLP simulations are important. The dc sweep doesn't show this gate-bounce effect.

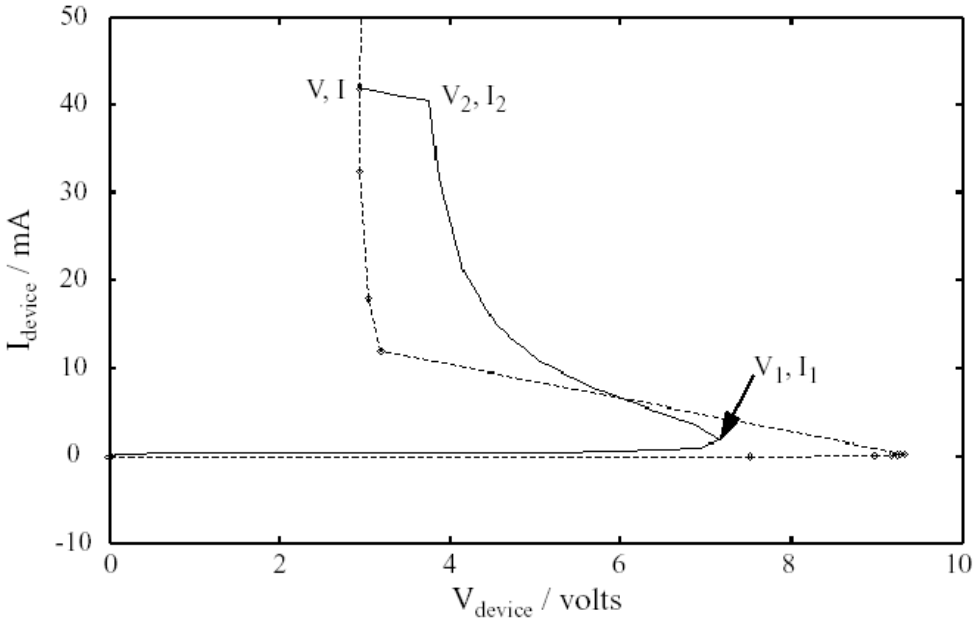


Figure. 6.4 Snapback I-V curve extracted from TLP simulation

In general for study of I-V characteristics of a simulated device first dc curve trace is used to extract the snapback voltage and snapback resistance and then transient TLP simulation is run to extract second breakdown or thermal failure point Fig. 6.4.

Most important application of including thermal-diffusion equation in device simulation is the modeling of device heating which leads to thermal runaway. In the simulation if the device is accurately modeled for heat conduction away from device and mobility, impact-ionization modeling accurately define the current through the device, electro-thermal simulations can accurately predict the time to failure of the device under given stress condition. Thermal failure is a three-dimensional process, since as soon as the hot spot gets created current rushes into the spot from all directions. Normally if current if flowing uniformly through the device then hot spot is created at the center of the device since it is the point of highest temperature. Sometimes thermal runaway may originate at the weak-spot too where the electric field is slightly higher [41].

The 2D device simulator can only model current rushing in from two dimensions only after formation of hot spots so it can't model the thermal runaway accurately. However by seeing the point where device voltage drops due to reduction in resistance the start of second breakdown can be predicted.

6.5. Simulation of Dielectric Failure

In case of ESD stress where high current flows for very short time i.e. CDM kind of ESD stress and protection circuits don't even trigger in the short duration we witness the dielectric

failure of the gate-oxide of the MOSFET. Applicability of the device simulators to dielectric failure is not as apparent as thermal failure still the electric field in the oxide region, lattice-temperature profile in silicon, and the hot-carrier injection current provides a qualitative examination of dielectric. This failure is a threat for both the protection devices and the device being protected but the input circuit device is much more likely to fail before protection circuit device. In the protection transistors basic reason of dielectric failure is the hot-carrier injection resulting from the high ESD current.

Device simulators can model the transport of charge carriers but there is no way to model the movement or melting of silicon lattice since the grid structure used for defining device is fixed. So it is assumed that when the modeled temperature exceeds some value (1688K for silicon) melting would occur [42]. Dielectric failure is defined by: injection of charge into the oxide and high electric-field stress across the oxide. The simplest analysis of dielectric failure is done by extracting electric field and voltage information across oxide which is not readily available from simulation but can be found from files containing potential and electric-field profiles. Figure 6.5 shows the typical simulated curve for electric-field vs. time in the 10nm thick oxide of a protection MOSFET subject to a square wave pulse.

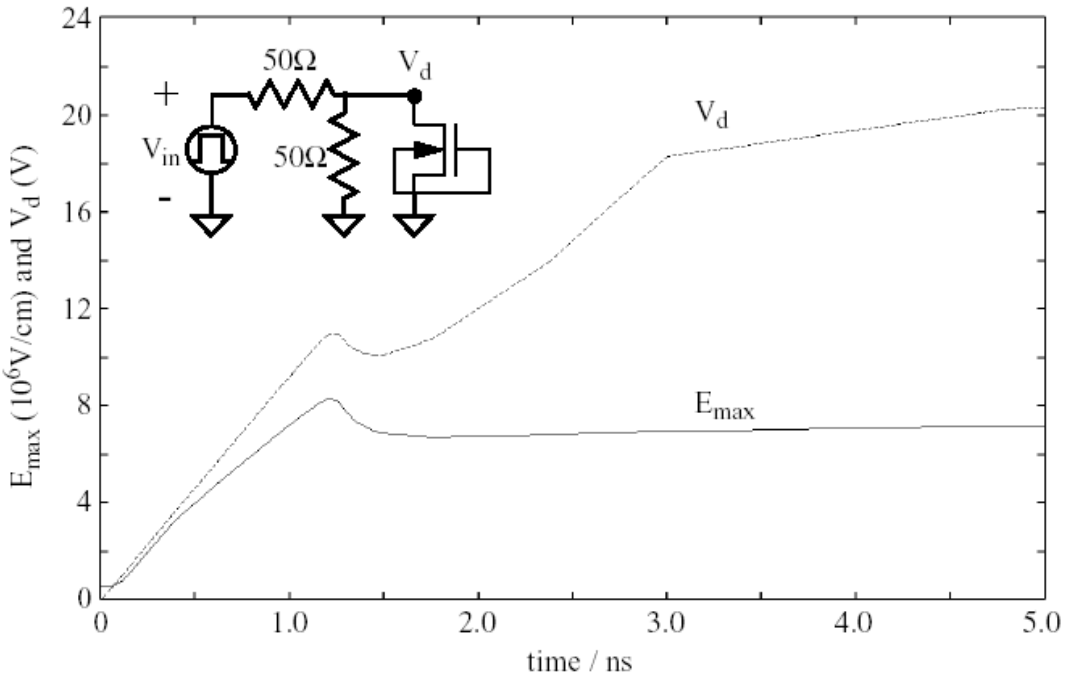


Figure. 6.5 Simulated response of max. electric field in gate oxide of an ESD-protection MOSFET

7. High Speed ESD Protection Scheme

It is well known that data transmission in high-speed operating chips suffer from waveform distortion by impedance mismatch in transmission line. The design of ESD protection for high-speed RF circuits poses a major design and reliability challenge due to the required transparency from the protection circuit in normal operating condition. The protections circuits consist of devices like MOSFET, diodes, silicon controlled rectifiers, which shunt the ESD current. However, under normal condition these devices present resistances and capacitances. At high frequencies the capacitances look like short circuits to ground. This may lead to impedance mismatch causing reflection of signals, corruption of signal integrity and inefficient power transfer between chips. In addition, while the operation frequency continues to rise, the size of protection devices and associated capacitances are not scaling down resulting in further worsening of power transfer behaviour. The obvious approach of minimizing capacitances is getting infeasible beyond a few GHz. This mandates new protection schemes such as inductor-based circuit and distributed transmission line protection scheme [43,44].

7.1. Inductor-Based and Distributed ESD Protection

As discussed most CMOS protection structures present parasitics, which is detrimental to the RF circuit performance. Fig. 7.1 shows the input section of a most common rf circuit the low noise amplifier (LNA) based on common source [46].

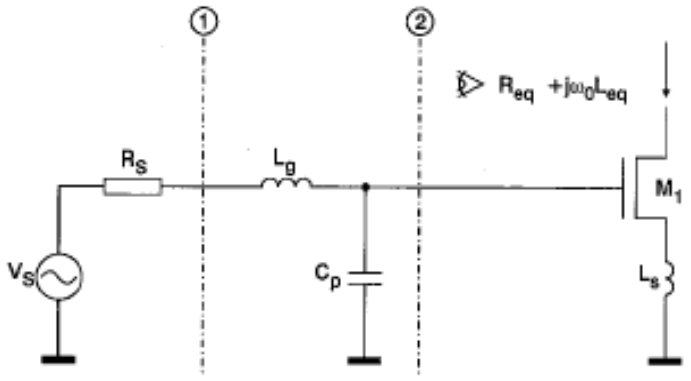


Figure. 7.1 Input section of an LNA

The parasitic capacitance C_p, which comes from bonding pad and ESD circuits, increases the noise figure of the LNA and decreases its gain. It places an upper bound on the input impedance

of the LNA, for higher frequency it becomes impossible to obtain an input match. The noise factor of LNA is given by [45]

$$F = 1 + (\frac{\omega_0}{\omega_\tau})^2 \frac{\gamma}{\alpha} g_m R_{eq} + \frac{\alpha\delta(1-|c|^2)}{k.g_m.R_{eq}} + (\frac{\omega_0}{\omega_\tau})^2 \frac{\gamma}{\alpha}\frac{2}{k} \quad 7.1$$

The noise factor has two terms dependent on g_mR_{eq} . The term, which is directly proportional to g_mR_{eq} results from the classical drain noise, whereas the one which is inversely proportional results from non-quasistatic noise. Due to ESD protection circuits C_p and in turn R_{eq} is large so the noise factor is mainly determined by the direct proportionality term. C_p also affects the power gain. High power gain requires high squared output current for a given source power. It is expressed as follows

$$|i_{out}|^2 = P_{av}(\frac{\omega_\tau}{\omega_0})^2 \frac{1}{R_{eq}} \quad 7.2$$

Again with increase in C_p, R_{eq} becomes large which reduces the output current and subsequently power gain is reduced.

Use of inductor as ESD protection can help alleviate the LNA power-gain and noise-factor reduction factor. Fig. 7.2 shows an LNA with on-chip inductor as an ESD protection [46].

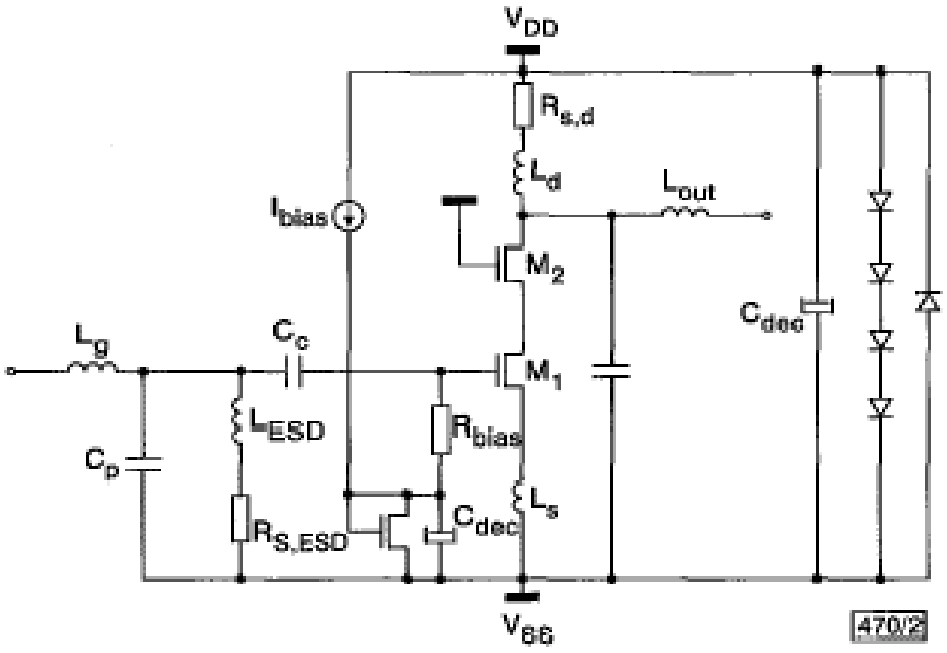


Figure. 7.2 An LNA with on-chip inductor as ESD protection circuit

The ESD currents have low frequency compared to rf signals so the protection network works as a low-pass filter. Contrary to the normal ESD protection circuits, which add parasitic capacitance to the input, this inductor tunes out parasitic capacitances at the input and doesn't show power gain reduction, in fact it improves the gain. Fig. 7.3 shows the comparison of LNA power gain, and noise figure with and without the protection inductor [46].

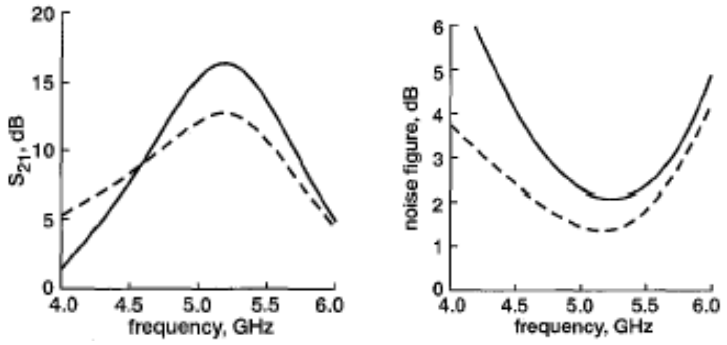


Figure. 7.3 LNA power gain and noise figure with and without inductor

Power consumption of both the circuits have been found same. The ESD protected LNA gives power gain of around 17 dB, which is 4dB higher than without protection LNA. This is according to expectation since inductor tunes out the parasitic capacitance. The ESD protection slightly deteriorates the noise figure.

Inductor based circuit provides a low impedance path at slow HBM ESD event, potentially providing very good protection levels. They show good compliance with 2kV HBM standard at multi GHz range of frequency of operation. However, CDM event spectrum spreads well into the RF range, near the resonant frequency of the L_{esd}-C_p resonator. This provides high impedance in the ESD conduction path and may also lead to oscillations since the resonator is undamped. A cancellation circuit as shown in Fig. 7.4 can solve this problem of CDM event [47].

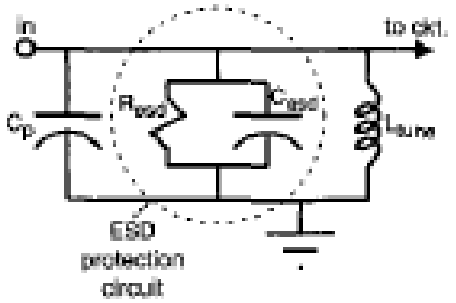


Figure. 7.4 Cancellation protection circuit for RF application

Here an explicit ESD protection circuit is added (modeled as R-C circuit in off state) to the circuit. Tuning out is now performed on total parasitic capacitance $C_p + C_{esd}$ by L_{tune}. During the normal operation ESD circuit remains off and the inductor is in resonance with the capacitances leaving only a large shunt resistor, which has minimal impact on the RF performance. When ESD event appears at the input, the protection circuit turns on to conduct the large stress current. Even in case of high frequency CDM stress the impedance in ESD path is low contrary to inductor-based protection and it effectively damps the high frequency oscillations. Fig. 7.5 presents the comparison of simulated response of inductor-based and cancellation circuit to a CDM event [47].

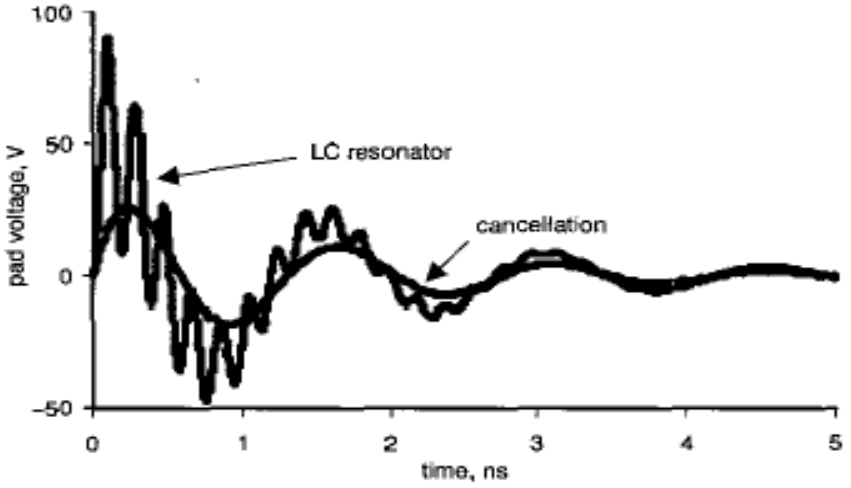


Figure. 7.5 Simulated response of L-C resonator and cancellation circuit for 500V CDM event

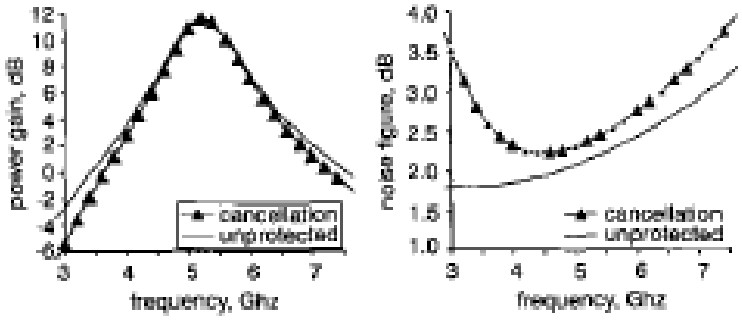


Fig. 7.6 Power gain and noise figure comparison for cancellation and unprotected circuit

Fig. 7.6 shows the power gain and noise figure comparison of cancellation and unprotected circuit [47]. Obviously the addition of ESD protection has minimal effect on the RF performance. For ESD protection in 10GHz range transmission line, like coplanar-waveguide (CPW), based protection circuit is employed.

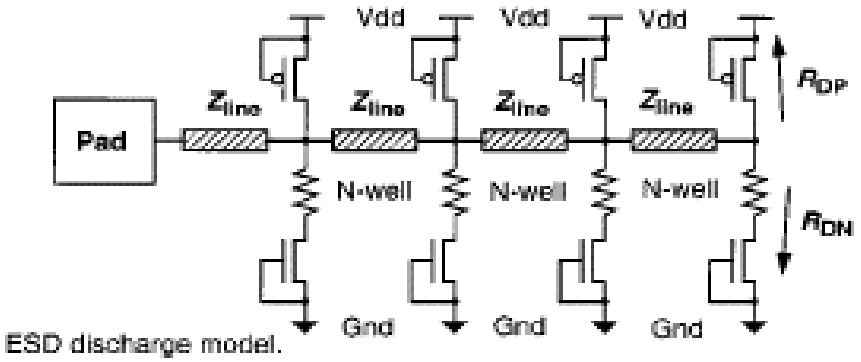


Figure. 7.7 Distributed ESD protection circuit

This provides accurate impedance matching depending on the number of CPW section used in protection circuit. It has been shown that use of more than four sections does not give significant improvement in the matching performance [48]. Fig. 7.7 shows the distributed ESD protection-circuit, here the required device has been divided in four small devices in parallel with a transmission line in between them [48].

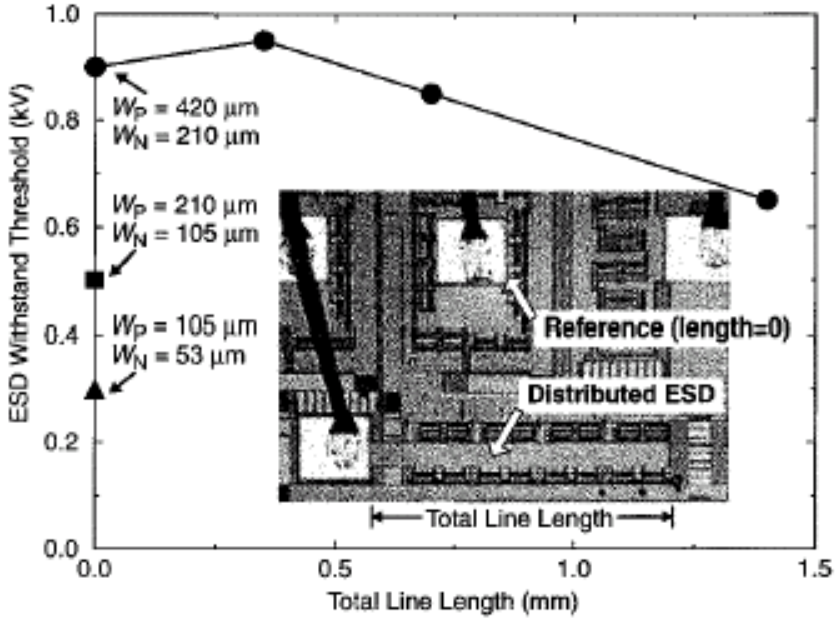


Figure. 7.8 Experimental CDM withstand threshold versus total line length

Fig. 7.8 shows the CDM withstand threshold for this distributed protection circuit [48]. Obviously with increase in interconnect length the threshold decreases but it provides better

impedance matching at the cost of reduced withstand threshold. In nutshell using appropriate circuits ESD protections can be provided for RF circuits till 10GHz frequency range without significantly affecting the normal RF performance of the circuit.

8. Conclusions

ESD is a destructive phenomenon that causes reliability problems and even permanent damage to the ICs. With the aggressive scaling of modern deep-submicron salicided CMOS technologies, the ICs are becoming more vulnerable to ESD. To prevent the ESD related failures, efficient on-chip protection structures are needed. However, a protection scheme that is suitable in the present technology may not be useful for the future generations of the technology. Therefore, a systematic approach is needed for ESD protection circuit design that can be transferred to the future technologies. To develop an ESD design methodology, a good understanding of the device physics under high current (electric field/temperature) conditions is essential. In this report, we presented a complete design methodology along with some new effects that are observed in 0.13 μm CMOS technology. We discussed the effects of bias conditions and layout parameters on the ESD robustness of an NMOS device for both silicided and non-silicided structures. We observed that an appropriate combination of bias conditions and layout parameters could maximize the ESD robustness of the device. This optimization can be achieved by developing simulation tools for ESD circuits. Circuit simulators such as SPICE have limitations in terms of operating voltages and currents. However, device simulators such as MEDICI can simulate the device behavior under ESD conditions. A combination of MEDICI and SPICE can provide a very useful simulation tool. MEDICI can simulate the device-under-test (DUT) and pass the nodal solution (voltages and currents) to SPICE, which can simulate the complete circuit along with the voltage/current sources and other circuit elements. MEDICI performs 2D simulation of a device, which may not be accurate enough for the future technologies (sub-100 nm). Hence, compact models that combine 3D and thermal effects are needed to achieve a fast and accurate simulation. Most of the device simulators available today do not simulate the inter-finger heat transfer and hence underestimate the ESD performance of a device. Therefore development of a simulator that incorporates inter-finger heat transfer will improve the accuracy of the simulation. The development of mixed-signal mixed-mode programming languages such as VHDL-AMS can play an important role in ESD protection circuit design since VHDL-AMS can simulate the interdependence of electric field, temperature and current using the model equations to provide a custom design environment.

9. References

[1] R. Merrill and E. Issaq, “ESD design methodology,” Proc. EOS/ESD Symp., pp. 233-237, 1993.

[2] D. L. Lin, “ESD sensitivity and VLSI technology trends: thermal breakdown and dielectric breakdown,” Proc. EOS/ESD Symp., pp. 73-81, 1993.

[3] S. H. Voldman and V. P. Grpss, “Scaling, optimization and design considerations of electrostatic discharge protection circuits in CMOS technology,” Proc. EOS/ESD Symp., pp. 251-260, 1993.

[4] C. Duvvury, R. McPhee, D. Baglee, and R. Rountree, “ESD protection reliability in 1 μm CMOS technologies,” Proc. IEEE Int. Reliability Physics Symp., pp. 199-205, 1986.

[5] F. Kuper, J. M. Luchies, and J. Bruines, “Suppression of soft ESD failures in a submicron CMOS process,” Proc. EOS/ESD Symp., pp. 117-122, 1993.

[6] T. L. Polgreen and A. Chatterjee, “Improving the ESD failure threshold of silicided n-MOS output transistors by ensuring uniform current flow, “ IEEE Trans. Elec. Dev., vol. 39, pp. 379-388, 1992.

[7] R. Rountree, “ESD protection for submicron CMOS circuits: issues and solutions,” IEDM Tech. Dig., pp. 580-583, 1988.

[8] http://public.itrs.net/Files/2001ITRS/Home.htm, The International Technology Roadmap for Semiconductors, 2001 edition.

[9] J. E. Vinson and J. J. Liou, “Electrostatic discharge in semiconductor devices: protection techniques,” Proc. IEEE, vol. 88, pp. 1878-1900, 2000.

[10] G. Krieger and P. Niles, “Diffused resistors characteristics at high current density levels: analysis and applications,” IEEE Trans. Elec. Dev., vol. 36, No. 2, pp. 416-423, 1989.

[11] G. Notermans, “On the use of N-well resistors for uniform triggering of ESD protection elements,” Proc. ESD/EOS Symp., pp. 221-229, 1997.

[12] A. Amerasekera, S. Ramaswamy, M-C Chang, and C. Duvvury, “Modeling MOS snapback and parasitic bipolar action for circuit-level ESD and high current simulations,” Proc. IRPS, pp. 318-326, 1996.

[13] H. Melchior and M. J. O. Strutt, “Secondary breakdown in transistors,” Proc. IEEE, pp. 439-440,1964.

[14] Y. Fong and C. Hu, “High-current snapback characteristics of MOSFETs,” IEEE Trans. Elec. Dev., vol. 37, pp. 2101-2103, 1990.

[15] K. Esmark, Device Simulation of ESD Protection Elements, Hartung-Gorre Verlag Konstanz, 2002.

[16] A. Amerasekera and C. Duvvury, ESD in Silicon Integrated Circuits, 2nd ed., Wiley, 2002.

[17] C. Russ, K. Bock, M. Rasras and I. D. Wolf, "Non-uniform triggering of gg-nMOST investigated by combined emission microscopy and transmission line pulsing," Proc. EOS/ESD Symp., pp. 177-186, 1998.

[18] T. J. Maloney and N. Khurana, "Transmission line pulsing techniques for circuit modeling of ESD phenomena," Proc. EOS/ESD Symp., pp. 49-54, 1985.

[19] H. Ishizuka, K. Okuyama and K. Kubota, "Photon emission study of ESD protection devices under second breakdown conditions," Proc. IRPS, pp. 286-291, 1994.

[20] N. Khurana, T. Maloney and W. Yeh, "ESD on CHMOS devices – equivalent circuits, physical models and failure mechanisms," Proc. IEEE Int. Reliability Physics Symp., pp. 212-223, 1985.

[21] C. Duvvury and C. Diaz, "Dynamic gate coupling of NMOS for efficient output ESD protection," Proc. IEEE Int. Reliability Physics Symp., pp. 141-150, 1992.

[22] C. Duvvury, C. Diaz and T. Haddock, "Achieving uniform nMOS device power distribution for sub-micron ESD reliability," IEDM Tech. Dig., pp. 131-134, 1992.

[23] M-D. Ker and T-Y. Chen, "Substrate-triggered ESD protection circuit without extra process modification," IEEE J. Solid-state Circuits, vol. 38, no. 2, pp. 295-302, Feb. 2003.

[24] K-H. Oh, C. Duvvury, C. Salling, K. Banerjee and R. W. Dutton, "Non-uniform bipolar conduction in single finger NMOS transistors and implications for deep submicron ESD design," Proc. IRPS, pp. 226-234, 2001.

[25] K-H. Oh, C. Duvvury, K. Banerjee and R. W. Dutton, "Analysis of non-uniform ESD current distribution in deep submicron NMOS transistors," IEEE Trans. Elec. Dev., vol. 49, no. 12, pp. 2171-2182, Dec. 2002.

[26] K-H. Oh, C. Duvvury, K. Banerjee and R. W. Dutton, "Gate bias induced heating effects and implications for the design of deep-submicron ESD protection," IEDM Tech. Dig., pp. 315-318, 2001.

[27] K-H. Oh, C. Duvvury, K. Banerjee and R. W. Dutton, "Analysis of gate-bias-induced heating effects in deep-submicron ESD protection designs," IEEE Trans. Dev. and Mat. Reliability, vol. 2, no. 2, pp. 36-42, Jun. 2002.

[28] T. Suzuki, S. Mitarai, S. Ito, H. Monma and N. Higashi, "A study of fully silicided 0.18 μm CMOS ESD protection devices," Proc. EOS/ESD Symp., pp. 78-87, 1999.

[29] K-H. Oh, C. Duvvury, K. Banerjee and R. W. Dutton, "Investigation of gate to contact spacing effect on ESD robustness of salicided deep submicron single finger NMOS transistors," Proc. IRPS, pp. 148-155, 2002.

[30] K-H. Oh, C. Duvvury, K. Banerjee and R. W. Dutton, "Impact of gate-to-contact spacing on ESD performance of salicided deep submicron NMOS transistors," IEEE Trans. Elec. Dev., vol. 49, no. 12, pp. 2183-2192, Dec. 2002.

[31] A. Amerasekera, V. Gupta, K. Vasanth and S. Ramaswamy, "Analysis of snapback behavior on the ESD capability of sub-0.20 μm NMOS," Proc. IRPS, pp. 159-166, 1999.

[32] K-H. Oh, K. Banerjee, C. Duvvury and R. W. Dutton, "Non-uniform Conduction Induced Reverse Channel Length Dependence of ESD Reliability for Silicided NMOS Transistors," IEDM Tech. Dig., pp. 341-344, Dec. 2002.

[33] V. M. Dwyer, A.J. Franklin, and D.S. Campbell, "Thermal Failure in Semiconductor Devices," Solid-State Electronics, vol. 33, 1990, pp. 553-560.

[34] "MEDICI two-dimensional Semiconductor Device Simulation, Version 1.1" Technology Modelign Associates, Inc. Palo Alto CA, 1993

[35] S. Selberher, Analysis and Simulation of Semiconductor Devices, Springer-Verlag, New York 1984.

[36] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, "Aphysically Based Mobility Model for Numerical Simulation of Nonplanar Devices," IEEE Trans. Computer Aided Design, vol. CAD-7 1998, pp. 1164-1171.

[37] Z. Yu, D. Chen, L. So, and R. W. Dutton, "PISCES-2ET Two Dimensional Device Simulation for Silicon and Heterostructures" Technical Report, Integrated Circuits Laboratory, Stanford University,1994.

[38] S. M. Sze, Physics of Semiconductor Devices, 2nd Ed., John Wiley, New York, 1981.

[39] J. G. Rollins and J. Choma, Jr., "Mixed-Mode PISCES-SPICE Coupled Circuit and Device Solver." IEEE Trans. Computer-Aided Design, vol. CAD-7, 1998,pp. 862-867.

[40] Z. Yu and R. W. Dutton, "A Modularized Mixed IC Device/Circuit Simulation System," Proc. Synthesis and Simulation Meeting and International Interchange, 1992, pp. 444-448.

[41] S. Ohtani, M. Yoshida, N. Kitagawa, and T. Saitoh, "Model of leakage current in LDD output MOSFET due to low-level ESD stress," Proc. 12th EOS/ESD Symp., 1990, pp. 177-181.

[42] S. M. Sze, VLSI Technology, 2nd Ed. McGraw-Hill, New York, 1988,p. 118.

[43] B. Kleveland, and T. lee, US Patent # 5929969, Oct. 1999.

[44] B. Kleveland, et al. "Distributed ESD protection for High Speed Integrated Circuits," IEEE Electron Device Lett., vol. 21, no. 8, pp. 390-392, 2000.

[45] Janssens J., and Steyaert M., “MOS noise performance under impedance matching constraints” Electron. Lett., 1999, 35, (15), pp. 1278-1280.

[46] P. Leroux and M. Steyaert, “High-performance 5.2 GHz LNA with on-chip inductor to provide ESD protection” IEEE Electronics Letters, 29 Mar., Vol.37 No.7 2001.

[47] S. Hyvonen, S. Joshi, and E. Rosenbaum, “ Cancellation technique to provide ESD protection for multi GHz RF inputs” IEEE Electronics Letters, 6 Feb., Vol. 39, No. 3, 2003.

[48] C. Ito, K. Banerjee, and R. Dutton, “ Analysis and Design of ESD Protection Circuits for High-Frequency/RF Applications” IEEE Letters 2001.