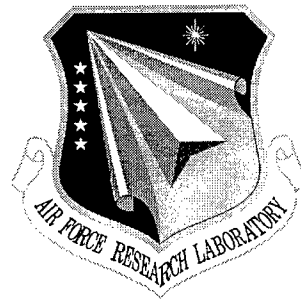


AFRL-IF-RS-TR-1998-116  
Interim Technical Report  
June 1998



## RELIABILITY-DRIVEN CAD SYSTEM FOR DEEP-SUBMICRON VLSI CIRCUITS

University of Illinois

S.M. Kang, E. Rosenbaum, Y.K. Cheng, L.P. Yuan, T. Li, C.H. Tsai and E. Li

*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

**DTIC QUALITY INSPECTED 1**

**AIR FORCE RESEARCH LABORATORY  
INFORMATION DIRECTORATE  
ROME RESEARCH SITE  
ROME, NEW YORK**

19980911 010

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

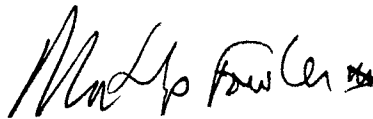
AFRL-IF-RS-TR-1998-116 has been reviewed and is approved for publication.

APPROVED:



MARTIN J. WALTER  
Project Engineer

FOR THE DIRECTOR:



NORTHROP FOWLER, III, Technical Advisor  
Information Technology Division  
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFTC, 26 Electronic Pky, Rome, NY 13441-4514. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

**REPORT DOCUMENTATION PAGE**

*Form Approved  
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 1998	3. REPORT TYPE AND DATES COVERED Interim Jul 96 - Jun 97	
4. TITLE AND SUBTITLE RELIABILITY-DRIVEN CAD SYSTEM FOR DEEP-SUBMICRON VLSI CIRCUITS			5. FUNDING NUMBERS G - F30602-94-1-0006 PE - 62702F PR - 4600 TA - A0 WU - A3	
6. AUTHOR(S) S.M. Kang, E. Rosenbaum, Y.K. Cheng, L.P. Yuan, T. Li, C.H. Tsai, and E. Li			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Illinois 109 Coble Hall, 801 S. Wright St. Champaign IL 61820			10. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-IF-RS-TR-1998-116	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFRL/IFTC 26 Electronic Pky Rome NY 13441-4514			11. SUPPLEMENTARY NOTES AFRL Project Engineer: Martin J. Walter/IFTC/(315) 330-4102	
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report describes the development of a hierarchical reliability-driven CAD system for deep-submicron VLSI/ULSI circuits. Three general issues are addressed in this report: layout extraction, circuit simulation, and experiment. Conventional layout extractors are not sufficient for use with the reliability-driven CAD tool. Device and interconnect coordinates must be recorded for simulation of the chip temperature profile; parasitic devices must be extracted for simulation of ESD events. The chip temperature profile, critical for accurate prediction of electromigration-induced failures and for accurate simulation of circuit delays, can be simulated for the case that the user supplies the input vectors. If the input vectors are unknown, the temperature profile is estimated using the average power consumption as derived through statistical analysis. I/O projection circuits were subjected to ESD stress and the results were used to verify the electrothermal circuit simulation iETSIM. Sub - 0.5 um PMOSFETS were subjected to hot carrier stress and it was observed that the dominant degradation mechanism is different from that seen in longer channel devices. The CAD system has been installed on the World Wide Web to provide greater ease of use and platform independence.				
14. SUBJECT TERMS Reliability-driven CAD, EOS/ESD, electromigration reliability, CMOS integrated circuits			15. NUMBER OF PAGES 96	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED			16. PRICE CODE	
18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED		19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED		20. LIMITATION OF ABSTRACT UL

# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
<b>2</b>	<b>CHIP TEMPERATURE PROFILE SIMULATION</b>	<b>4</b>
2.1	Introduction . . . . .	4
2.1.1	Overview of ILLIADS-T . . . . .	4
2.1.2	Organization of this chapter . . . . .	7
2.2	Temperature-Dependent MOS Device Modeling . . . . .	7
2.3	Temperature-Dependent RWQ Modeling . . . . .	8
2.3.1	Mobility and RWQ fitting results . . . . .	10
2.4	3-D Thermal Simulator - iTEMP . . . . .	13
2.4.1	Modeling . . . . .	13
2.4.2	Problem formulation . . . . .	14
2.5	Verification of ILLIADS-T and Simulation Results . . . . .	19
2.5.1	Tester chip design and calibration . . . . .	19
2.5.2	Verification of ILLIADS-T . . . . .	22
2.5.3	ILLIADS-T simulation example . . . . .	29
2.6	Timing Analysis Using ILLIADS-T . . . . .	32
2.6.1	Introduction . . . . .	32
2.6.2	Temperature-dependent gate and RC delays . . . . .	33
2.6.3	Monte-Carlo power estimation for on-chip temperature profiling . . . . .	33
2.6.4	Thermal simulation for timing analysis . . . . .	35
2.7	Simulation Results . . . . .	36

<b>3</b>	<b>STATISTICAL POWER ANALYSIS OF SEQUENTIAL CIRCUITS</b>	<b>39</b>
<b>4</b>	<b>ESD PROTECTION CIRCUIT DESIGN AND ANALYSIS USING iETSIM</b>	<b>46</b>
4.1	Electrothermal Models in iETSIM . . . . .	46
4.1.1	MOS snapback model . . . . .	46
4.1.2	Thermal model . . . . .	47
4.1.3	Diffused resistor model . . . . .	49
4.1.4	SCR model . . . . .	49
4.2	Simulation and Optimization of Multi-Finger Protection NMOSFETS . . . . .	51
4.2.1	Introduction . . . . .	51
4.2.2	Simulation and experimental results . . . . .	52
4.2.3	Discussion . . . . .	54
4.2.4	Current profile . . . . .	55
4.2.5	Layout optimization methodology . . . . .	56
4.2.6	Design example . . . . .	57
4.3	Simulation of a Two-Stage Protection Circuit . . . . .	58
4.3.1	Circuit description . . . . .	58
4.3.2	Device characterization . . . . .	60
4.3.3	Simulation and experiment . . . . .	65
4.3.4	Failure analysis . . . . .	68
<b>5</b>	<b>HOT CARRIER INDUCED DEGRADATION OF DEEP-SUBMICRON PMOSFETS</b>	<b>70</b>

5.1	Introduction . . . . .	70
5.2	Experiment . . . . .	70
5.3	Results . . . . .	73
5.4	Conclusions . . . . .	76

# Reliability-Driven CAD System for Deep-Submicron VLSI Circuits

## Abstract

This report describes the development of a hierarchical reliability-driven CAD system for deep-submicron VLSI/ULSI circuits. Three general issues are addressed in this report: layout extraction, circuit simulation, and experiment.

Conventional layout extractors are not sufficient for use with the reliability-driven CAD tool. Device and interconnect coordinates must be recorded for simulation of the chip temperature profile; parasitic devices must be extracted for simulation of ESD events.

The chip temperature profile, critical for accurate prediction of electromigration-induced failures and for accurate simulation of circuit delays, can be simulated for the case that the user supplies the input vectors. If the input vectors are unknown, the temperature profile is estimated using the average power consumption as derived through statistical analysis.

I/O protection circuits were subjected to ESD stress and the results were used to verify the electrothermal circuit simulation iETSIM. Sub - 0.5  $\mu\text{m}$  PMOSFETs were subjected to hot carrier stress and it was observed that the dominant degradation mechanism is different from that seen in longer channel devices.

The CAD system has been installed on the World Wide Web to provide greater ease of use and platform independence.

**Subject Terms:** Reliability-driven CAD, EOS/ESD, electromigration reliability, CMOS integrated circuits.

# 1 INTRODUCTION

Using circuit reliability prediction and design tools, one can identify and remedy potential reliability problems before circuit fabrication commences. This is a cost- and time-saving alternative to first identifying problems after manufacture, during testing and/or burn-in. This work addresses reliability problems that threaten integrated circuits, such as electrostatic discharge/electrical overstress (ESD/EOS), time-dependent dielectric breakdown, hot-carrier-induced degradation, and electromigration.

Useful reliability design tools should have the three features listed below:

- Major reliability problems and performance issues are considered concurrently.
- Entire chips can be analyzed quickly.
- Computer models for reliability problems are physics-based and experimentally verified.

The reliability-driven CAD system is being developed based on these three principles. Figure 1 shows the hierarchy of the reliability-driven CAD system.

During this year, the third, it was proposed in the original statement of work that the following tasks be completed.

- Develop a layout extractor for I/O and ESD protection circuits.
- Add ESD reliability design rules to the rule checker.
- Formulate a procedure for calculating substrate current in a circuit simulator which sidesteps the time step quantization problem.
- Integrate all of the reliability software into one user-friendly package.

Significant progress has been made on item #1. This task will be completed during the 9-month extension of this project. Items #2 and #3 have been replaced with new tasks which arose after the initiation of this project, namely, development of an I/O circuit

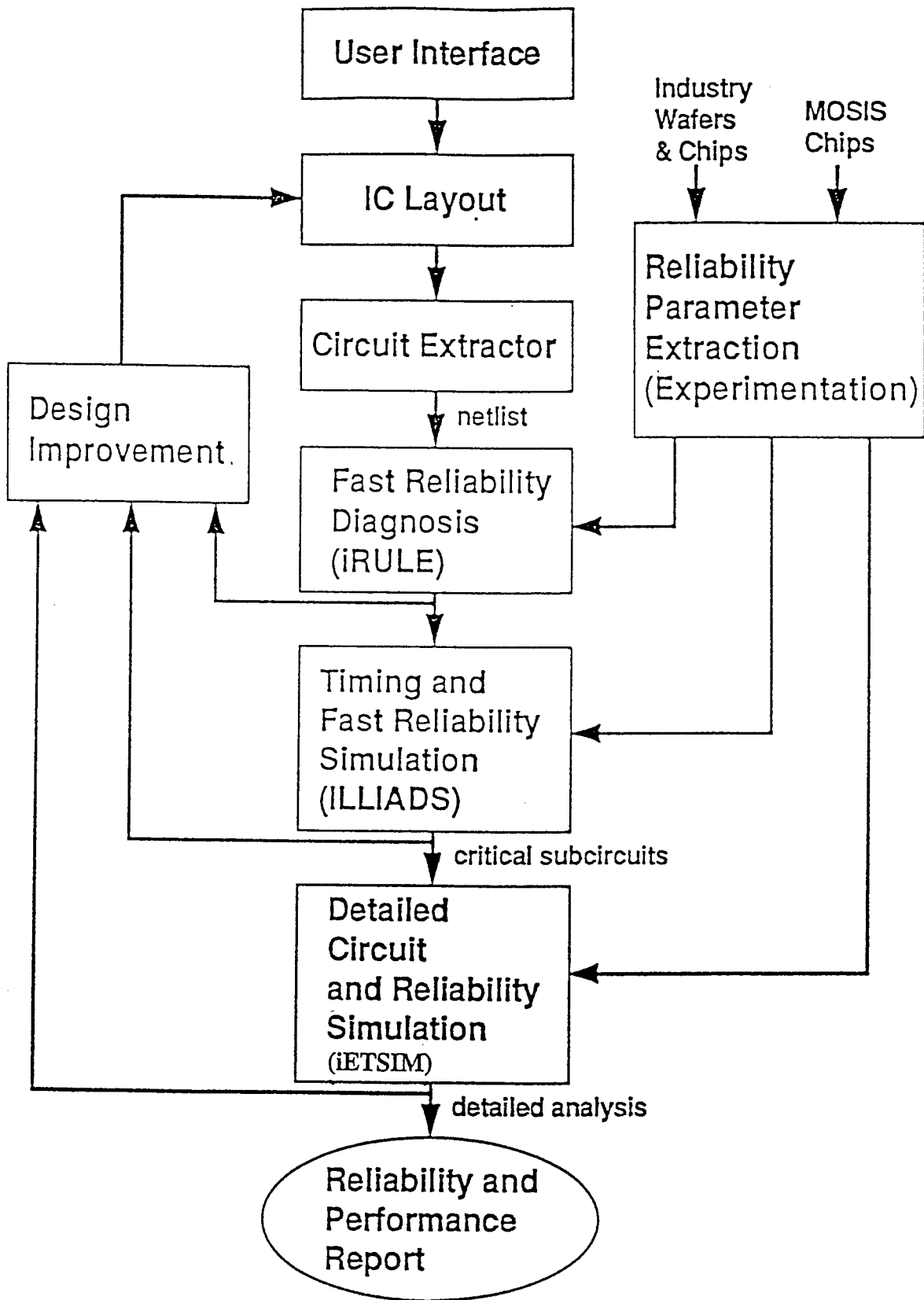


Figure 1: Block diagram of the hierarchy of the reliability-driven CAD system.

synthesis tool and investigation of hot carrier reliability in 0.25  $\mu\text{m}$  CMOS technology. Task #4 was addressed by placing the reliability tools on the World Wide Web.

We included temperature estimation in our tool for electromigration diagnosis of integrated circuits. This task was not envisioned in the original proposal but is described in this yearly report.

## 2 CHIP TEMPERATURE PROFILE SIMULATION

### 2.1 Introduction

Due to the increasing component density, higher operating speed, and larger scale of integration, the power density and on-chip temperature in ICs continue to increase. Furthermore, the temperature of packaged VLSI circuits can vary by as much as a few tens of degrees from the center to an edge of the chip. Because the failure rate of microelectronic devices depends heavily on the localized operating temperature, hot spots due to high local-power dissipation have become a long-term IC reliability concern.

Furthermore, the circuit delay and the interconnect reliability are functions of the operating temperature.

This chapter describes a tool, ILLIADS-T, which is used for temperature-profile estimation, hot-spot identification, and circuit reliability prediction for CMOS VLSI chips.

#### 2.1.1 Overview of ILLIADS-T

Simply put, electrothermal simulation consists of electrical and thermal simulations. The electrical simulation is used to obtain the information on power dissipation and performance of devices or circuits. On the other hand, the thermal simulation is used to find the temperature profile and to update all the temperature-dependent physical parameters of the device or circuit model. A simplified flowchart of ILLIADS-T simulation procedure is shown in Fig. 2.2. The main features of ILLIADS-T are listed below.

1. To achieve the computational time efficiency required for large circuits, ILLIADS-T uses a fast-timing simulator, ILLIADS, to calculate the power dissipated by each *logic gate*. Each gate is then viewed as a heat source in our thermal simulation.
2. Existing *coupled* electrothermal simulators are inefficient in nature [1][2][3]. The total simulation time is first divided into many small time intervals, then the power and temperature values are updated and coupled for each time interval. This approach

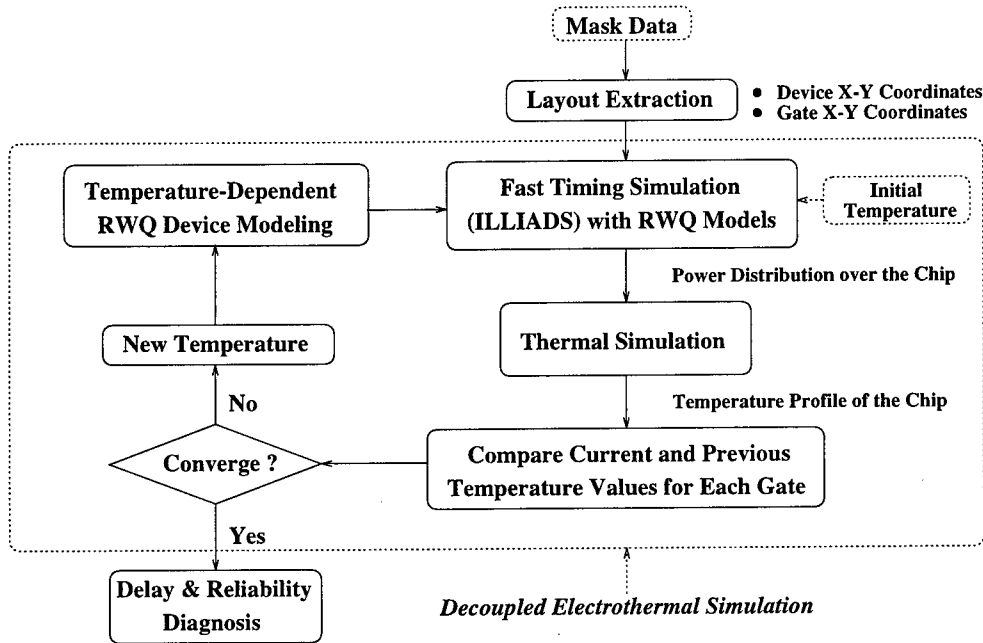


Figure 2.2: Flowchart of ILLIADS-T electrothermal simulation.

is only suitable for the transient simulation on small circuits. ILLIADS-T, which is designed to find the chip-level steady-state temperature distribution and the resulting circuit performance, uses a much more efficient approach. It starts with an initial guess of the average chip temperature and then calculates the average power for each gate based on the current waveform drawn from the power supply. Next, the gate power values are fed to the thermal simulator to estimate the temperature profile. The temperature profile is then used to update the device model parameters for the second round of power calculation. This process continues until convergence is obtained and the steady-state temperature profile has been found. Note that our approach *decouples* the power and temperature calculations. This approach is justified because the chip temperature does not follow the instantaneous power dissipation; instead, it becomes virtually constant after reaching the steady state. Therefore, the average power rather than the instantaneous power is used in our temperature calculation.

3. Because existing electrothermal simulators were developed mainly for the temperature profile estimation of SSI or MSI circuits [1][2][3], thermal boundary conditions (BCs) were simplified. Moreover, the 1-D/2-D thermal simulations were usually adopted. For VLSI/ULSI chips with complex packaging structures, the simplified BCs and 1-D/2-D approaches are not usually valid. To handle this problem, we have developed a new thermal simulator, iTEMP, to solve 3-D heat equations for the chip substrate, while modeling the packages and heat sinks as *effective* thermal resistances.
4. By using the region-wise quadratic (RWQ) modeling technique instead of the complex MOS models as in [4], temperature-dependent power and delay estimation can be done in ILLIADS-T even when only measured data are available and the MOS models have not been fully developed or characterized. This makes ILLIADS-T device-model-independent and thus, applicable to advanced CMOS technologies.

Referring back to Fig. 2.2, the primary input to ILLIADS-T is the layout description file of the target VLSI chip. A layout extractor has been developed to obtain the electrical circuit that the layout represents, as well as to identify the location of each device. A standard device specification in the netlist generated by our layout extractor is shown as follows:

```
MOS_name ND NG NS NB MODEL_name <L=VAL> <W=VAL> <AD=VAL>
<AS=VAL> <PD=VAL> <PS=VAL> XMIN YMIN XMAX YMAX,
```

where XMIN, YMIN, XMAX, and YMAX define the bounding box of a MOS device, and MODEL\_name specifies a particular RWQ model for a MOS device.

ILLIADS-T then calculates the bounds of each logic gate according to the coordinates of the bounding boxes of MOS devices within this gate. Next, the average power dissipation from each gate at the initial temperature is calculated by ILLIADS. Then iTEMP will take as input the power values and the coordinates of heat sources to calculate the on-chip

temperature profile by solving the heat equations, in particular, the average temperature of each gate is found. At this stage, the transistors in each gate have the updated local temperature and ILLIADS is rerun to find the new average power values under the new temperature distribution. This iterative procedure stops when the updated temperature of each gate no longer has any significant change from the previous value. Empirical results shown in Section 2.5 indicate that this process is efficient and usually converges after two or three iterations. Note that in CMOS circuits, the short-circuit power usually accounts for approximately 25% of the total IC power consumption [5]. The temperature-induced variations of the short-circuit power and/or the switching activity are what necessitate a few iterations during ILLIADS-T simulation.

### **2.1.2 Organization of this chapter**

The remainder of this chapter is organized as follows. In Section 2.2, we present the temperature-dependent MOS device models for fast-timing simulation. A 3-D thermal simulator is described in Section 2.4. In order to verify the simulation results, we have designed a tester chip and had it fabricated and packaged by MOSIS. The tester chip design and the comparison between measurement and ILLIADS-T simulation results are presented in Section 2.5. Additional ILLIADS-T simulation examples are also given in Section 2.5. Application of ILLIADS-T to timing analysis is presented in Section 2.7. ILLIADS-T is also used in the temperature-dependent electromigration diagnosis tool, iTEM; this work was presented in our 1996 Report to Rome Laboratory.

## **2.2 Temperature-Dependent MOS Device Modeling**

Most timing simulators lack simulation accuracy because of the use of simplified MOS transistor models for submicron devices. To solve this problem, the regionwise-quadratic (RWQ) modeling technique [6] was introduced for submicron device modeling. In order to accurately capture the temperature effect on power and gate delay, we have developed a *temperature-dependent* RWQ model for MOS devices.

The RWQ modeling procedure takes a set of data points  $(V_{ds}, V_{gs}, I_{ds})$  as input, which have been obtained either from measured data of a device or by exercising a particular analytical or empirical MOS I-V model. Defining  $V_{gse} = V_{gs} - V_{T0}$ , the  $(V_{ds}, V_{gse})$  plane is optimally partitioned into a number of regions, and a quadratic model of  $I_{ds}$  is fit in terms of  $V_{ds}$  and  $V_{gse}$  in each region using the data points in that region. Specifically, the following quadratic model of  $I_{ds}$  is fit to the data in the  $k^{th}$  region,

$$\frac{I_{ds}}{\beta} = (\alpha_0^{(k)} + \alpha_1^{(k)}V_{gse} + \alpha_2^{(k)}V_{ds} + \alpha_3^{(k)}V_{gse}^2 + \alpha_4^{(k)}V_{gse}V_{ds} + \alpha_5^{(k)}V_{ds}^2), \quad k = 1, \dots, n_r \quad (2.1)$$

where  $\beta = \frac{1}{2}\mu_0 C_{ox} \frac{W}{L}$ ,  $n_r$  is the number of regions chosen for best fitting, and  $\alpha$ 's are the fitting parameters for the  $k^{th}$  region. In (2.1), there are two temperature-dependent physical parameters,  $\mu_0$  and  $V_{T0}$ .

### 2.3 Temperature-Dependent RWQ Modeling

The temperature dependency of  $V_{T0}$  is relatively weak compared to  $\mu_0$ , and we adopt the temperature model of  $V_{T0}$  which is used in the SPICE Level 1 MOS model. In contrast, it is difficult to devise an analytical formula to accurately calculate the channel mobility because of the complex physical effects involved [7]. Therefore, we have developed a physically based, semi-empirical mobility model which has been verified for the temperature range of 300 - 400 K. Since this model is to be used in a fast timing simulator, it must be simple yet accurate. In addition, this model should be scaled only with temperature but not with the transverse electric field  $E_{eff}$  (although the physical channel mobility actually depends on both temperature and transverse electric field). This is because the transverse field dependency is automatically taken into account by  $V_{gse}$  in (2.1), i.e.,  $E_{eff} \propto V_{gse}$  [8].

The carrier mobility is directly related to the mean free time between collisions which, in turn, is determined by the various scattering mechanisms. In silicon devices, the three most important scattering mechanisms are Coulomb, lattice, and surface-roughness scattering. According to the scattering theory [9], one may show that

$$\frac{1}{\tau_c} = \frac{1}{\tau_{c,C}} + \frac{1}{\tau_{c,L}} + \frac{1}{\tau_{c,SR}}$$

or equivalently

$$\frac{1}{\mu_n} = \frac{1}{\mu_C} + \frac{1}{\mu_L} + \frac{1}{\mu_{SR}} \quad (2.2)$$

We will next examine the temperature dependency of each of these terms. The  $E_{eff}$  dependency will be also noted.

### A. Coulomb (Impurity) Scattering

Coulomb scattering results when a charged carrier interacts with an ionized impurity. The Coulomb scattering limited mobility  $\mu_C$  can be described as [10]

$$\mu_C \propto T/N_I \quad (2.3)$$

where  $T$  is the temperature and  $N_I$  is the charge density near the Si-SiO<sub>2</sub> interface.

### B. Lattice (Phonon) Scattering

Lattice scattering results from thermal vibrations of the lattice atoms at any temperature above zero. For intermediate inversion layer concentrations ( $Q_N/q = 0.5 \sim 5 \times 10^{12}/cm^2$ ), the channel mobility has been observed to have the following dependencies on  $T$  and  $E_{eff}$  [11]

$$\mu_L \propto T^{-n} E_{eff}^{-1/\gamma} \quad (2.4)$$

where  $\gamma$  and  $n$  are process dependent parameters.

### C. Surface-Roughness Scattering

Surface-roughness scattering results from interactions with the asperities at the Si-SiO<sub>2</sub> interface. It is not a function of temperature [12]. The surface-roughness scattering limited mobility  $\mu_{SR}$  depends on  $E_{eff}$  according to [12]

$$\mu_{SR} \propto E_{eff}^{-2} \quad (2.5)$$

Based on (2.3), (2.4), (2.5) and (2.2), we propose the following temperature-dependent mobility model:

$$U(T) = \mu_0^{-1}(T)$$

$$\begin{aligned}
&= a_1 T^{-1} + a_2 T^{a_3} + a_4 \\
&= A_1 \left[ \left( \frac{T}{300} \right)^{-1} - 1 \right] + A_2 \left[ \left( \frac{T}{300} \right)^{A_3} - 1 \right] + A_4
\end{aligned} \tag{2.6}$$

The symbol  $U(T)$ , defined as the inverse of mobility  $\mu_0(T)$ , is used for convenience.  $A_1$ ,  $A_2$ ,  $A_3$  and  $A_4$  are fitting parameters and will be determined by using the nonlinear least-square fitting technique to match the extracted  $\mu_0(T)$ . The  $E_{eff}$  dependencies in (2.4) and (2.5) are merged into  $A_2$  and  $A_4$  in (2.6).

Let  $I_{ds}^{(k)}(\mathbf{x})$  and  $\mu_0(T)f^{(k)}(\mathbf{x})$  be the experimental and RWQ-fitted drain currents in the  $k$ -th region, respectively.  $\mathbf{x}$  stands for the data point vector  $(V_{ds}, V_{gse})$ . We extract the optimized mobility by minimizing the objective function

$$\sum_{k=1}^{n_r} \sum_{i=1}^{N_k} (I_{ds}^{(k)}(\mathbf{x}_i) - \mu_0(T)f^{(k)}(\mathbf{x}_i))^2 \tag{2.7}$$

where  $n_r$  is the number of regions and  $N_k$  is the number of data points in region  $k$ . This provides us with a best fit  $\mu_0(T)f^{(k)}$  from which  $\mu_0(T)$  is extracted as follows:

$$\mu_0(T) = \frac{\sum_{k=1}^{n_r} \sum_{i=1}^{N_k} I_{ds}^{(k)}(\mathbf{x}_i) f^{(k)}(\mathbf{x}_i)}{\sum_{k=1}^{n_r} \sum_{i=1}^{N_k} (f^{(k)}(\mathbf{x}_i))^2} \tag{2.8}$$

Once  $\mu_0(T)$  values are extracted at several temperatures, (2.6) is used to find the optimized  $A_1$ ,  $A_2$ ,  $A_3$  and  $A_4$ .

### 2.3.1 Mobility and RWQ fitting results

To evaluate the mobility modeling procedure described by (2.6) - (2.8), we extracted  $\mu_0(T)$  for an NMOSFET with dimensions  $L = 1 \mu\text{m}$  and  $W = 12.5 \mu\text{m}$ . Nonlinear least-square fitting of the  $A_1$ ,  $A_2$ ,  $A_3$ , and  $A_4$  parameters was accomplished using the Levenberg-Marquart algorithm [13]. The results are shown in Fig. 2.3 and the extracted parameters are given in the inset. The RWQ model at  $30^\circ\text{C}$  is compared with measured data in Fig. 2.4(a). The mobility model in (2.6) was used to predict  $\mu_0(T = 90)$  as  $482.3 \text{ cm}^2/(\text{V} \cdot \text{s})$ , which is very close to  $483.5 \text{ cm}^2/(\text{V} \cdot \text{s})$  obtained by extraction. The  $I_{ds}$ - $V_{ds}$  curves at  $T = 90^\circ\text{C}$  were constructed by using the RWQ fitting parameters  $\alpha_0 - \alpha_5$  obtained at room temperature and

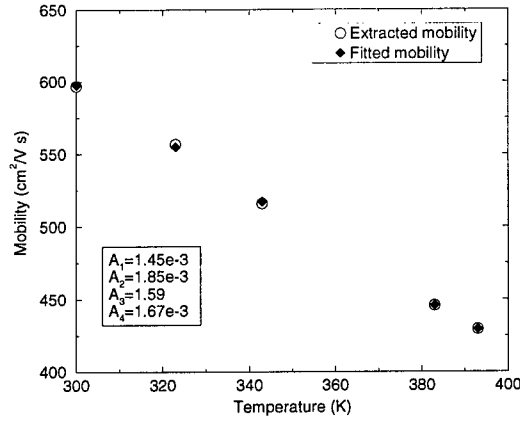


Figure 2.3: Fitted  $\mu_0(T)$  vs. extracted  $\mu_0(T)$ .

the value of  $\mu_0(T = 90)$ . The results are compared with experimental data in Fig. 2.4(b). The predicted  $I_{ds}-V_{ds}$  plot at 130 °C is shown in Fig. 2.5.

Finally, to demonstrate the simulation accuracy of ILLIADS-T in which a temperature-dependent RWQ model was implemented, we simulated a 9-stage inverter chain using both ILLIADS-T and SPICE, i.e., the BSIM3 MOSFET model was used in SPICE. The output waveforms at different temperatures are compared in Fig. 2.6.

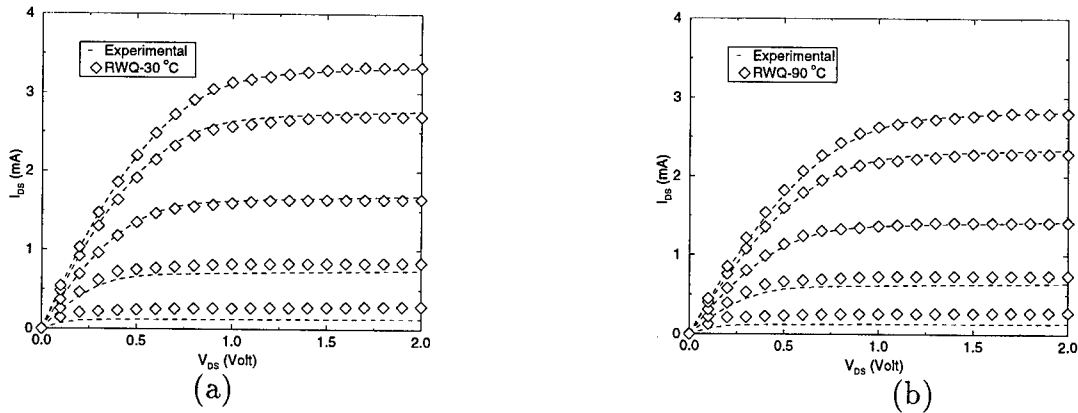


Figure 2.4: (a) RWQ fitting result at 30 °C, and (b) RWQ fitting result at 90 °C with mobility optimization.

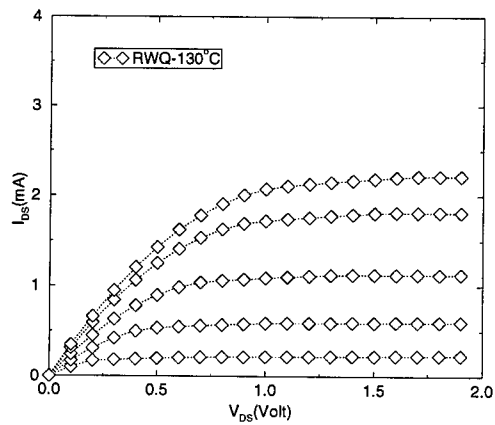


Figure 2.5: Predicted  $I_{ds}$ - $V_{ds}$  at 130 °C using RWQ model.

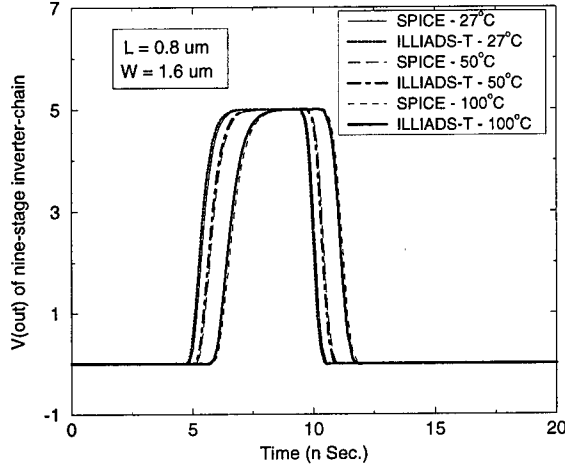


Figure 2.6: Waveform comparison between SPICE and ILLIADS-T for a 9-stage inverter chain.

## 2.4 3-D Thermal Simulator - iTEMP

### 2.4.1 Modeling

The heat diffusion equation is the governing equation for heat conduction and temperature calculation. The general equation is written as [14]

$$\rho c_p \frac{\partial T(x, y, z, t)}{\partial t} = \nabla \cdot [k(x, y, z, T) \nabla T(x, y, z, t)] + g(x, y, z, t) \quad (2.9)$$

subject to the general thermal BC:

$$k(x, y, z, T) \frac{\partial T(x, y, z, t)}{\partial n_i} + h_i T(x, y, z, t) = f_i(x, y, z) \quad (2.10)$$

where  $T$  is the temperature ( $^{\circ}\text{C}$ ),  $g$  is the power density of the heat source(s) ( $\text{W}/\text{m}^3$ ),  $k$  is the thermal conductivity ( $\text{W}/(\text{m}^{\circ}\text{C})$ ),  $\rho$  is the density of material ( $\text{Kg}/\text{m}^3$ ),  $c_p$  is the specific heat ( $\text{J}/(\text{Kg}^{\circ}\text{C})$ ),  $h_i$  is the heat transfer coefficient ( $\text{W}/(\text{m}^2^{\circ}\text{C})$ ),  $f_i(x, y, z)$  is a function of position, and  $n_i$  is the outward direction normal to the surface  $i$ . For a steady-state case, the  $\frac{\partial T}{\partial t}$  term is zero. Three kinds of thermal BC's derived from (2.10) can be applied to the chip boundaries, depending on the packaging materials and the surrounding environment.

They are isothermal (Dirichlet), insulated (Neumann) and convective (Robin) BC's [14]. In summary,  $T = f_i(x, y, z)$  for isothermal condition,  $\frac{\partial T}{\partial n_i} = 0$  for insulated condition, and  $k_i \frac{\partial T}{\partial n_i} = h_i(T - T_a)$  for convective condition, where  $T_a$  is the ambient temperature.

Traditional 1-D/2-D thermal simulators as we mentioned earlier, cannot provide the required accuracy for VLSI chips due to the finite 3-D structures. iTEMP, however, simulates a chip by using the following mixed 3-D/1-D strategy: (i) 3-D simulation is performed for the chip substrate for accuracy; (ii) To enhance the simulation efficiency, packages and heat sinks are treated as *1-D thermal resistances* by which the computational complexity is reduced substantially. We will henceforth call strategy (ii) the effective heat transfer macromodeling. The idea is to serially combine the thermal resistance of the *bulk of packages or heat sinks* ( $R_k$ ) with the one *from packages to ambience* ( $R_h$ ), and to find the effective heat transfer coefficient  $h^e$  as given by

$$h^e = \frac{1}{A_c} \frac{1}{(R_h + R_k)} \quad (2.11)$$

where  $R_k = \frac{L}{k_p A_c}$ ,  $R_h = \frac{1}{h_p A_c}$ ,  $L$  is the thickness and  $k_p$  is the thermal conductivity of the package or heat sink,  $h_p$  is the heat transfer coefficient from the package or heat sink to ambience, and  $A_c$  is the chip area normal to the direction of heat flow. In other words, we merge the package and heat sink effects into the  $h$  term in (2.10) and form an effective  $h^e$ . The advantage of effective heat transfer macromodeling is threefold. First, it enhances the efficiency of iTEMP. Second, it removes the difficulty of analytically solving heat conduction problems of multilayered chip structures [15]. Third, it allows iTEMP to easily handle complicated chip structures, such as pins, by replacing  $k_p$  in  $R_k$  with  $k_{eff} = X k_{pin} + (1-X)k_p$ , where  $k_{pin}$  is the thermal conductivity of pins, and  $X = \frac{(Area\ of\ pins)}{(Total\ package\ area)}$ .

#### 2.4.2 Problem formulation

We solve (2.9) and (2.10) by using the numerical 3-D finite-difference (FD) technique. Since the gate count in a VLSI chip is large, it is impractical to allocate one or more grids to each gate in the FD method. Instead, the grid number and spacings in iTEMP are determined by taking into account the chip size, the gate density, and the temperature field density. We employ an *adaptive* meshing technique to determine the grid spacing in iTEMP. First, iTEMP uniformly deploys the on-chip grids according the user-specified initial grid number. After obtaining an initial estimate of the temperature distribution, iTEMP

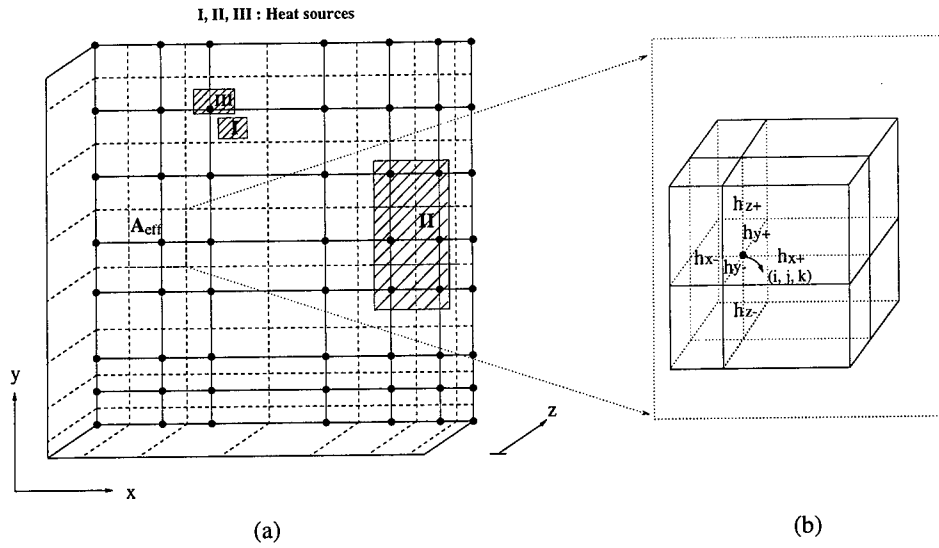


Figure 2.7: (a) Top view of a part of the chip containing heat sources, and (b) 3-D view of grid point  $(i,j,k)$ .

further refines or redistributes the grids by sensing the temperature gradient and adding extra grids in the regions with larger gradient, based on the following *weight function* (2.12) and *equidistribution criterion* (2.13) [16]:

$$w(r) = \sqrt{1 + \alpha^2 \left( \frac{\partial T}{\partial r} \right)^2} \quad (2.12)$$

$$\int_{r_i}^{r_{i+1}} w(r) dr = \text{Constant} \quad (2.13)$$

where  $\alpha$  is a user-specified parameter and  $r$  denotes the  $x$  or  $y$  axis. Temperature solutions found using the current and previous grid systems are compared. If the percentage difference is less than a prescribed threshold, then the grid refinement process is terminated. The stopping criterion can also be the user-specified maximum number of grids. According to our empirical observation, tens of neighboring gates can be covered by a single grid rectangle given a 1% error bound. Only a few grids are placed in the  $z$ -direction (thickness) with most grids concentrated at the chip surface near the heat sources. This is because the temperature drops rapidly in the  $z$ -direction away from the surface and larger grid sizes can be used.

A schematic representation of a part of a chip containing several heat sources and the variable grid system is shown in Fig. 2.7. After the coordinates of each heat source are identified, the corresponding grid points into which the heat flows and the proportionate

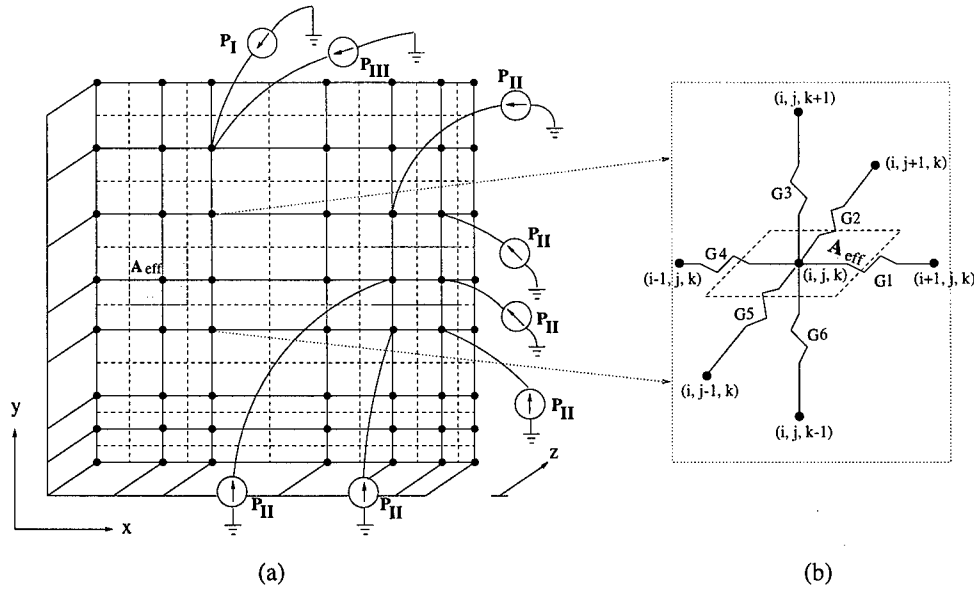


Figure 2.8: (a) Analogous thermal circuit to Fig. 2.7(a), and (b) thermal conductances from  $(i, j, k)$  to adjacent grids.

power values in the analogous thermal circuit are found as shown in Fig. 2.8.  $P_i$  in Fig. 2.8 denotes the heat flow from source  $i$ . The solid lines in Fig. 2.7(a) and Fig. 2.8(a) represent the chosen grid lines and dashed ones are in the middle of two adjacent grid lines.  $A_{eff}$  is the effective area of a grid point. Every heat source that overlaps the effective area of a grid point serves as a power source feeding into that grid, and the corresponding power value is calculated based on the ratio of the source area within  $A_{eff}$  to the total area of the source. In Fig. 2.7(b),  $h_{x+}$ ,  $h_{x-}$ ,  $h_{y+}$ ,  $h_{y-}$ ,  $h_{z+}$ , and  $h_{z-}$  are halves of the distances from grid  $(i, j, k)$  to grid  $(i+1, j, k)$ ,  $(i-1, j, k)$ ,  $(i, j+1, k)$ ,  $(i, j-1, k)$ ,  $(i, j, k+1)$ , and  $(i, j, k-1)$ , respectively. The thermal conductances  $G_1$ ,  $G_2$  and  $G_3$  in Fig. 2.8(b) can be found by applying the first law of thermodynamics on the grid point  $(i, j, k)$ :

$$\begin{aligned}
 & k(h_{y+} + h_{y-})(h_{z+} + h_{z-}) \frac{T_{i+1,j,k} - T_{i,j,k}}{2h_{x+}} + k(h_{y+} + h_{y-})(h_{z+} + h_{z-}) \frac{T_{i-1,j,k} - T_{i,j,k}}{2h_{x-}} + \\
 & k(h_{x+} + h_{x-})(h_{z+} + h_{z-}) \frac{T_{i,j+1,k} - T_{i,j,k}}{2h_{y+}} + k(h_{x+} + h_{x-})(h_{z+} + h_{z-}) \frac{T_{i,j-1,k} - T_{i,j,k}}{2h_{y-}} + \\
 & k(h_{x+} + h_{x-})(h_{y+} + h_{y-}) \frac{T_{i,j,k+1} - T_{i,j,k}}{2h_{z+}} + k(h_{x+} + h_{x-})(h_{y+} + h_{y-}) \frac{T_{i,j,k-1} - T_{i,j,k}}{2h_{z-}} \\
 & = 0
 \end{aligned} \tag{2.14}$$

where  $k$  is the thermal conductivity. From (2.14), we recognize that the heat conduction in a thermal circuit is similar to the current conduction in an electrical circuit with the analogy shown in Fig. 2.9. In other words, we can always map a finite-difference heat conduction problem into an electrical RC network problem. The thermal conductances connected to

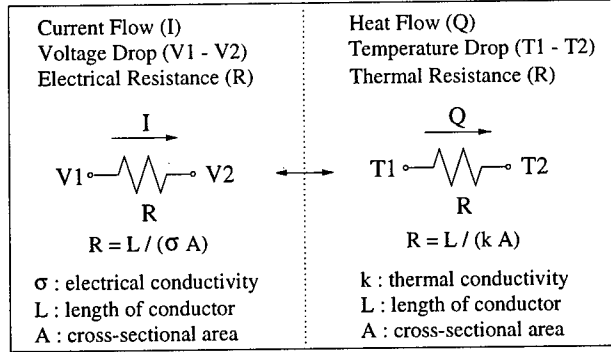


Figure 2.9: Analogy between thermal and electrical circuits.

grid  $(i, j, k)$  can therefore be found as

$$G_1 = \frac{k \cdot (h_{y+} + h_{y-})(h_{z+} + h_{z-})}{2h_{x+}} \quad (2.15)$$

$$G_2 = \frac{k \cdot (h_{x+} + h_{x-})(h_{z+} + h_{z-})}{2h_{y+}} \quad (2.16)$$

$$G_3 = \frac{k \cdot (h_{x+} + h_{x-})(h_{y+} + h_{y-})}{2h_{z+}} \quad (2.17)$$

Similar expressions for  $G_4$ ,  $G_5$  and  $G_6$  can be derived. The resulting analogous electrical circuit is solved by either the sparse matrix or the successive-over-relaxation (SOR) technique, and the on-chip temperatures are obtained. We compute the average temperature of each gate by averaging the temperature values of grids that a gate covers, and then use the updated values as the input to the fast timing simulator for the next simulation run.

The previous discussion describes temperature calculation at the interior grids. However, special care must be taken at the chip boundaries. Figure 2.10(a) illustrates the thermal circuit used to model the top of a chip with the convective boundary condition, where  $T_a$  is the ambient temperature. To find the equivalent thermal resistances for this system, we again apply the first law of thermodynamics on the grid point  $(i, j, k)$ :

$$\begin{aligned} & k(h_{y+} + h_{y-})h_{z-} \frac{T_{i+1,j,k} - T_{i,j,k}}{2h_{x+}} + k(h_{y+} + h_{y-})h_{z-} \frac{T_{i-1,j,k} - T_{i,j,k}}{2h_{x-}} + \\ & k(h_{x+} + h_{x-})h_{z-} \frac{T_{i,j+1,k} - T_{i,j,k}}{2h_{y+}} + k(h_{x+} + h_{x-})h_{z-} \frac{T_{i,j-1,k} - T_{i,j,k}}{2h_{y-}} + \\ & k(h_{x+} + h_{x-})(h_{y+} + h_{y-}) \frac{T_{i,j,k-1} - T_{i,j,k}}{2h_{z-}} + h^e(T_a - T_{i,j,k})(h_{x+} + h_{x-})(h_{y+} + h_{y-}) \end{aligned}$$

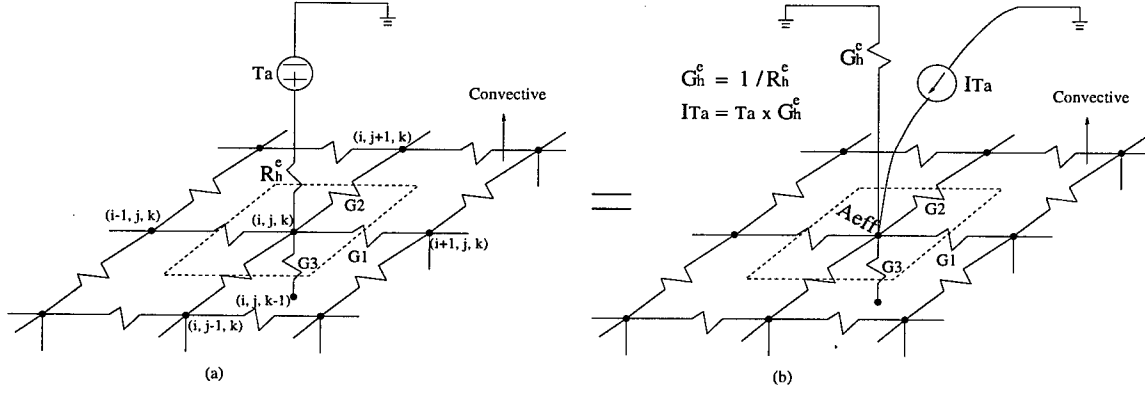


Figure 2.10: Equivalent thermal circuit at the convective boundary.

$$= 0 \quad (2.18)$$

where  $h^e$  is the effective heat transfer coefficient. Using the analogy in Fig. 2.9, we recognize the thermal resistances  $G_1$ ,  $G_2$  and  $G_3$  in Fig. 2.10(a) as

$$G_1 = \frac{kh_{z-}(h_{y+} + h_{y-})}{2h_{x+}} \quad (2.19)$$

$$G_2 = \frac{kh_{z-}(h_{x+} + h_{x-})}{2h_{y+}} \quad (2.20)$$

$$G_3 = \frac{k(h_{x+} + h_{x-})(h_{y+} + h_{y-})}{2h_{z-}} \quad (2.21)$$

Defining  $A_{eff} = (h_{x+} + h_{x-})(h_{y+} + h_{y-})$  as the effective area of the grid  $(i, j, k)$ , we obtain the thermal resistance related to the convective heat transfer ( $R_h^e$  in Fig. 2.10(a)) as

$$R_h^e = \frac{1}{h^e A_{eff}} \quad (2.22)$$

We solve this boundary value problem as follows. First, the circuit in Fig. 2.10(a) is transformed to the equivalent circuit in Fig. 2.10(b), and a 3-D network containing only resistive elements and independent current sources is obtained (capacitive elements are open-circuited in steady state). Next, nodal analysis of this network is performed and the admittance matrix is constructed. Finally, the admittance matrix is solved and the temperature of each node is found. For boundary conditions other than the convective condition, a similar procedure follows after replacing  $h^e$  in (2.22) with  $\infty$  for the isothermal condition and with 0 for the insulated condition.

## 2.5 Verification of ILLIADS-T and Simulation Results

### 2.5.1 Tester chip design and calibration

A tester chip was designed for the verification of simulation accuracy. It was fabricated using 0.8  $\mu\text{m}$  CMOS technology and packaged by MOSIS. Figure 2.11 shows the layout of the chip, where the blocks I, III and V are high-frequency 3-stage ring oscillators designed in a standard super-buffer configuration. Blocks II and IV are 149-stage ring oscillators, and the three small dots (D1, D2, D3) are diodes. Henceforth, we denote the 3-stage and 149-stage oscillators as Rosc3s and Rosc149s, respectively. Each ring oscillator has an enable signal which is used to activate or deactivate the oscillator. Because the operating frequency of the Rosc149s is much lower than that of Rosc3s, power is mainly dissipated from the Rosc3s. The on-chip temperature can be determined by measuring the voltage drop across the forward-biased diodes according to  $V_F = (kT/q) \ln(I_F/I_s(T)+1)$ , where  $I_F$  is the forward-bias current provided by a constant current source. The oscillation frequency of the Rosc149s can also be measured before and after Rosc3s are turned on to observe any change due to the on-chip temperature rise.

Figure 2.12 shows the diode circuit designed for the temperature measurement. Because the voltage drop across the lead resistance of the diode is also a function of temperature, a four-terminal configuration is used to cancel out the voltage drop in the test leads. The diodes are calibrated individually by measuring  $V_F$  at different temperatures. The diode temperature is controlled by placing the chip upside-down on a hot plate after the chip lid has been removed. The temperatures on the surface of the hot plate are accurately determined by placing a *thermistor* on the plate and measuring its resistance values. These values are then translated to the temperatures of the thermistor, namely, the temperatures of the hot plate. The  $I_F$  values are kept small, so that self-heating from the diode may be ignored. An example of the calibration data is shown in Fig. 2.13. When the tester chip is operating, the local temperature near the diode is determined by comparison with the calibration data. The package thermal parameters are also calibrated based on the MOSIS

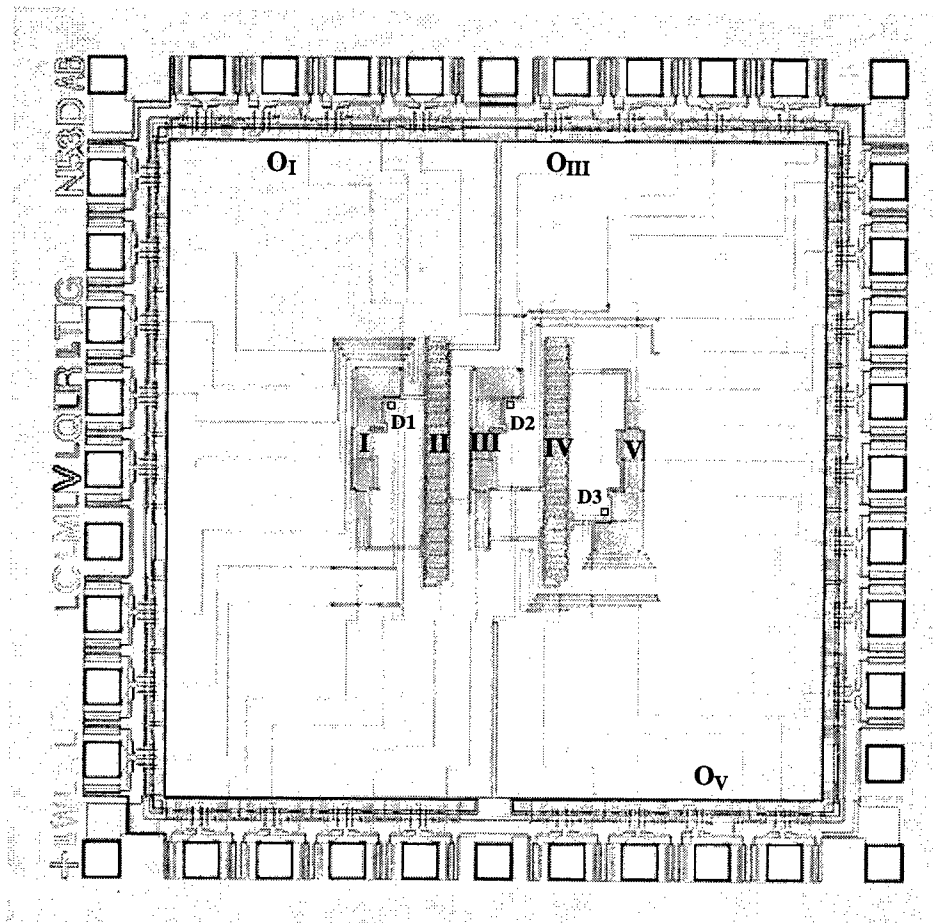


Figure 2.11: Layout of tester chip, where long blocks are Roscl49s and short blocks are Roscl3s.

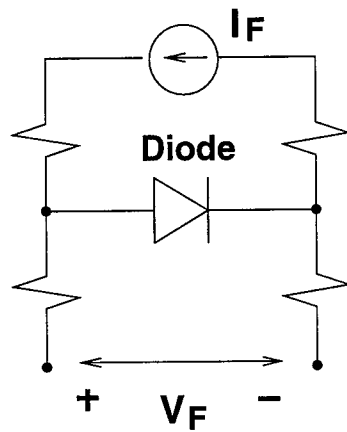


Figure 2.12: Four-terminal configuration for diode measurement.

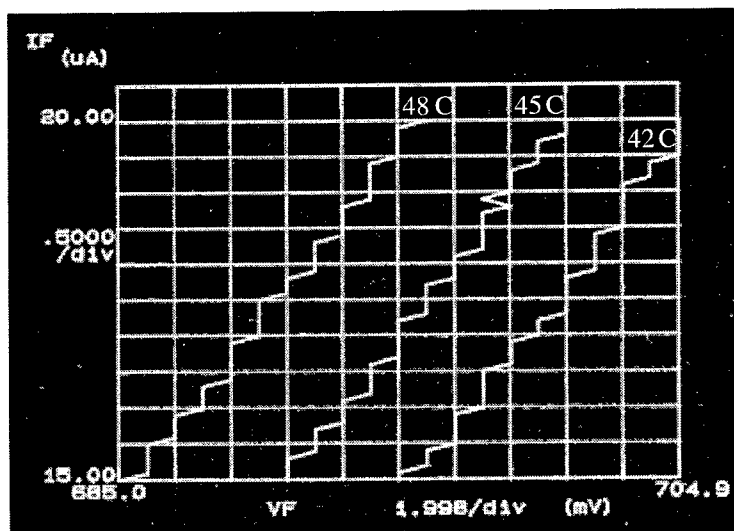


Figure 2.13: Diode calibration example (D1).

handbook for the DIP40 package. The effective heat transfer coefficient of the chip bottom ( $h^e$  in (2.11)) is determined to be 8,689 (W/(m<sup>2</sup> °C)) with all other sides insulated.

### 2.5.2 Verification of ILLIADS-T

During the tester chip experiments, Rosc149s were always activated while the chip power consumption was varied by activating different Rosc3s. Depending on the on/off status of the Rosc3s, there are eight unique experiments as shown in Table 2.1. For example, the ILLIADS-T-simulated temperature profiles for Expt. 2 (block I and III on,<sup>†</sup> block V off) and Expt. 1 (all blocks on) are shown in Figs. 2.14 and 2.15, respectively. Note that the power dissipation from the output buffers of blocks I, III, and V ( $O_I$ ,  $O_{III}$ , and  $O_V$  in Fig. 2.11) was also taken into account. The simulated and measured diode temperatures for all eight experiments are compared in Figs. 2.16 - 2.18. In these figures, the error bars show the spread of the measured data. Good agreement between measured and simulated temperatures was found.

ILLIADS-T was also used to predict the frequency shift of Rosc149s due to the local temperature rise. The mobility-temperature relationship was extracted from frequency measurements on block II, and the mobility model (2.6) was employed to obtain the optimized fitting parameters  $A_1 - A_4$ . Next, the mobility model was used in ILLIADS-T to predict the frequency shift of block IV for the eight experiments, and the results are compared with the measured data as shown in Figs. 2.19 - 2.22. Additional simulation results are presented in Table 2.2, where  $P_{avg}$  is the average power consumption of the chip (including output buffers), and  $T_{blk4}$  is the average temperature of block IV. The oscillation frequencies of block IV before and after electrothermal simulation are shown in the fourth column of Table 2.2. Note that as the temperature increases, the oscillation frequency is significantly

Table 2.1: Activation status of Rosc3s.

Expt. #	1	2	3	4	5	6	7	8
Block I III V	111	110	101	100	011	010	001	000

†: on 0: off.

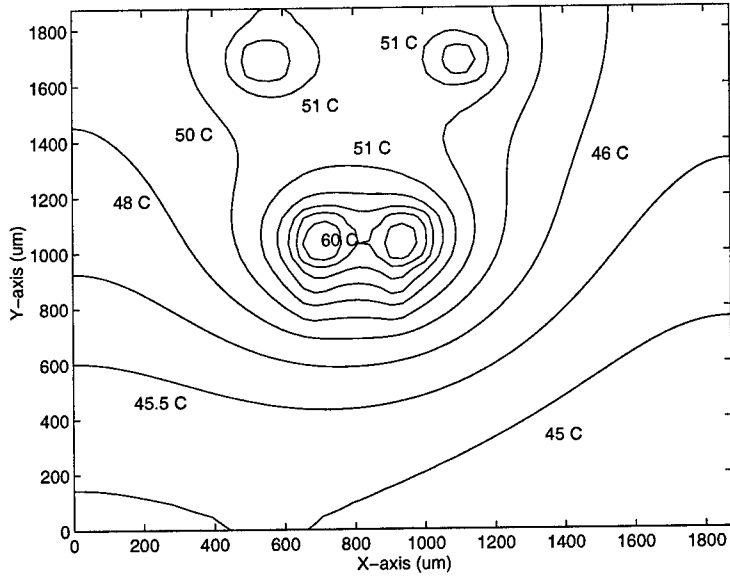


Figure 2.14: Simulated temperature profile for Expt. 2.

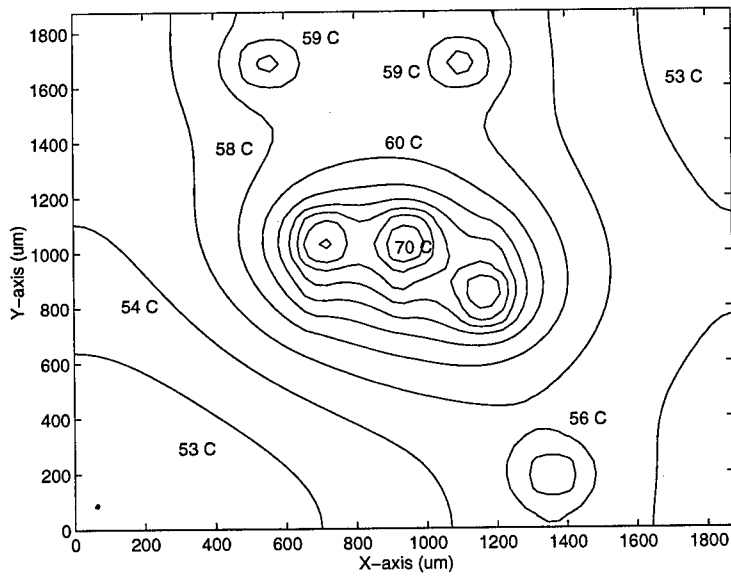


Figure 2.15: Simulated temperature profile for Expt. 1.

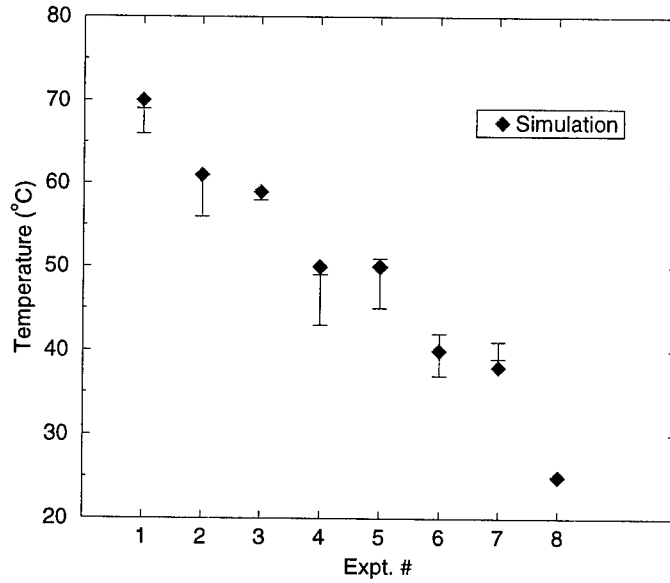


Figure 2.16: Comparison between simulated and measured temperatures for D1.

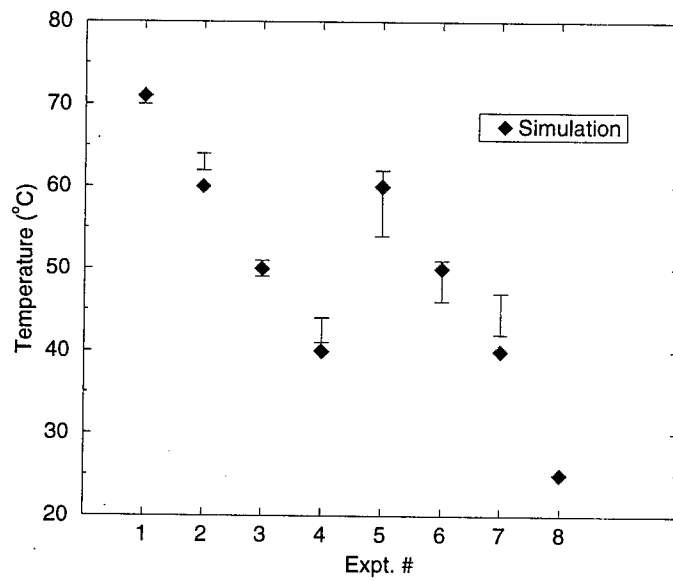


Figure 2.17: Comparison between simulated and measured temperatures for D2.

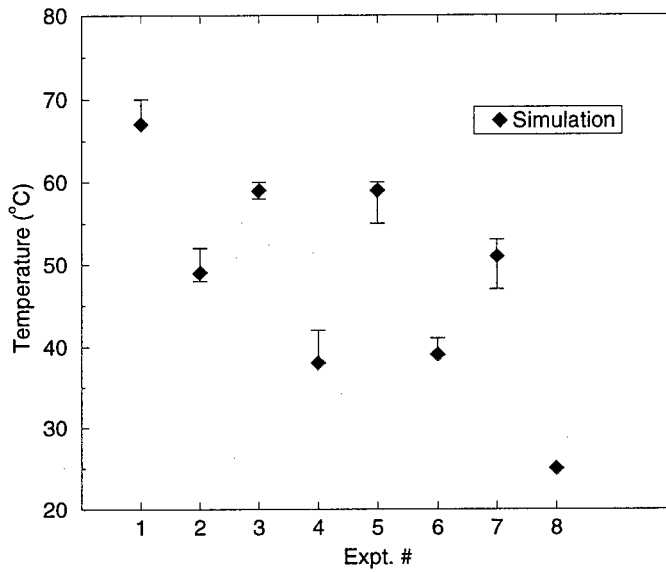
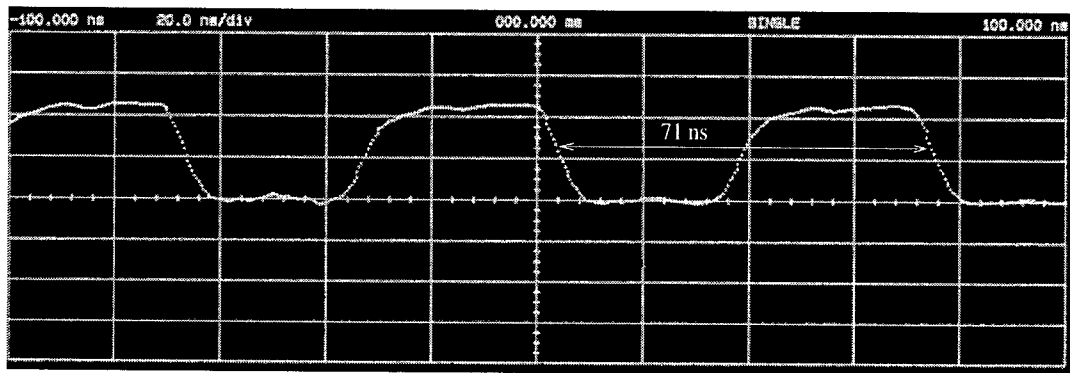


Figure 2.18: Comparison between simulated and measured temperatures for D3.

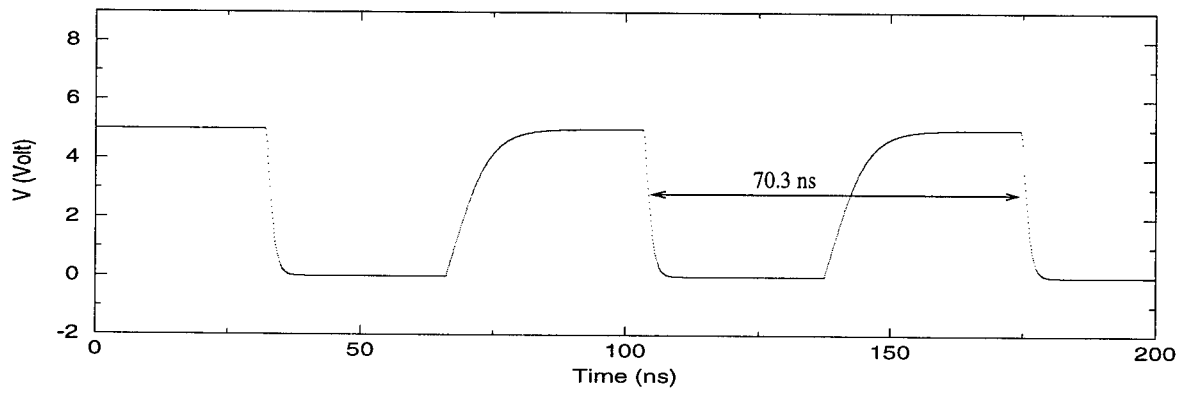
Table 2.2: ILLIADS-T simulation results of the tester chip.

Tester chip	$P_{avg}$ [Watt]	$T_{blk4}$ [°C]	Freq. shift [MHz]	CPU time <sup>1</sup> [sec]
Expt. 7	0.350	44.17	14.07 → 11.53	422
Expt. 5	0.636	56.03	14.07 → 10.38	650
Expt. 1	0.882	63.43	14.07 → 9.62	822

<sup>1</sup>On SUN SPARCstation 10.

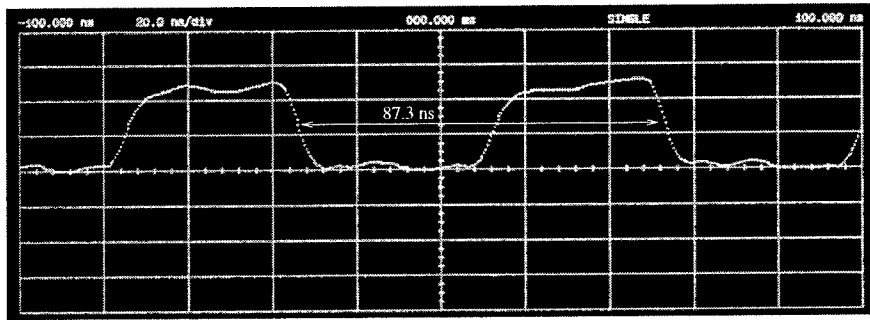


(a)

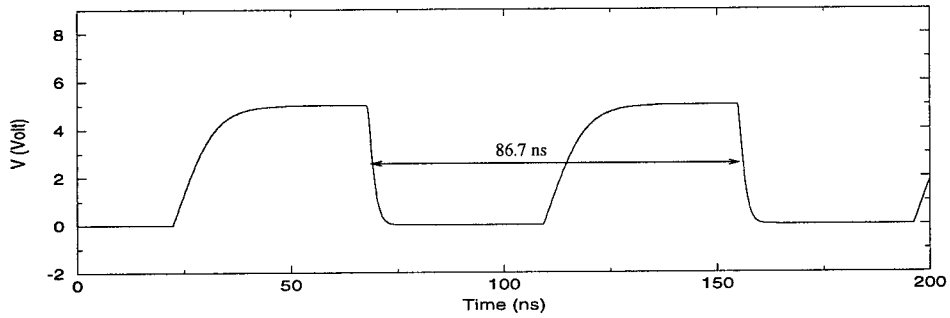


(b)

Figure 2.19: (a) Measured and (b) simulated waveforms for Expt. 8.

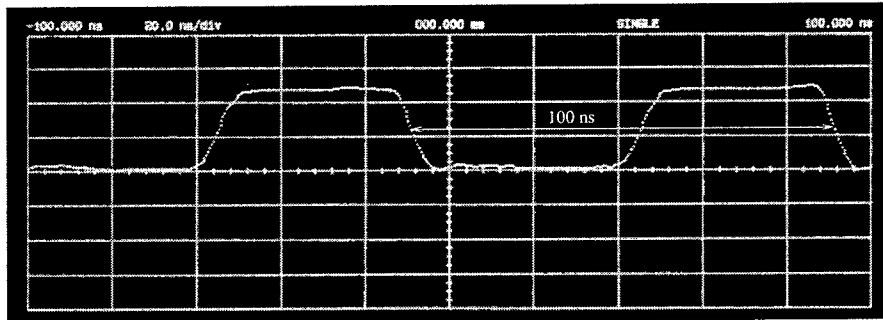


(a)

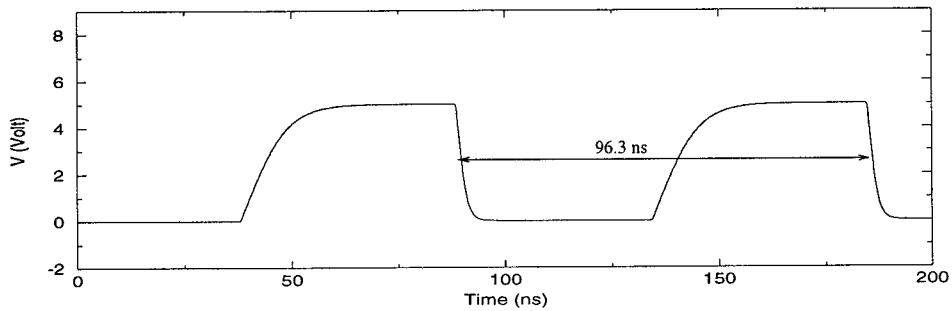


(b)

Figure 2.20: (a) Measured and (b) simulated waveforms for Expt. 7.

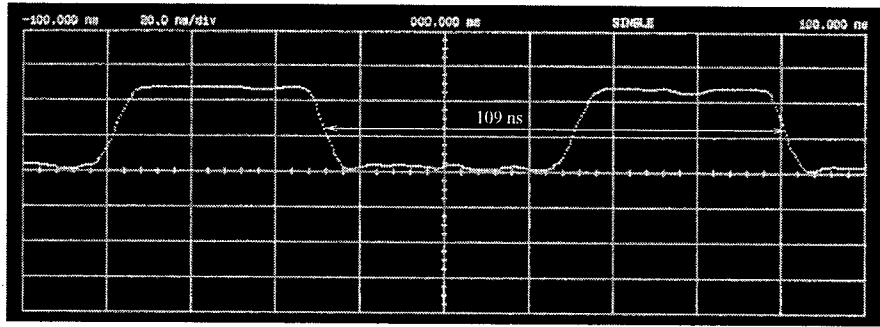


(a)

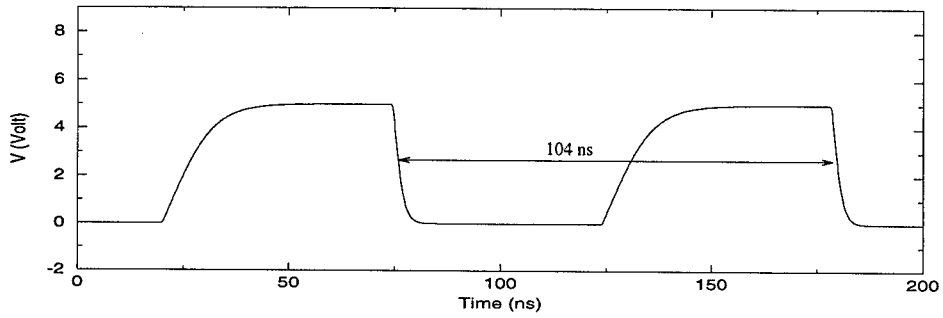


(b)

Figure 2.21: (a) Measured and (b) simulated waveforms for Expt. 5.



(a)



(b)

Figure 2.22: (a) Measured and (b) simulated waveforms for Expt. 1.

lowered and, consequently, so is the power. Therefore, the power values listed in Table 2.2 were calculated at the simulated operating temperature.

Table 2.3: Materials and thermal parameters for the packaging structure.

Parameter Description	Units	Bulk	Bump	Polymer	Substrate	Gel	Bottom Sink
Material	-	doped Si	Sn/Pb	polyimide	doped Si	silicone gel	Al
Therm. Cond. <sup>1</sup>	$W/(mK)$	98.40	53.4	0.25	98.40	0.4	216.5
Thickness	$mm$	0.25	0.025	0.005	0.5	0.005	1.0

<sup>1</sup>Data from [17].

### 2.5.3 ILLIADS-T simulation example

Here, we demonstrate ILLIADS-T simulation results for the example shown in Fig. 2.23. This chip contains two ISCAS85 benchmark circuits, C3540 and C6288, one negative adder, and two three-stage ring oscillators identical to Rosc3. The packaging structure for the chip is shown in Fig. 2.24 and the corresponding thermal parameters are given in Table 2.3. The heat transfer coefficient between the heat sink and the ambience is assumed to be 12,000 ( $W/(m^2K)$ ). Simulation results are presented in Table 2.4, where  $T_{max}$  and  $T_{min}$  are the pinpointed maximum and minimum temperatures of individual circuits. To demonstrate the importance of performing the temperature-dependent simulation, the output waveforms at bit ten of the negative adder, with and without electrothermal simulation, are compared in Fig. 2.25. A logic fault due to a timing problem is identified via the electrothermal simulation. Note that even when the temperature of the negative adder is assigned to 35°C, which is the average chip temperature that would be used in conventional simulations, the thermally induced fault still cannot be detected as shown in Fig. 2.25. This suggests that the on-chip temperature variation must be considered in timing verification, and ILLIADS-T may serve as a useful tool to ensure that the specified timing constraints are met.

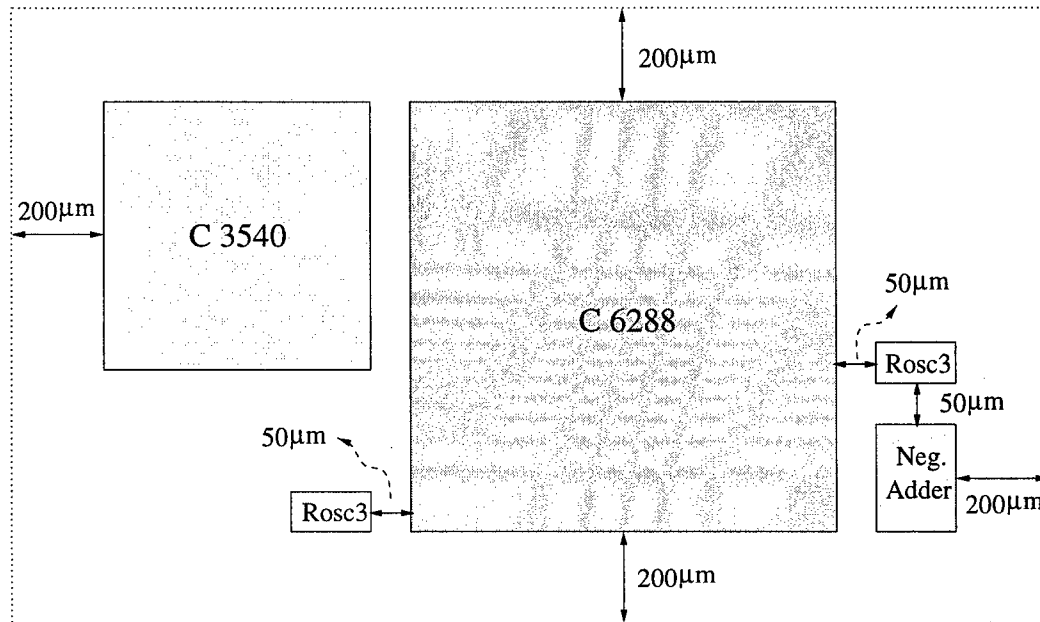


Figure 2.23: Layout of the simulated chip.

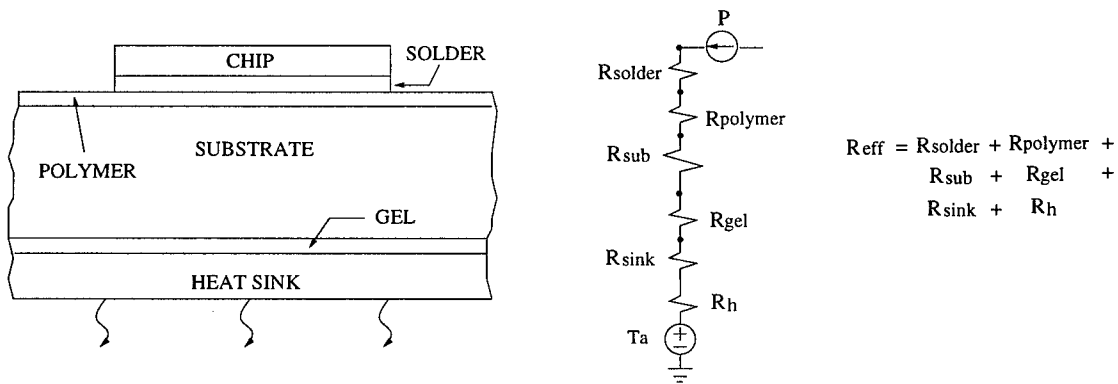


Figure 2.24: Packaging structure used in the simulation example.

Table 2.4: ILLIADS-T simulation results.

Circuit	Ckt. size	$n_{tran}$	$n_{hsrc}$	Freq.	$T_{max}$	$T_{min}$	$n_{run}^1$	CPU <sup>2</sup>
Unit	$\mu\text{m}^2$	-	-	MHz	$^{\circ}\text{C}$	$^{\circ}\text{C}$	-	sec
10-bit Neg. Adder	450 × 300	868	216	100	47.17	38.07	-	-
C3540	1730 × 1540	5842	1656	200	36.82	31.61	-	-
C6288	2180 × 2330	11016	3001	100	44.55	33.10	-	-
Complete Chip	-	-	-	-	-	-	3	3142

<sup>1</sup>Convergence criterion:  $(\Delta T/T_{rise}) < \%1$ .    <sup>2</sup>On SUN SPARCstation 10.

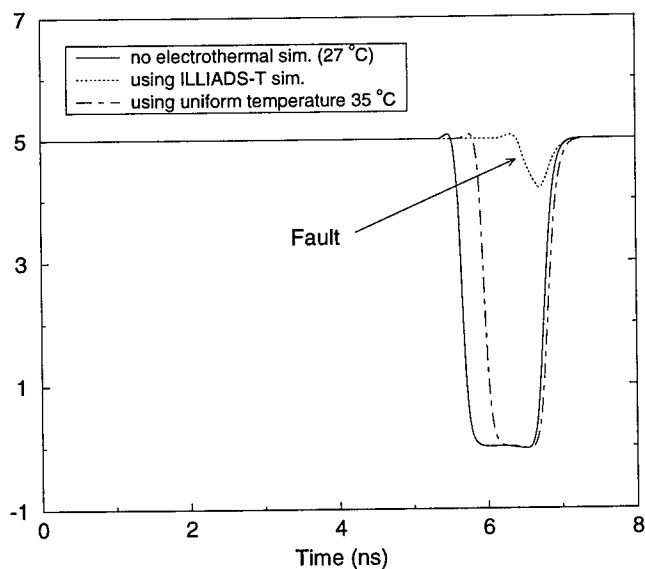


Figure 2.25: Output waveforms of the 10-bit negative adder.

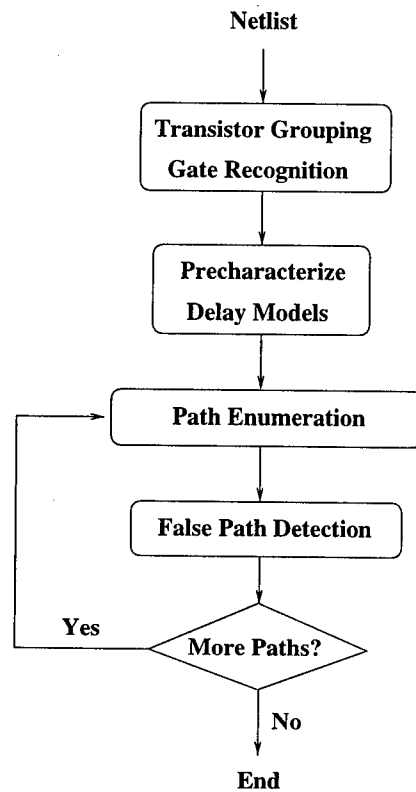


Figure 2.26: Generic flowchart of timing analysis.

## 2.6 Timing Analysis Using ILLIADS-T

### 2.6.1 Introduction

Timing analysis is an important issue in high-performance ULSI circuit design. Over the last decade, designers have increasingly resorted to timing analysis tools to check if a given circuit meets the performance goal (i.e., clock speed). The general timing analysis flow is illustrated in Fig. 2.26.

### 2.6.2 Temperature-dependent gate and RC delays

From both the experimental and simulation results in Section 2.5, we observe that the on-chip temperature profile substantially affects the circuit delay. The critical path timing, affected by the delays of logic gates and interconnects, is also strongly temperature-dependent. Temperature-sensitive timing analysis is thus important for high-performance ULSI chip development.

The temperature-dependent gate delay can be calculated using the RWQ and mobility models introduced in Section 2.2. As for interconnects at given temperatures, we use the following equation to find the resistance value for the temperature-dependent RC delay estimation:

$$R(T) = R_0[1 + \alpha_T(T - T_0)]. \quad (2.23)$$

$R(T)$  in (2.23) is the resistance at temperature  $T$ ,  $R_0$  is the resistance at room temperature  $T_0$ , and  $\alpha_T$  is the temperature coefficient of resistivity (e.g.,  $0.004 \text{ }^\circ\text{C}^{-1}$  for Aluminum). As a general rule of thumb, the RC delay increases about 5% for a  $10 \text{ }^\circ\text{C}$  interconnect temperature rise. To facilitate the RC delay calculation, we have extended the layout extractor developed earlier [20] to extract the signal-line interconnect resistance in the form of a distributed RC tree. More description of our delay modeling will be presented in Section 2.6.3.

### 2.6.3 Monte-Carlo power estimation for on-chip temperature profiling

The power estimation method introduced in Section 2.1.1 is strongly input pattern-dependent since it requires the user to specify complete information about the input patterns. However, input signals are generally unknown during the early design phase and it is practically impossible to estimate the power by simulating the circuit for all possible inputs. In order to estimate the steady-state temperature profile for the temperature-dependent timing analysis, it is more useful to obtain the average power in a statistical manner. A Monte-Carlo based approach is thus more suitable for finding the typical on-chip temperature profile.

Here we employ the technique called MED [21] for the average power estimation. Ac-

According to the central limit theorem,  $\bar{x}$  is a value of a random variable with mean  $\eta$  whose distribution approaches the normal distribution for large  $N$ . With  $(1 - \alpha)$  confidence it then follows that [22]

$$-z_{\alpha/2} \leq \frac{\bar{x} - \eta}{\sigma/\sqrt{N}} \leq z_{\alpha/2}, \quad (2.24)$$

where  $\sigma$  is the standard deviation, and  $z_{\alpha/2}$  is defined so that the area to its right under the standard normal distribution curve is equal to  $\alpha/2$ .  $\bar{x}$  here is defined as the sample mean of the *transition density* or the power value of a given DCCB in the circuit. For sufficiently large number of  $N$  (i.e.,  $N \geq 30$ ),  $\sigma$  can be approximated by the sample standard deviation  $s$ . By using (2.24), one can show that the number of samples needed is

$$N \geq \left( \frac{z_{\alpha/2} s}{\bar{x} \epsilon_1} \right)^2 \quad (2.25)$$

such that we have  $(1 - \alpha)$  confidence that the following equation is satisfied.

$$\frac{|\bar{x} - \eta|}{\eta} \leq \frac{\epsilon_1}{1 - \epsilon_1} = \epsilon, \quad (2.26)$$

where  $\epsilon$  is defined to be a user-specified error tolerance. Thus (2.25) provides a stopping criterion to yield the accuracy specified in (2.26) with confidence  $(1 - \alpha)$ . It is clear from (2.25) that for a small value of  $\bar{x}$ , the number of samples required can be very large to meet the specified accuracy level. In MED-like approaches, the stopping criterion (2.25) is used for the DCCBs which have  $\bar{x}$  larger than a user-specified threshold value,  $\eta_{min}$ . Those DCCBs are referred to as regular-density DCCBs. A different stopping criterion is used for the DCCBs which have  $\bar{x}$  less than  $\eta_{min}$ :

$$N \geq \left( \frac{z_{\alpha/2} s}{\eta_{min} \epsilon_1} \right)^2. \quad (2.27)$$

Those DCCBs are referred to as low-density DCCBs. From (2.27) an absolute error bound can be provided for the low-density DCCBs. Although it usually takes longer time for those DCCBs to converge, they have the least effect on the circuit power and reliability [21].

Additional remarks are noted here. Firstly, we do not consider the external spatial correlation of the input signal, and the circuit is given a sequence of two input vectors

for one iteration of logic simulation. All possible input patterns (high, low, high-to-low, low-to-high) are assumed to have equal probability to occur. Secondly, our logic simulator calculates the power and delay for each gate by taking into account different input slopes, load capacitances, MOS device parameters, temperatures, and interconnects. Each signal-line interconnect network is transformed to an equivalent  $\pi$ -model and is lumped to the corresponding driving gate. The state equation (Riccati differential equation) of the gate is then solved analytically to give the accurate power and delay values [23].

#### **2.6.4 Thermal simulation for timing analysis**

To determine the typical steady-state temperature profile of the chip substrate, the gate power values obtained by using the Monte-Carlo simulation are inputted to our 3-D thermal simulator. Recall that in order to find the steady-state temperature, an iterative procedure should be invoked between the power and temperature calculations (i.e., power and temperature are functions of each other, and the details were discussed in Chapter 2.1). From our empirical observation, when the iteration count of the Monte-Carlo power and temperature calculations exceeds two, the resulting temperature values are actually very close to the final steady-state values. Therefore, here we limit the number of iterations to two as our empirical convergence criterion for the Monte-Carlo power and temperature simulations.

To find the signal-line interconnect temperature, we first extract the coordinates of each metal. Next we assign the localized substrate temperature found earlier to the metals at the same X-Y location. In this simple approach, the temperature difference between the substrate and the multi-layered signal-line metals is ignored. This is reasonable for a typical signal line carrying a normal amount of current since the temperature rise due to Joule heating is relatively small because of the small current density. However, if the line width decreases or the current increases substantially, the Joule heating needs to be taken into account for the multi-layered interconnect system.

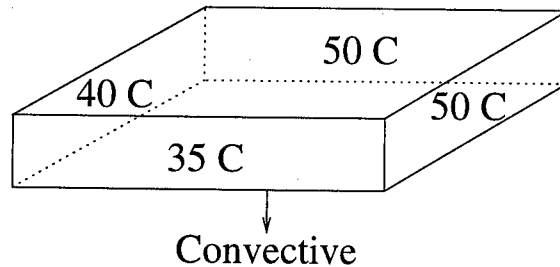


Figure 2.27: Thermal boundary conditions used for the temperature-dependent timing simulation.

## 2.7 Simulation Results

Finding the critical path and the possible input patterns which trigger this path involves the work of path enumeration and false path detection. In the current implementation, we do not attempt to identify the real critical path. Instead, we assume that the input pattern(s) which triggers the critical path, i.e., critical pattern, is given. The temperature-dependent critical path is then identified based on the provided critical pattern.

During the Monte-Carlo power simulation phase, we concurrently search for the longest path delay and its corresponding input pattern. In other words, if the number of samples needed in the Monte-Carlo simulation is  $N$ , the longest delay and its triggering pattern are found out of the  $N$  input patterns. We then consider the pattern thus obtained as the critical pattern and it will be used to identify and report the DCCBs along the longest path. The critical pattern is acquired as the byproduct of the Monte-Carlo power simulation.

In the remainder of this section, we will use ISCAS85 benchmark circuits to demonstrate our simulation results. Figure 2.27 shows the thermal boundary conditions used for all circuits under simulation: The four sides are set to be in the isothermal condition, i.e., constant temperatures, and the top is perfectly insulated and the bottom is convective to room temperature with the heat transfer coefficient 5,000. Simulation results for the

Table 2.5: Simulation results of ISCAS85 benchmark circuits.

Circuit	Power	$T_{dccb\_max}$	$T_{dccb\_min}$	Delay	MC CPU	Therm CPU
Unit	mW	$^{\circ}C$	$^{\circ}C$	ns	sec	sec
C432	7.1	47.69	36.15	6.68	228.4	63.8
C499	15.0	49.90	37.14	5.51	397.6	131.6
C880	11.1	49.43	38.29	4.67	340.7	96.3
C1355	12.8	51.02	36.20	5.52	656.8	413.9
C3540	41.1	49.88	35.38	8.39	1522.9	1331.2
C6288	380.7	50.04	35.23	17.8	10094	3494

temperature-dependent Monte-Carlo power estimation and critical path delay calculation are demonstrated in Table 2.5. The 95% confidence ( $1 - \alpha = 0.95$ ), 5% error tolerance ( $\epsilon = 0.05$ ), and  $\eta_{min} = 0.2$  are used in the Monte-Carlo power simulation. The estimated circuit powers are shown in the second column in Table 2.5.  $T_{dccb\_max}$  and  $T_{dccb\_min}$  are the simulated maximum and minimum temperatures of the DCCBs on the longest path, respectively. The estimated temperature-dependent longest path delays are shown in the fifth column. The CPU time (on SUN SPARCstation 10) used for the Monte-Carlo power and thermal simulation are also given in Table 2.5. Finally, the temperature profile of C6288 is demonstrated in Fig. 2.28. The DCCBs on the longest path of C6288 are also shown in Fig. 2.28.

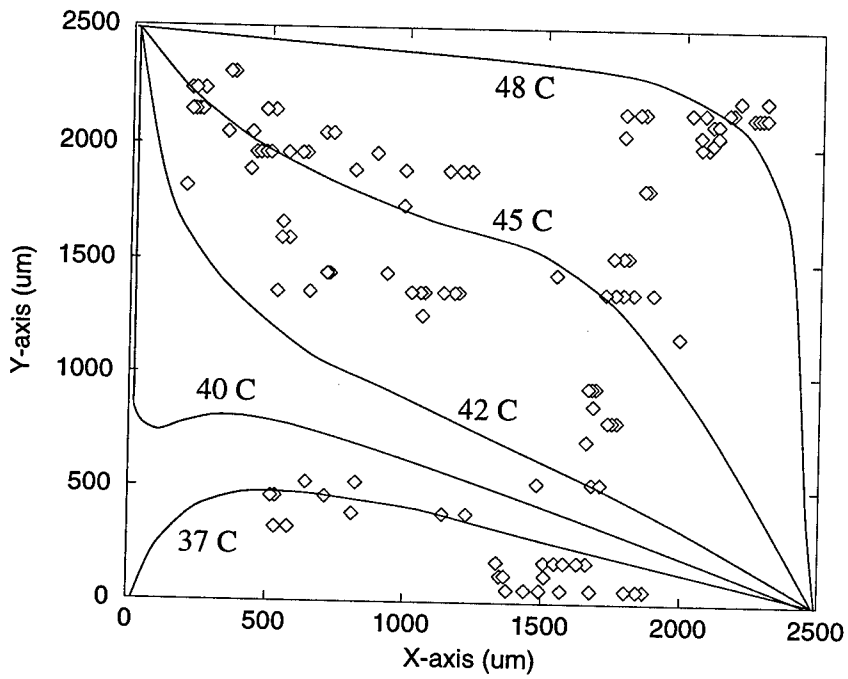


Figure 2.28: The temperature profile and the DCCB distribution on the longest path of C6288: The solid lines are the simulated temperature contour and the small diamonds are DCCBs on the longest path.

### 3 STATISTICAL POWER ANALYSIS OF SEQUENTIAL CIRCUITS

Pattern-independent, temperature-dependent electromigration diagnosis requires average power dissipation for substrate temperature profiling and average current density through each segment of on-chip interconnect. Thus, accurate power/current analysis is crucial to the success of the EM diagnosis tool. Circuits may be classified into two categories: combinational circuits and sequential circuits. Techniques for accurate power analysis are available for combinational circuits (e.g., [27, 28]). Unfortunately, not much success has been achieved for such analysis in sequential circuits due to the difficulty in collecting meaningful power data. Here, power analysis is performed using a statistical approach since the accuracy of probabilistic approaches is likely to be poor. A sequential circuit contains both primary inputs and latch inputs. While the switching characteristics of primary inputs are determined by the operating environment, those of the latch inputs depend on the finite-state machine (FSM) that is implemented by the circuit. This causes spatial and temporal correlations among latch input signals, which further induces temporal correlations among power data. On the other hand, statistical techniques require a random sample, i.e., a sample of *independent and identically distributed* (iid) power data, for mean estimation. Thus, the feedback mechanism characteristic of sequential circuits greatly increases the complexity of the power estimation task.

As a solution to this problem, we propose a new statistical approach, as depicted in Fig. 3.29. This approach takes full account of signal correlations among latches as well as internal nodes. In sequential circuits, power dissipation in consecutive clock cycles are temporally correlated. In order to construct a random power sample, we need to find a proper *independence interval* over which the circuit should be simulated between two power sampling cycles such that the acquired power data are iid. As shown in Fig. 3.30, the independence interval is selected via a *randomness test*. The randomness test examines the validity of the hypothesis that a power sequence is composed of iid's by *accepting* or *rejecting*

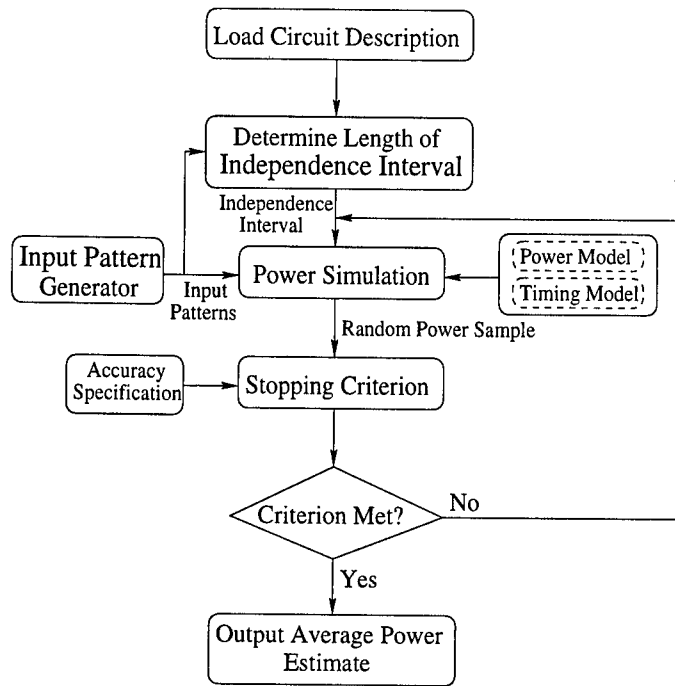


Figure 3.29: Flowchart of the proposed power estimation approach.

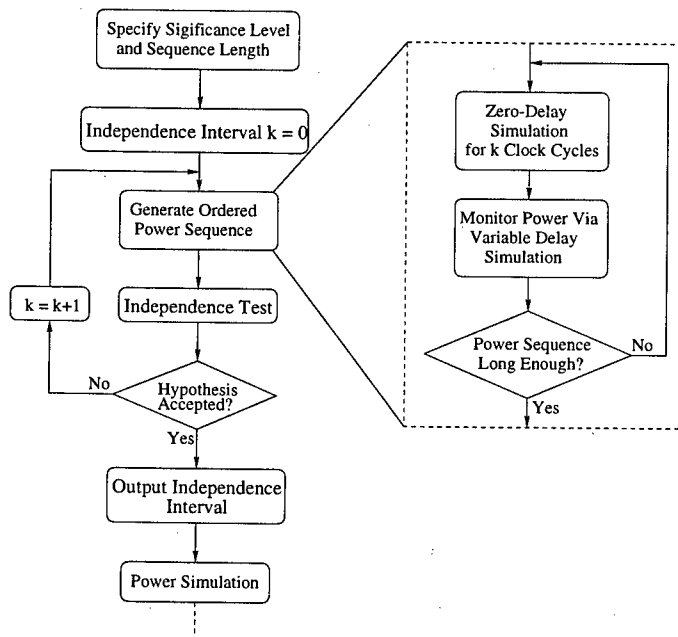


Figure 3.30: Iteration procedure for selecting a proper independence interval.

the hypothesis according the statistical evidence gathered from the sequence. At a trial independence interval, if the hypothesis is accepted with a user-specified *significance level*, the sequence can be viewed as a random sample. Otherwise, the trial interval is incremented and another power sequence is collected. This procedure continues until the hypothesis is accepted and the associated independence interval is used hereafter to generate random power samples. A distribution-independent stopping criterion is then used to continuously analyze the power sample data and control the sample size until the desired accuracy is achieved. In addition to the high estimation accuracy achieved by considering all signal correlations, the simulation efficiency is also greatly improved by the dynamic selection mechanism of the independence interval.

We tested the proposed approach on a set of ISCAS89 benchmark circuits on a SPARC 20 workstation with 244 MB memory. All circuits are assumed to operate at a clock frequency of 20MHz with a 5V power supply. The significance level of the randomness test is set to 0.20 while the maximum error allowed was specified as 5% with 0.99 confidence. The signals at the primary inputs are assumed to be mutually independent and have probability of 0.5. However, correlated input streams can also be handled without any extra effort since no assumption is made on input pattern statistics. The power sequence length for the randomness test should be carefully selected. It should not be too long because simulation efficiency may be degraded by a lengthy independence interval selection procedure. Neither can it be too short because statistical fluctuations in hypothesis test results reduce with increasing test sample size. In the following experiments, the power sequence length is chosen to be 320 because the gain in statistical stability of the test results is marginal if it is any longer.

Table 3.6 shows the power estimation results for the test circuits. In Table 3.6, SIM is the sample average power obtained from finding the average of power over one million consecutive clock cycles. It is deemed a sufficiently accurate estimate of the real average power, and is used as the reference for all experiments. In the table, *I.I.* is the independence interval determined by the randomness test. The average power estimate  $\bar{\mu}_p$  is obtained

Circuit Name	SIM (mW)	<i>I.I.</i>	$\bar{\mu}_p$ (mW)	Sample Size	CPU Time (sec)
s208	0.276	2	0.276	4928	138.8
s298	0.430	2	0.429	2816	73.6
s344	0.751	1	0.751	960	14.6
s349	0.785	2	0.785	1088	21.8
s382	0.433	2	0.433	2176	75.6
s386	0.519	1	0.518	1728	35.4
s400	0.418	2	0.420	2272	52.7
s420	0.353	2	0.354	4576	195.0
s444	0.427	3	0.427	2400	69.9
s510	1.175	1	1.175	3168	114.7
s526	0.443	1	0.434	2176	53.1
s641	0.786	1	0.787	1088	26.1
s713	0.804	1	0.804	1088	26.2
s820	0.957	1	0.957	1952	58.2
s832	0.941	3	0.941	2080	75.1
s838	0.443	3	0.443	2272	149.4
s1196	3.080	1	3.079	608	26.7
s1238	3.009	0	3.010	576	24.4
s1423	2.773	1	2.774	2368	275.0
s1488	1.844	2	1.844	4000	293.0
s1494	1.735	5	1.735	3936	392.5
s5378	6.667	2	6.659	352	51.9
s9234	2.008	1	2.008	704	79.6
s15850	5.939	1	5.938	896	462.8

Table 3.6: Power estimation results.

by taking the average of the sample whose size is listed in column Sample Size. The CPU time usage is reported in the last column. For all test circuits, the proposed approach produces accurate average power estimates within a reasonable amount of CPU time. It is also observed that usually an independence interval of a few clock cycles is sufficient for the randomness hypothesis to be accepted with the specified significance level. This observation agrees with [29] on that a small unrolling factor of a FSM is generally enough for accurate power estimation. Furthermore, the duration of the independence interval is determined dynamically and varies with the target circuit. Hence simulation efficiency is greatly improved by not assigning a pessimistic warm-up period *a priori* [30].

To understand the average performance of the proposed technique, we conducted 1,000 simulation runs for every circuit and summarized the results in Table 3.7. In this table,  $I.I_{min}$ ,  $I.I_{max}$  and  $I.I_{avg}$  are the minimum, maximum and average independence interval, respectively.  $S_{avg}$  is the average sample size and  $D_{avg}$  is the average percentage deviation of the estimation results from the reference value. In order to consider the deviation of both polarities,  $D_{avg}$  is estimated by

$$D_{avg} = \frac{1}{N} \sum_{i=1}^N \frac{|P_{exact} - P_{estimate}|}{P_{exact}} \times 100\%, \quad (3.28)$$

where  $N$  is the number of simulation runs. In this table, the independence interval varies somewhat because the randomness test provides a statistical rather than a deterministic conclusion on whether or not the power sequence is “random enough.” Thus, we do obtain a fixed independence interval. Nevertheless, Table 3.7 shows that the estimation results indeed meet the accuracy specification with very low average deviation. The accuracy and robustness of the technique are therefore demonstrated.

Circuit	$II_{min}$	$II_{max}$	$II_{avg}$	$S_{avg}$	$D_{avg}$	Err(%)
s208	0	5	1.60	5015	0.87	0.0
s298	1	2	1.60	2650	1.13	0.0
s344	1	3	1.45	964	0.99	0.0
s349	1	3	1.40	959	1.03	0.0
s382	0	3	1.65	2247	1.14	0.0
s386	1	2	1.20	1788	1.03	0.0
s400	1	4	1.85	2291	1.08	0.0
s420	0	6	1.39	4265	1.25	0.9
s510	0	5	1.20	3141	0.99	0.0
s526	1	3	1.20	2230	1.12	0.0
s641	0	3	0.85	1077	0.96	0.0
s713	0	2	0.70	1098	1.01	0.0
s820	0	2	1.20	1951	1.04	0.0
s832	0	3	1.30	2044	1.03	0.0
s838	1	3	1.40	2833	1.79	1.4
s1196	0	3	0.85	586	0.97	0.0
s1238	0	1	0.40	567	1.00	0.0
s1423	1	3	1.60	2416	1.07	0.0
s1488	1	4	2.20	4012	1.27	0.1
s1494	1	5	2.60	4009	1.14	0.0
s5378	1	10	2.40	352	1.40	0.7
s9234	1	6	1.95	894	0.91	0.0
s15850	1	3	1.20	900	1.15	0.0

Table 3.7: Large number simulation summary.

The proposed approach can be used to estimate the average power dissipation and used by a heat equation solver for substrate temperature profiling. It can be also used to estimate the average current densities through interconnects, which is necessary to calculate the mean time-to-failure and evaluate the impact of joule heating on EM lifetime. Extension and application of this work to temperature-dependent, pattern-independent EM diagnosis is the subject of future research.

## 4 ESD PROTECTION CIRCUIT DESIGN AND ANALYSIS USING iETSIM

### 4.1 Electrothermal Models in iETSIM

#### 4.1.1 MOS snapback model

The MOS snapback model [31][32] is represented in Fig. 4.31. The BSIM3 model is used to describe the operation outside the snapback regime. The five current sources in Fig. 4.31 are  $I_{ds}$  the drain current used in the BSIM3 MOS model [33],  $I_{ii}$  the impact ionization generation current,  $I_b$  the base current,  $I_c$  the collector current, and  $I_{th}$  drain-body thermal generation current.

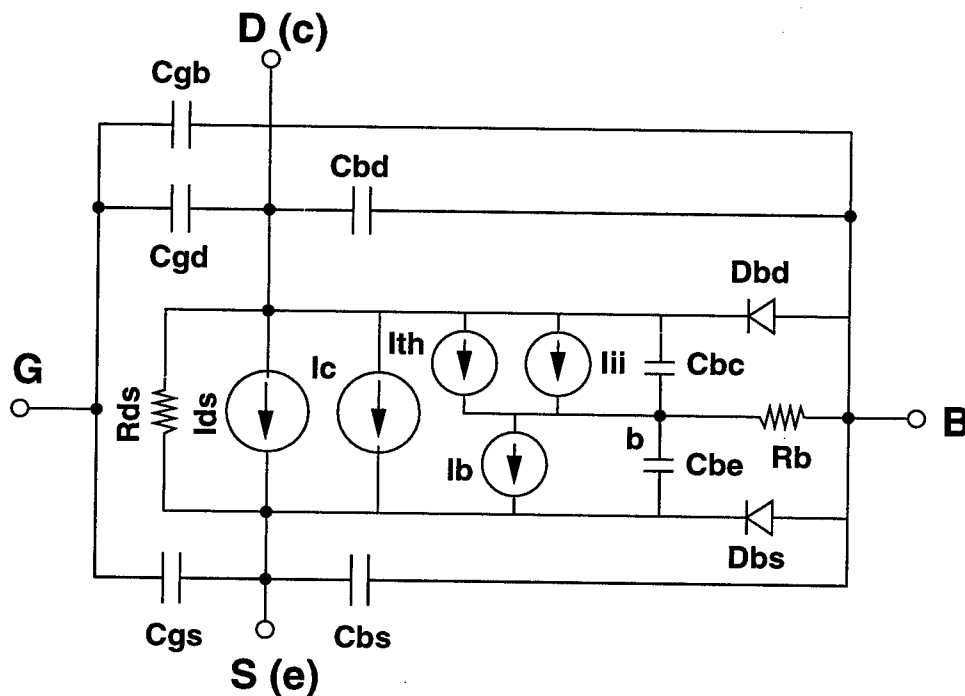


Figure 4.31: The MOS model for snapback simulation.

The circuit elements  $C_{be}$  and  $C_{bc}$  account for the finite diffusion time of the minority electrons across the base region. These capacitances control the turn-on time of the parasitic

bipolar device which is important for simulation of fast ESD transients.

The current source  $I_{th}$  is added between the drain and internal base node for electrothermal simulations. The thermal generation current depends on the drain junction temperature as

$$I_{th}(T) = I_{th}(T_o) \left( \frac{T}{T_o} \right)^{3+\gamma/2} \exp \left( \frac{qE_g(T_o)}{kT_o} - \frac{qE_g(T)}{kT} \right) \quad (4.29)$$

where  $E_g(T) = 1.17 - \frac{4.73 \times 10^{-4} T^2}{T+636}$ , and  $\gamma$  is a constant.

#### 4.1.2 Thermal model

By modeling device behavior up to the onset of second breakdown, we can determine the safe operating limits for a given protection circuit. Second breakdown is thermally originated and its model should be based on the solution of the heat diffusion equation. The temperature at the drain junction (TDJ) controls the thermal generation current ( $I_{th}$ ) while the source junction temperature (TSJ) affects the saturation currents  $I_{oe}$  and  $I_{oc}$ .

The temperature distribution in the vicinity of a heat source is obtained by solving the heat diffusion equation

$$\frac{\partial T}{\partial t} - \frac{\kappa}{\rho C_p} \nabla^2 T = \frac{P(t)}{\Delta \rho C_p} = \frac{I(t) V_H}{\Delta \rho C_p} \quad (4.30)$$

where  $T$  is the temperature,  $\kappa$  is the thermal conductivity,  $\rho$  is the density,  $C_p$  is the specific heat,  $P(t)$  is the power generated by a source in a volume  $\Delta$ ,  $I(t)$  is the current flowing through the device and  $V_H$  is the holding voltage. For a rectangular box with volume  $\Delta = abc$  which dissipates power  $P$ , the solution of the heat diffusion equation [34] yields

$$T(\vec{r}, t) = T_o + \frac{\zeta}{\rho c_p \Delta} \times \int_0^t PG(x, a, \tau) G(y, b, \tau) G(z, c, \tau) d\tau \quad (4.31)$$

where

$$G(x, a, \tau) \triangleq \frac{1}{2} \left\{ \operatorname{erf} \left( \frac{a/2+x}{2\sqrt{\tau\kappa/\rho c_p}} \right) + \operatorname{erf} \left( \frac{a/2-x}{2\sqrt{\tau\kappa/\rho c_p}} \right) \right\}$$

and the factor  $\zeta = 2$  for a semi-infinite medium. The integral over time in (4.31) can be evaluated using an electrical equivalent integrator circuit [3] as shown in Fig. 4.32. The time

dependent resistor  $R(t)$  is given by the following equation which can be derived from (4.31). A similar circuit can estimate the temperature for the source junction.

$$R(t) = \frac{\rho c_p}{\zeta G(x, a, t) G(y, b, t) G(z, c, t)} \quad (4.32)$$

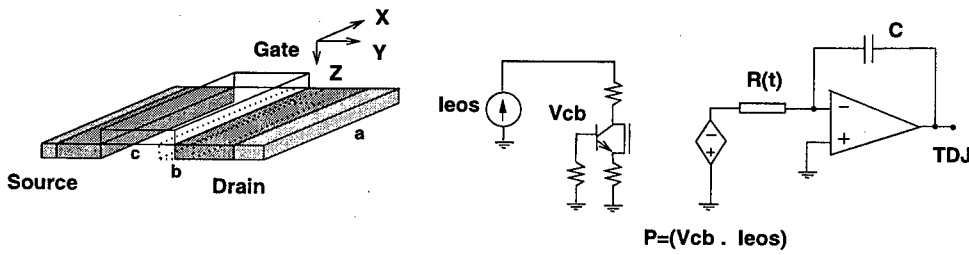


Figure 4.32: 3D view of a MOS transistor and the integrator circuit to estimate the drain junction temperature.

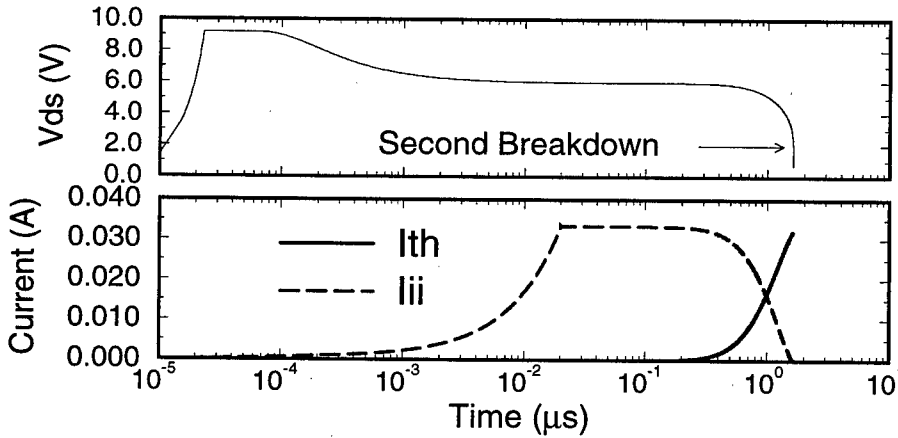


Figure 4.33: iETSIM simulation of second breakdown in a grounded gate NMOS device ( $W=20\mu\text{m}$  and  $L=0.4\mu\text{m}$ ).

Figure 4.33 shows the simulated transient response of an NMOS device subjected to a drain current stress of 120mA. At around  $1.7\mu\text{s}$ , the device enters the second breakdown and the pad voltage collapses. The second breakdown is detected when the thermal generation current increases significantly.

### 4.1.3 Diffused resistor model

Low-field transport is described by the ideal ohmic relation

$$I_n = qn\mu_n W X_j V / L \quad (4.33)$$

where  $I_n$  is the electron current,  $V$  is the voltage across the resistor,  $L$  is the effective resistor length,  $X_j$  is the effective junction depth, and  $\mu_n$  is the electron mobility which depends on the doping level ( $N_d$ ). At high fields, velocity saturation occurs and the current is limited to a saturation value given by

$$I_n = qnW X_j v_{sat} \quad (4.34)$$

where  $v_{sat}$  is the saturation velocity of electrons in silicon. A model that accounts for both the low and high field effects is written as

$$I_n = \frac{qnW X_j V}{L \left( \frac{1}{\mu_n} + \frac{V}{v_{sat}L} \right)} \quad (4.35)$$

### 4.1.4 SCR model

The circuit in Fig. 4.34 is used for modeling the SCR. It consists of two bipolar transistors, well and substrate resistances. The bipolar model is based on the Gummel-Poon model and is enhanced to incorporate the avalanche breakdown effect. The breakdown parameters of the lateral NPN transistor are tuned to match the experimental trigger voltage [31].  $R_{on}$  is used to adjust the SCR on-resistance.

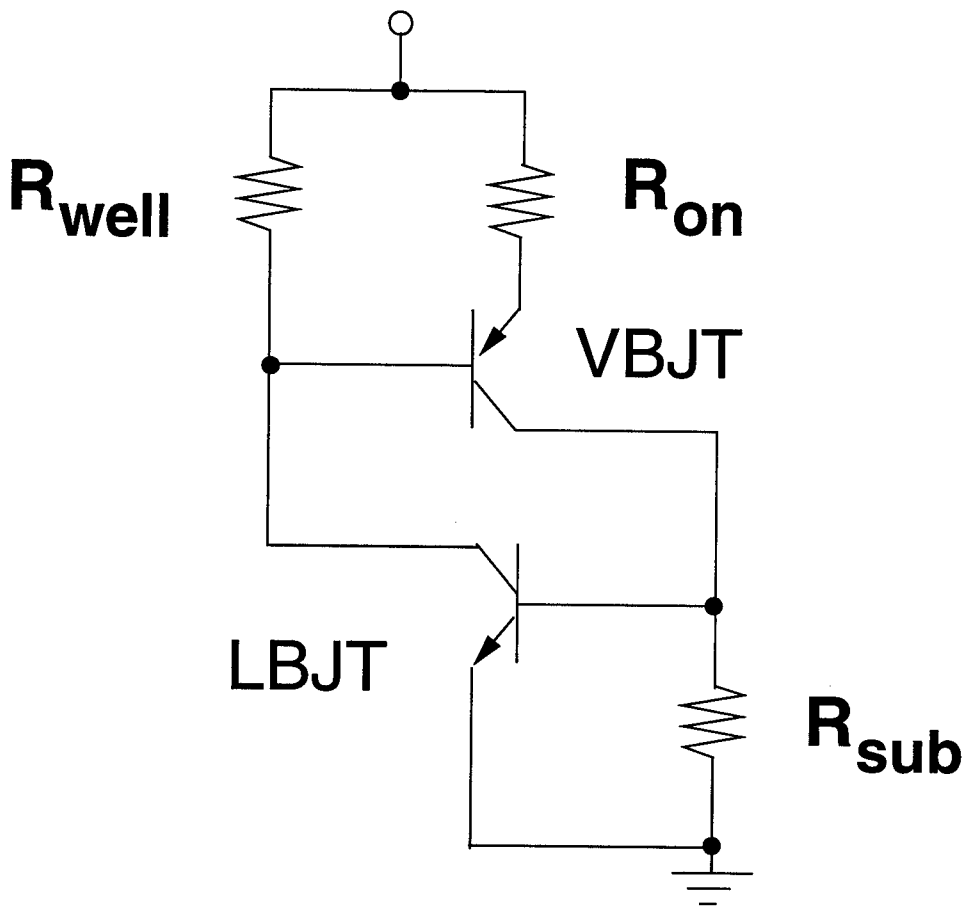


Figure 4.34: The circuit model for an SCR.

## 4.2 Simulation and Optimization of Multi-Finger Protection NMOS-FETS

### 4.2.1 Introduction

The electrothermal circuit simulator iETSIM [3][35], which solves the heat diffusion equation simultaneously with the device electrical models, was used to study the thermal coupling effect and to determine the layout dependent behavior of NMOSFETS. For a silicided CMOS technology, the contact-to-gate spacing is usually kept at a minimum. In this section, we propose an optimization methodology for the layout of multiple-gate-finger NMOS devices in silicided technologies.

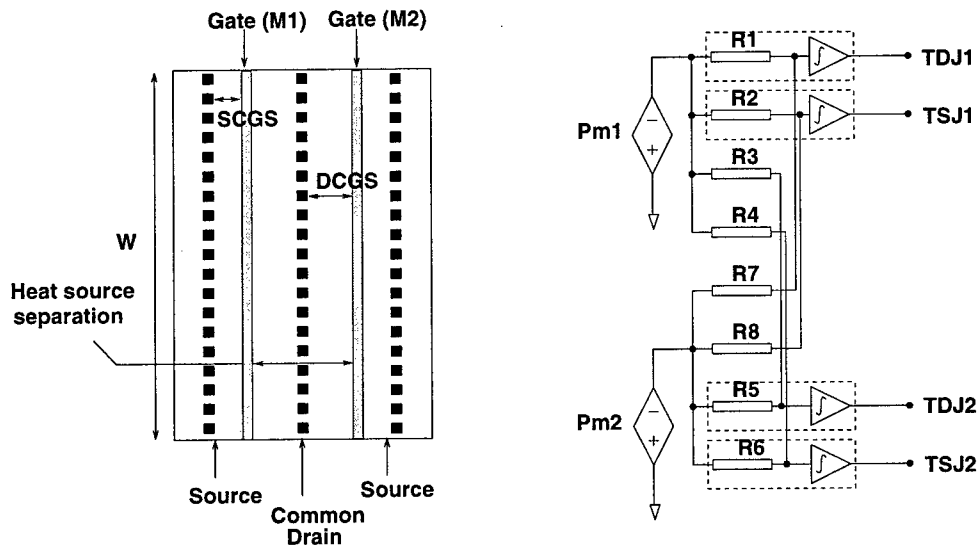


Figure 4.35: A two-finger NMOS device layout and the temperature monitoring circuit for the fully coupled electrothermal simulation.

To accurately study the effect of layout on the NMOS device's protection performance, the cross coupling of the heat sources at each drain finger needs to be modeled. The current summation property at the iETSIM power integrator input models the complete coupling effects between various heat sources. As an example, a two-finger device and the temperature monitoring circuit are shown in Fig. 4.35. The power dissipated in each finger is monitored

by the elements  $Pm1$  and  $Pm2$ . The heat coupling due to the other heat source is obtained using the elements  $R3, R4, R7, R8$ ; they are added to the drain and source temperature integrators associated with  $R1, R2, R5, R6$ . The temperatures  $TDJ1, TDJ2, TSJ1, TSJ2$  (drain junction temperature of device 1, etc.) are then fed back to the electrical models to yield a fully coupled electrothermal simulation.

To concentrate on the heat-coupling effect, we use the same device parameters (e.g., substrate resistance, drain and source resistances, drain and source diffusion area, etc.) for all the NMOS fingers. The drains of the NMOS fingers are tied together to form a single node, as are the NMOS gates and sources. The simulation is performed for grounded-gate NMOS (GGNMOS) devices in a  $0.4\mu\text{m}$  salicided, LDD bulk silicon process. The NMOS snap back model parameters are extracted by an automatic extraction program *BPE1.0* [36], while the drain and source resistances are extracted from the transmission line pulsing (TLP) measurement.

#### 4.2.2 Simulation and experimental results

Simulation results for an eight-finger device under a 0.8A stress current are plotted in Fig. 4.36. Initially all the fingers conduct the stress current equally. However, as the temperature rises, the center fingers become hotter than the outer ones. The non-uniform temperature distribution across the fingers causes a redistribution of the current after approximately 150ns. The thermal generation current in the central fingers increases due to the higher temperature with the central fingers conducting more of the stress current. This causes a further increase in the temperature of the central fingers and eventually leads to thermal failure of the center fingers at around 330ns. As seen from the simulation results, for stresses longer than 150ns, the effective total device width is reduced due to the non-uniform current distribution.

Figure 4.37 compares the current level per finger from 100ns TLP measurements with that from simulations. For this technology, all the fingers turn on without extra ballasting resistances. The thermal model parameters [3] were extracted to fit the single-finger

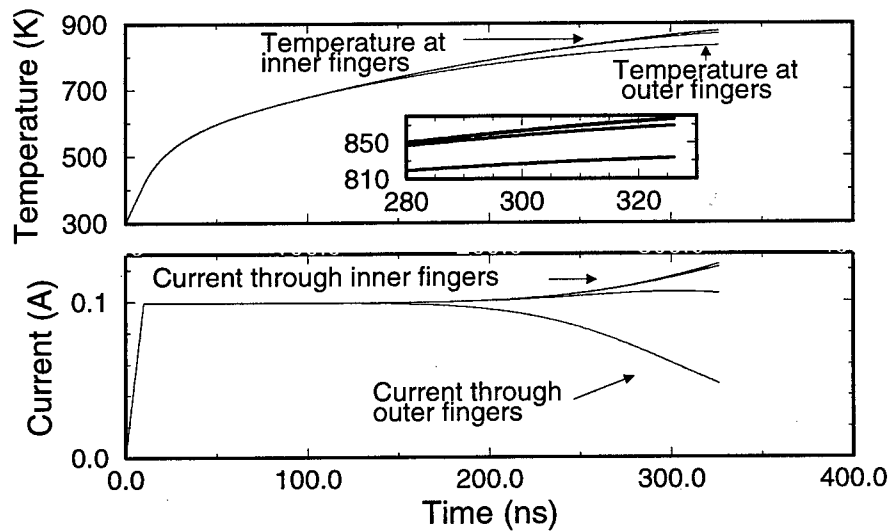


Figure 4.36: Transient temperature and current responses under a 0.8A EOS current input. The source contact-to-gate spacing (SCGS) and drain contact-to-gate spacing (DCGS) are  $1.55\mu\text{m}$ , and contact size (CS) is  $0.5\mu\text{m}$ .

measurement. When the heat coupling effect is not considered, the protection level shown in dashed lines is overestimated. However, simulation results agree well with the measured data when heat coupling is considered. A significant reduction in protection level is observed when the number of fingers is increased from one to four, but not when the number is further increased. This can be explained as follows. For the single-finger device, only self-heating contributes to the thermal failure; but for the two-finger device, each finger has one neighboring heat source. For four fingers and more, each finger has two nearby heat sources except edge fingers. As shown in the figure, simulation predicts that the current protection level is reduced when the heat source separation distance, i.e., CGS, decreases.

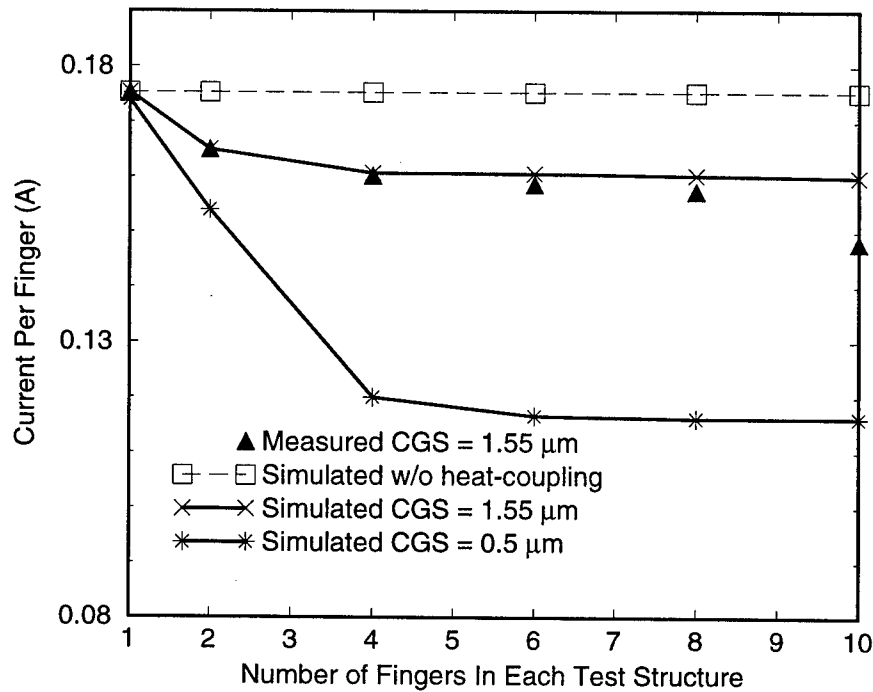


Figure 4.37: Measurement and simulation results under 100ns current pulses for various NMOS fingers. (finger  $W=25\mu\text{m}$ ,  $L=0.4\mu\text{m}$ ,  $\text{SCGS}=\text{DCGS}=\text{CGS}$ ,  $\text{CGS}=0.5\mu\text{m}$  )

### 4.2.3 Discussion

A small discrepancy between measured and simulated data may be for the devices with a large number of fingers seen in Fig. 4.37. Duvvury [37] showed that ground bus resistance can cause current non-uniformities as shown in Fig. 4.38; the finger which is closest to the ground contact tends to fail before others. Also, since the outer contacts of the two edge fingers are not shared by their neighbors, the corresponding finger resistances may be lower. All these non-uniform effects become more important as the finger count increases. With all the above effects modeled accurately, simulations can predict the failure location under a specified stress condition. Simulation results indicate that the protection level and the failure location are primarily affected by electrical non-uniformities for short duration stresses. For

EOS events, failure sites are often observed around the center of the device due to the heat-coupling effect.

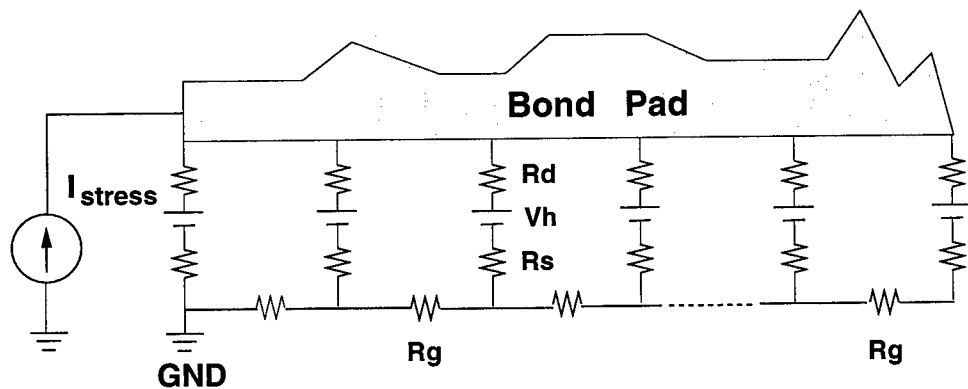


Figure 4.38: The equivalent circuit for the multifinger device.  $R_d$ ,  $R_s$ ,  $R_g$  represent drain, source and ground bus resistances respectively.  $V_h$  represents the NMOS holding voltage.

To improve the protection level, it is desirable to reduce the non-uniform current effects. Increasing the bus metal width will effectively reduce the bus resistance, thus increasing the ratio of finger resistance to bus resistance [37].

#### 4.2.4 Current profile

A current profile is a plot of current-to-failure vs. time-to-failure for a protection circuit or device. It has been shown experimentally to be an effective measure of the circuit EOS/ESD robustness [3]. Figure 4.39 shows the current profiles of two devices with different contact-to-gate spacings. It should be noted that the protection level enhancement obtained by increasing the contact-to-gate spacing varies with the stress time.

The above observation suggests that the layout optimization needs to target a specified stress pulse width. Typically, a designer is provided with a minimum HBM-ESD (human-body model) protection level for an I/O protection circuit. Although pulse current stresses are used in this work, it has been shown that good correlation exists between the HBM-ESD and the current protection level [38]. Thus, HBM-ESD specification can be translated into a current level for a specified pulse width (typically 100ns).

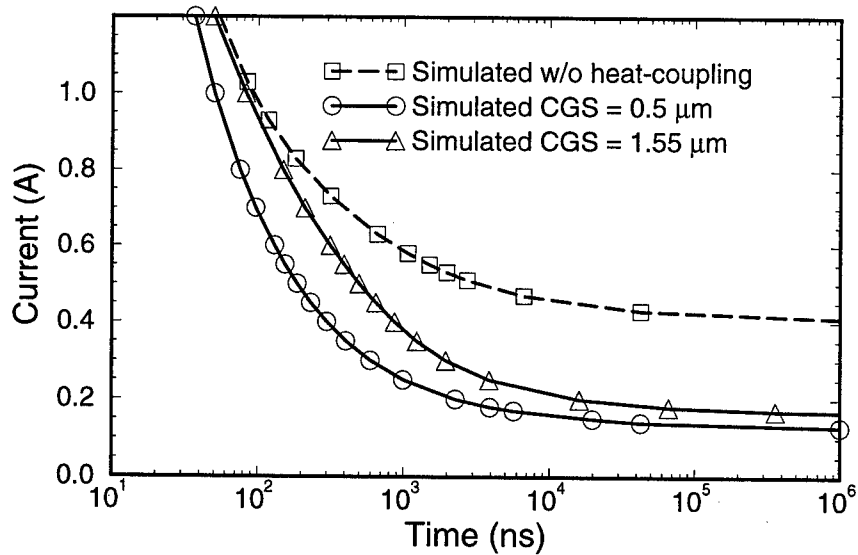


Figure 4.39: Simulated current profiles for two designs with six fingers. (finger  $W=25\mu\text{m}$ ,  $L=0.4\mu\text{m}$ ,  $CS=0.5\mu\text{m}$ )

#### 4.2.5 Layout optimization methodology

Due to the limited number of test structures and large layout design space, the design process can be expedited by using simulation tools. The optimization flow based on simulations is shown in Fig. 4.40. To derive the model parameters, the process technology needs to be characterized. Simulation can be used for the design of test structures. Design specifications are often specified in terms of the silicon area, HBM-ESD level, etc., either as optimization objectives or constraints. For a given process technology with minimum channel length and contact size, design space is explored by varying NMOS finger width, source contact-to-gate spacing (SCGS) and drain contact-to-gate spacing (DCGS). Even with a few design variables, the number of design alternatives can be enormous. It is imperative to reduce the design space rather than enumerating all the possibilities.

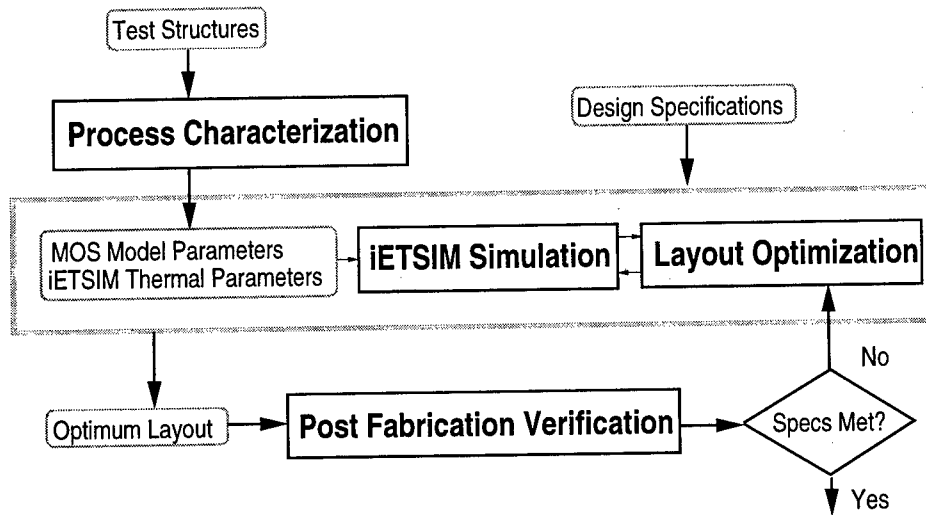


Figure 4.40: Protection Circuit Optimization Flow.

#### 4.2.6 Design example

For high speed applications, it is desirable to minimize the driver output capacitance. As an example, we will find the optimum layout with minimum drain area, so that the output capacitance is minimized. The area budget is  $750\mu m^2$ . The device must pass the 2kV HBM-ESD level, which translates to a 1.33A current level during 100ns pulse testing. The finger width is chosen and fixed at  $25\mu m$  to reduce the design space. We explore the design space by enumerating the contact-to-gate spacing in  $0.2\mu m$  increments. The designs in Fig. 4.41 are the largest possible ones within the area budget for the corresponding DCGS and SCGS. The current levels are obtained from simulations of 100ns pulse stress. There are 11 designs (shaded) which satisfy the HBM requirement; the design which has DCGS= $0.4\mu m$ , and SCGS= $1.2\mu m$  is the best solution among them.

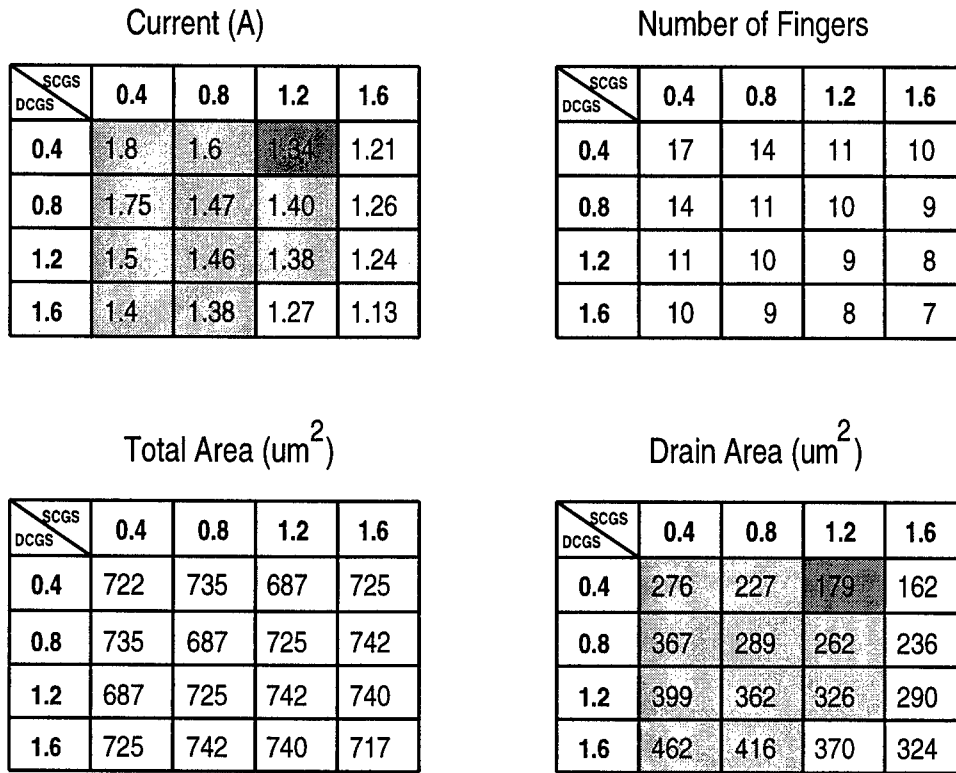


Figure 4.41: The design space with the variations of DCGS and SCGS. (finger  $W=25\mu\text{m}$ ,  $L=0.4\mu\text{m}$ ,  $CS=0.5\mu\text{m}$ )

### 4.3 Simulation of a Two-Stage Protection Circuit

To verify the simulation capability of iETSIM, we performed transmission line pulse (TLP) measurements on a two-stage I/O protection circuit. The simulated behavior of the protection circuit showed good agreement with the measurements.

#### 4.3.1 Circuit description

The two-stage circuit studied in this work consists of an SCR as the primary protection device, gate-grounded NMOS (GGNMOS) transistors as the secondary protection devices, and silicided n-diffusions as series resistors. The circuit schematic is shown in Fig. 4.42. The layout floorplan of the circuit is illustrated in Fig. 4.43 with the critical dimensions indicated. The SCR is an effective ESD protection device due to its low holding voltage

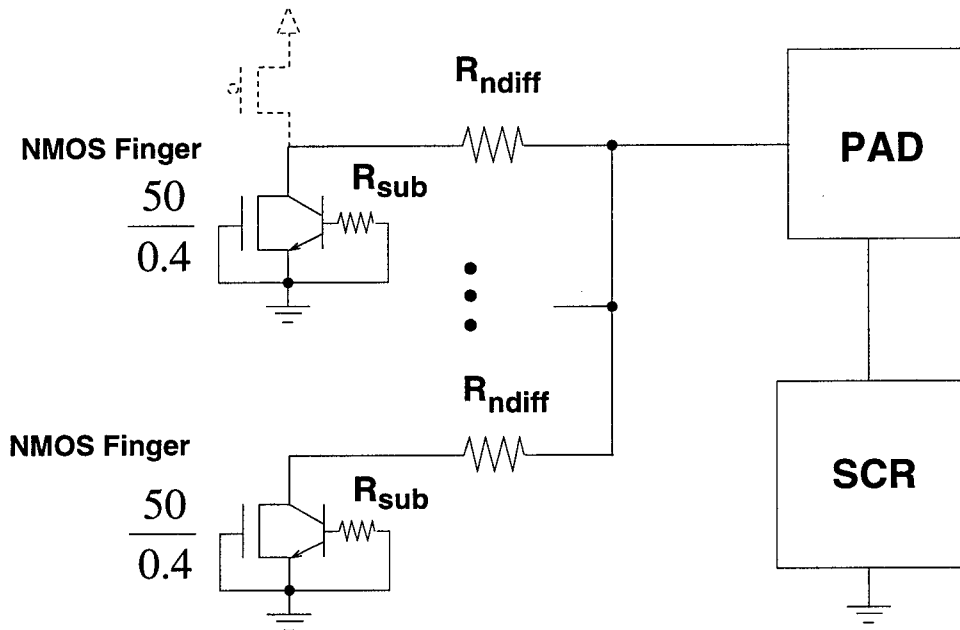


Figure 4.42: Circuit schematic of the two-stage protection circuit with various numbers of NMOS fingers. An NMOS leg is defined as one NMOS transistor, and an NMOS finger is composed of two drain-connected NMOS legs which share one resistor.

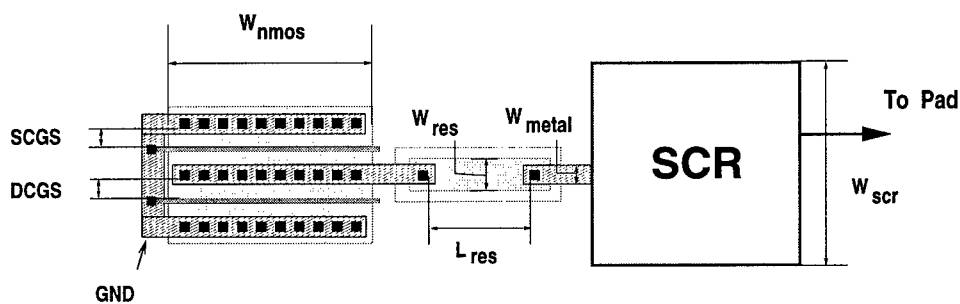


Figure 4.43: Floorplan of the two-stage protection circuit with one NMOS finger. The figure is not drawn to scale.  $L_{res}$  is the distance measured between the closest contact hole edges of the two terminals.

[39][40][41][42]. However, the triggering voltage of an SCR is usually higher than that of a gate-grounded NMOS transistor. We add one silicided n-diffusion resistor for each NMOS finger to raise the voltage at the SCR anode under ESD events, so that the triggering of the SCR is guaranteed. Furthermore, the extra drain ballasting resistance provided by the resistors ensures the uniform turn-on of the NMOS fingers. In addition, the current flowing through the NMOS transistors is limited due to current saturation in the diffusion resistors under high current levels [43].

The test structures were fabricated in a 0.4  $\mu\text{m}$  silicided LDD bulk silicon process without using a silicide block. The silicided n-diffusion is made in the n-well. In the I/O cells, the gate of a protection NMOS transistor is either connected to the pre-driver output when belonging to the driver, or tied to ground when added in parallel with the driver NMOS transistors for improving the ESD protection level. In the test structures, the gates of all NMOS legs are grounded as shown in Fig. 4.42. We will later justify using test structures to predict the circuit protection level in the I/O cells.

### 4.3.2 Device characterization

The transmission line pulsing (TLP) technique was employed to characterize the devices under high currents [44]. Fig. 4.47 illustrates the setup of the TLP system. The system provides a square current pulse for wafer-level measurement of the I-V curves. The variation of the pulse amplitude  $I$  is produced using different supply voltages. The length of the transmission line determines the pulse width  $t$  (approximately 3 ns/foot).

The current-at-failure for a pulse width of about 100 ns correlates with the HBM-ESD level [38][45]; therefore, a fixed pulse width of 100 ns was used in this study. Given the required HBM-ESD level such as 1.5 kV, the equivalent current protection level is calculated as  $1.5 \text{ kV} / 1.5 \text{ k}\Omega = 1 \text{ A}$ , where 1.5 k $\Omega$  is the series resistance in the Human Body Model circuit.

NMOS devices were stressed using the TLP system to determine the current at second breakdown,  $I_{t2}$ . The BSIM3 MOS model parameters were extracted using the extraction

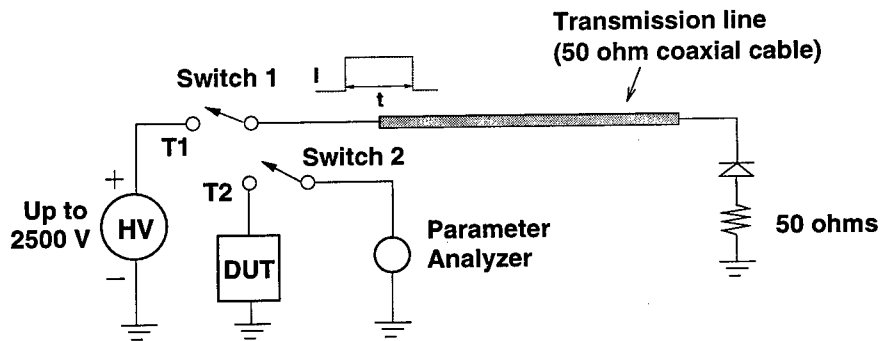


Figure 4.44: Transmission line pulsing measurement setup. HV is the high voltage supply, and DUT stands for the Device Under Test.

program BSIMPro [46], and the snapback model parameters were extracted using the extraction program BPE1.0 [32][36]. The high current I-V characteristic of a five finger NMOS device subject to 100 ns pulses is shown in Fig. 4.48. For this technology, all the fingers turn on without extra ballasting resistance. The device is under short circuit failure when the second breakdown occurs. The device  $I_{t2}$  scales well with the number of NMOS fingers. Each finger can sustain approximately 0.3 A current (finger  $I_{t2}$ ). The avalanche breakdown voltage for the gate-grounded NMOS device is around 9.6 V.

Measured and simulated I-V characteristics of a 50  $\mu\text{m}$  wide SCR device are given in Fig. 4.49. The device can sustain current above 3.1 A without damage; the I-V curve was not measured beyond 3.1 A as it has passed the normally required current level. The holding voltage is modeled well by the circuit in Fig. 4.37. Note that the SCR voltage is almost linearly proportional to its current in the high current regime; a constant resistance  $R_{on}$  is therefore used to fit the slope of the measured I-V curve. The simulated I-V curve matches well with the measurements for currents below 2.5 A. Beyond 2.5 A, the deviation of the measurement from the simulation, which is attributed to the heating effect, can introduce simulation errors. The exact triggering voltage of the SCR can not be determined using the TLP setup because it is dependent on the slope of the input signal [41]. This is confirmed by the observation that the triggering voltage of the SCR in the two-stage circuit ranged from 8.5 to 11.5 V depending on the number of fingers; the results suggest that the rising slope of the stress signal following the avalanche breakdown of NMOS fingers depends on

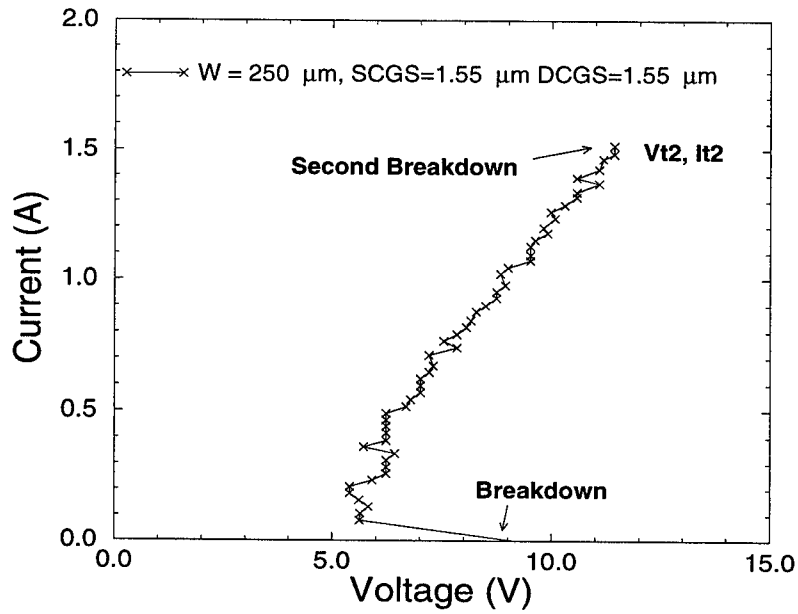


Figure 4.45: The high current I-V curve for a GGNMOS device with five fingers.

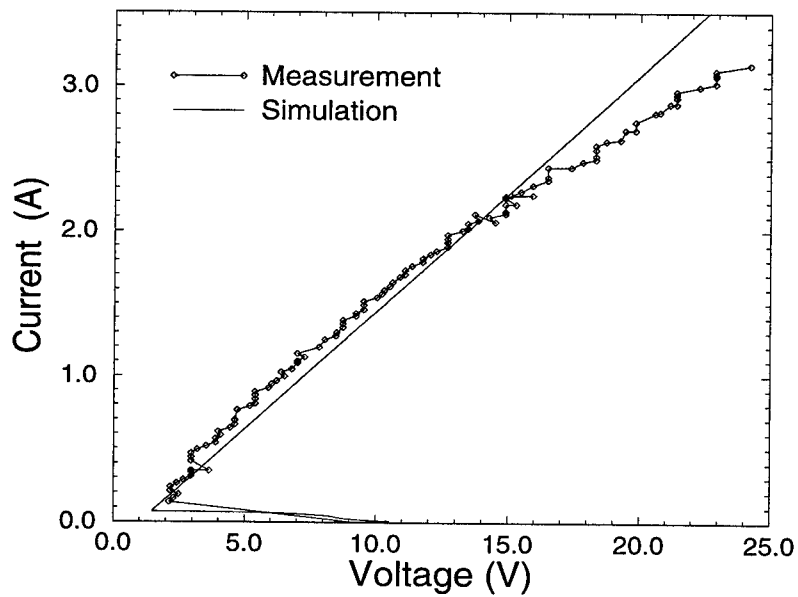


Figure 4.46: The high current I-V curve for an SCR device with 50  $\mu\text{m}$  width.

the number of NMOS fingers.

Model parameters for the resistor used in the protection circuit must also be extracted. Silicided n-diffusion resistors were characterized as shown in Figs. 4.50 and 4.51. Comparison

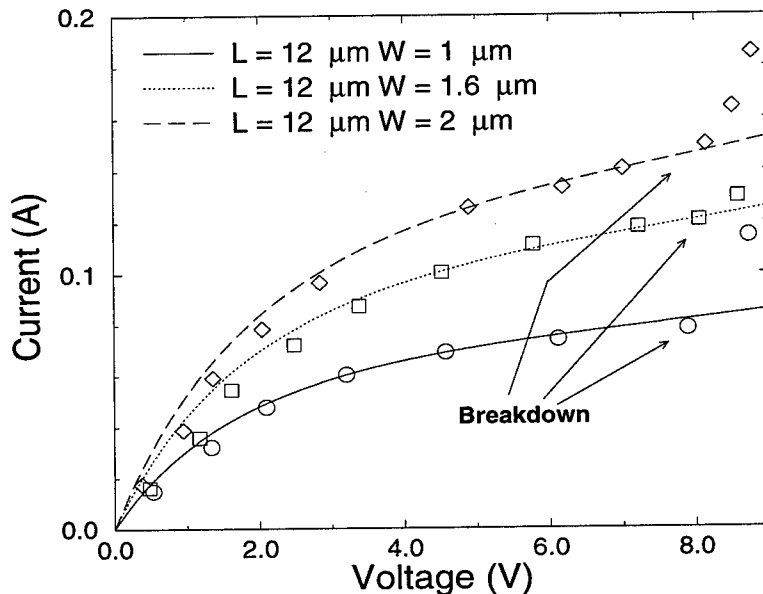


Figure 4.47: Measured and simulated I-V characteristics of silicided diffusion resistors with various widths (symbols denote the measurements).

of the measured and simulated results in the figures indicates that an excellent match can be achieved with the model given by Eq. (4.35).

The model parameters include the doping concentration, electron mobility, and the physical dimensions of the resistor. In addition, an offset  $W_D$  is introduced to account for lateral diffusion;  $W$  in Eq. (4.35) is replaced by  $W_D + W_{res}$  where  $W_{res}$  is the physical layout width of the resistor. The extraction of  $W_D$  from measurements is illustrated in Fig. 4.52. The saturation current for resistors with fixed length and variable width is plotted as shown. The line which fits the measured data is extrapolated to intersect the x-axis. The offset from the origin determines the value for  $W_D$ .

Although the diffusion resistor model in iETSIM can capture breakdown [47], we choose

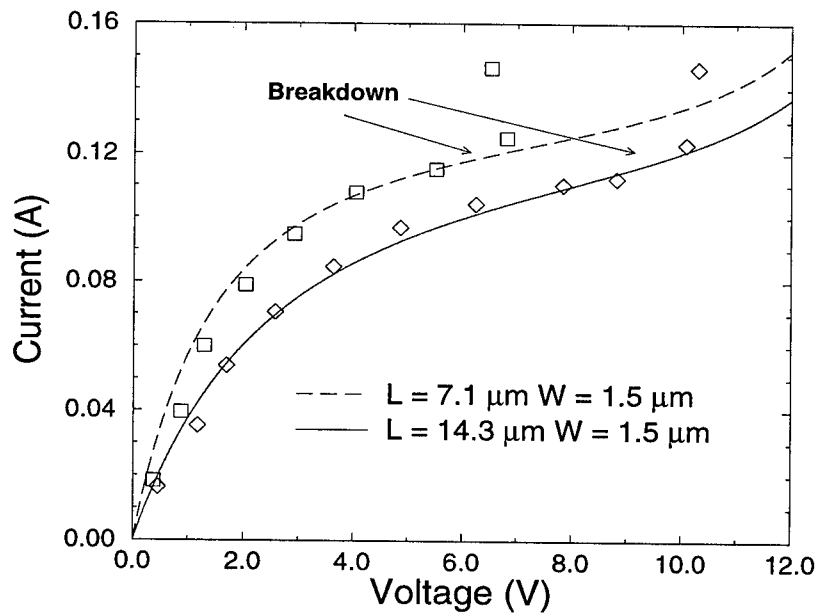


Figure 4.48: Measured and simulated I-V characteristics of silicided diffusion resistors with various lengths (symbols denote the measurements).

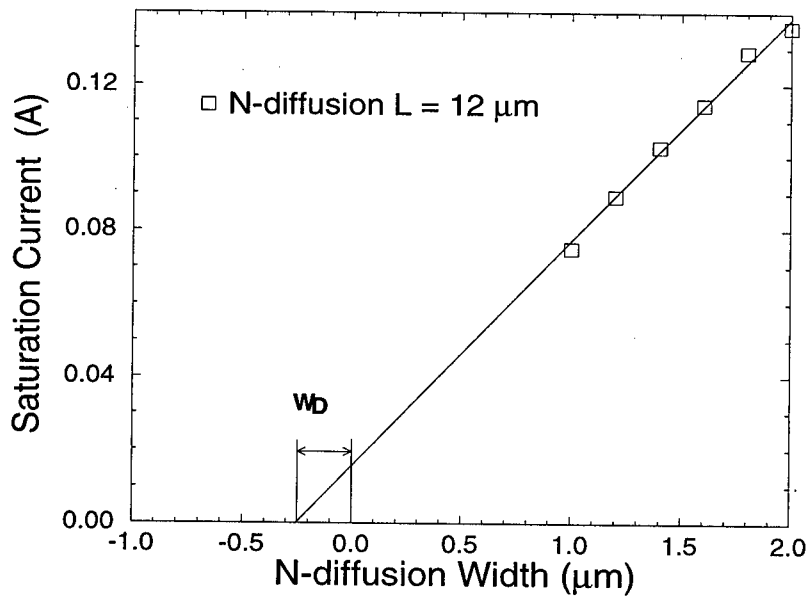


Figure 4.49: The saturation current (symbols) for silicided n-diffusions with various widths.  $W_D$  is extracted as  $0.25 \mu\text{m}$ .

not to use this feature to achieve a better fit in the saturation region. Therefore, the circuit failure due to the resistor breakdown must be monitored explicitly. The resistor breakdown voltage is extracted from measurements. As shown in Fig. 4.51, the breakdown voltage of a resistor is dependent on its physical length. It has been shown that both avalanche and thermal generation of minority carriers contribute to breakdown [48]. For a fixed voltage, the electrical field and the number of avalanche generated carriers decrease with increasing resistor length.

### 4.3.3 Simulation and experiment

We simulated the protection circuit shown in Fig. 4.46. The length and width of each resistor are  $7.1 \mu\text{m}$  and  $1.5 \mu\text{m}$ , respectively, providing approximately  $10 \Omega$  resistance per finger. The width of the NMOS finger is  $50 \mu\text{m}$ , and the SCR width is  $50 \mu\text{m}$ . To simplify the simulation deck, the netlist for multiple fingers and resistors are transformed into an equivalent circuit with a single finger and resistor. The transient responses of a five-finger circuit subjected to a 1.5 A current pulse are plotted in Fig. 4.53. The GGNMOS turns on first due to its lower avalanche breakdown voltage.

Next, the pad voltage ramps up and triggers the SCR. Current saturation in the diffusion resistor limits the stress current through the GGNMOS device so that most of the current flows through the SCR.

The saturation current of the diffusion resistor is around 0.12 A as shown in Fig. 4.51, which is less than the NMOS finger  $I_{t2}$ , 0.3 A; this indicates that the protection circuit will fail after the n-diffusion breaks down. With this observation in mind, we monitor the voltage drop across the resistor (indicated in Fig. 4.53) when we increase the current stress level. Figure 4.54 plots the simulated voltage drop across the resistor versus the stress current level for the protection circuit with four NMOS fingers. As the breakdown voltage of the resistor is approximately 6.5 V which is estimated from Fig. 4.51, the current protection level predicted by iETSIM is 2.58 A as illustrated in Fig. 4.54.

Figure 4.55 shows the high current measurement for the circuit with four fingers. As

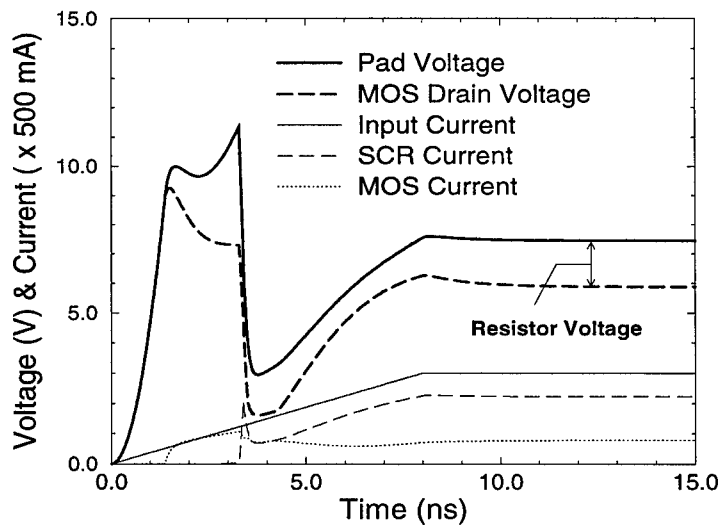


Figure 4.50: Simulated transient response of a two-stage circuit to a 1.5 A current stress input. The voltage across the diffusion resistor is the voltage difference between the pad and the NMOS drain.

seen in the figure, the I-V curve switches to the SCR curve at the 2.55 A current level; this suggests that the other current path has open circuited. Indeed, failure analysis indicates that the metal lines connected with the diffusion resistors experience open failures. Circuit open failure is not electrically detectable by monitoring the leakage current; therefore, the TLP measurement continued until the supply voltage limit (2 kV) was reached. Under the high current stress, the resistors break down first, then the metal lines connected to resistors blow open by the high current flowing through the individual NMOS fingers. Recall that simulation predicted a failure current of 2.58 A; this is in excellent agreement with the measured value.

Figure 4.56 compares the measured and simulated protection level for the protection circuit with various numbers of NMOS fingers. The simulation results agree well with the measurements, with less than 5% error. Note that the current level is linearly proportional to the number of NMOS fingers, and is increased by the value of the resistor saturation current with the addition of every finger. This indicates that the current level enhancement

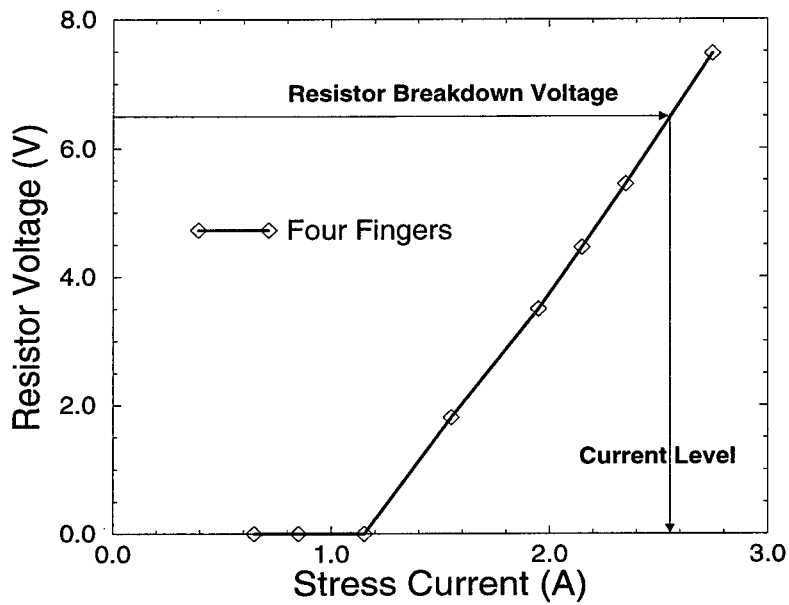


Figure 4.51: Simulated voltage drop across the resistor versus the stress current level for the protection circuit with four NMOS fingers.

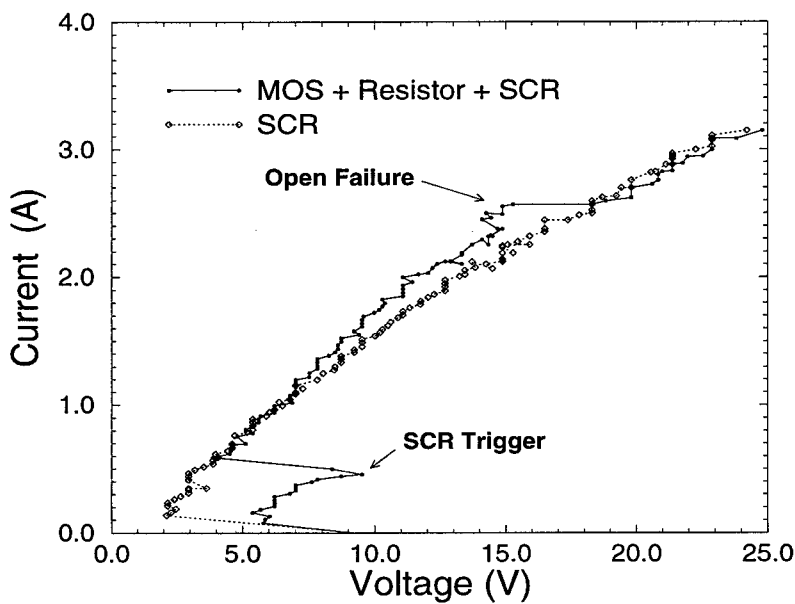


Figure 4.52: Measurement of the protection circuit in Fig. 4.41 which contains four NMOS fingers.

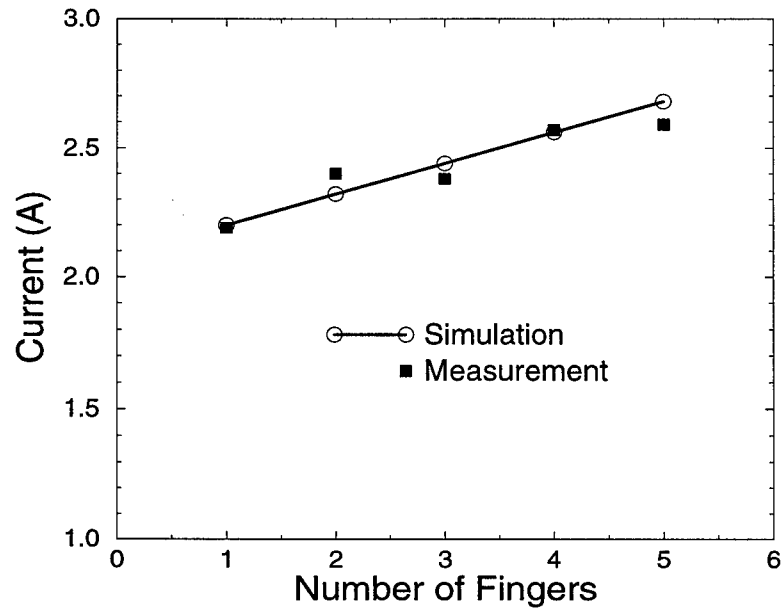


Figure 4.53: Measured and simulated results of ESD current level for the protection circuit with various numbers of NMOS fingers.

gained by increasing the finger numbers is limited by the resistor saturation current when the NMOS finger  $I_{t2}$  is larger than the resistor saturation current.

The key difference between test structures and I/O circuits is the NMOS gate voltage  $V_g$ ; in I/O circuits,  $V_g$  may not be 0 for all fingers. The NMOS transistor with its gate biased can trigger at a lower voltage than a gate-grounded NMOS transistor. Note that the pad voltage at failure is around 15 V (see Fig. 4.55), which is higher than the avalanche breakdown voltage of a gate-grounded NMOS transistor; this indicates that all the fingers will be turned on at the steady state. Therefore, the GGNMOS test structures can be used to predict the current protection level for the I/O circuits.

#### 4.3.4 Failure analysis

In designs with the metal line width ( $W_{metal}$  in Fig. 4.46) increased to avoid open failures, we instead observed failures in the GGNMOS device. The most commonly observed failure site for this type of I/O protection circuit is in one of the central NMOS fingers, as opposed

to the edge fingers reported by other researchers [37].

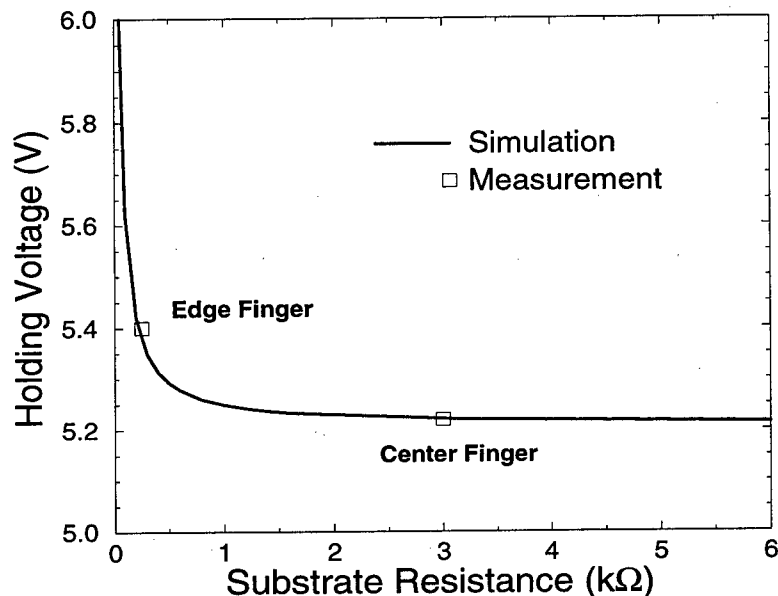


Figure 4.54: Simulated and measured holding voltage versus the substrate resistance for a 20  $\mu\text{m}$  wide gate-grounded NMOS transistor.

This observation can be explained by the relationship between the NMOS holding voltage and its substrate resistance. The simulation results in Fig. 4.57 indicate that the NMOS holding voltage decreases with the substrate resistance [49]. Each finger in a multifinger structure sees a different substrate resistance; therefore, the holding voltage of center and edge fingers for a multifinger NMOS device may differ. We measured the holding voltages on an eight leg NMOS test structure, where the NMOS drains are not connected so they can be probed separately. A 0.18 V holding voltage difference from center to edge was observed (center finger has the lower holding voltage). As a result, the NMOS transistor with the highest substrate resistance will fail because the diffusion resistor for this finger sustains the highest voltage between the pad and the NMOS drain, and will break down before the others. Usually the center finger has higher substrate resistance, but this is dependent on the placement of substrate contacts.

# 5 HOT CARRIER INDUCED DEGRADATION OF DEEP-SUBMICRON PMOSFETS

## 5.1 Introduction

Previous studies of hot carrier induced degradation in PMOSFETs have concluded that the worst case degradation occurs when the device is biased at the gate current ( $I_g$ ) peak and is due to electron trapping in the gate oxide [58][59][60][61][62]. However, some recent studies of deep-submicron ( $\sim 0.25\mu\text{m}$ ) PMOSFETs suggest that, in advanced technologies, peak  $I_g$  is no longer the worst case stress condition [63][64]. If this assertion is correct, then the procedure for accelerated lifetime testing should be modified as should the lifetime model implemented in circuit reliability simulators such as ILLIADS-R. In this chapter, we present an experimental investigation of hot carrier induced degradation in deep-submicron PMOSFETs.

## 5.2 Experiment

The devices used in this work are surface channel PMOSFETs from a  $0.35\mu\text{m}$  twin-well CMOS technology with oxide thickness of 7nm. Hot carrier lifetime was defined as the time required for a 10% change in the saturation drain current, i.e.,  $|\frac{\Delta I_{dsat}}{I_{dsat}}| = 0.1$ .  $I_{dsat}$  was measured at the specified operating voltage, 2.7V.

Charge pumping characteristics were measured to provide additional information about the degradation mechanism [65]. Charge pumping measurements were performed using a square voltage waveform at the gate with a fixed amplitude of 2V and a frequency of 1MHz. The signal rise and fall times were 50ns. The source and drain were held at 0V. The peak charge pumping current is proportional to the number of interface states; any shift in the gate-pulse voltage at this maximum is indicative of charge trapping in the oxide [66].

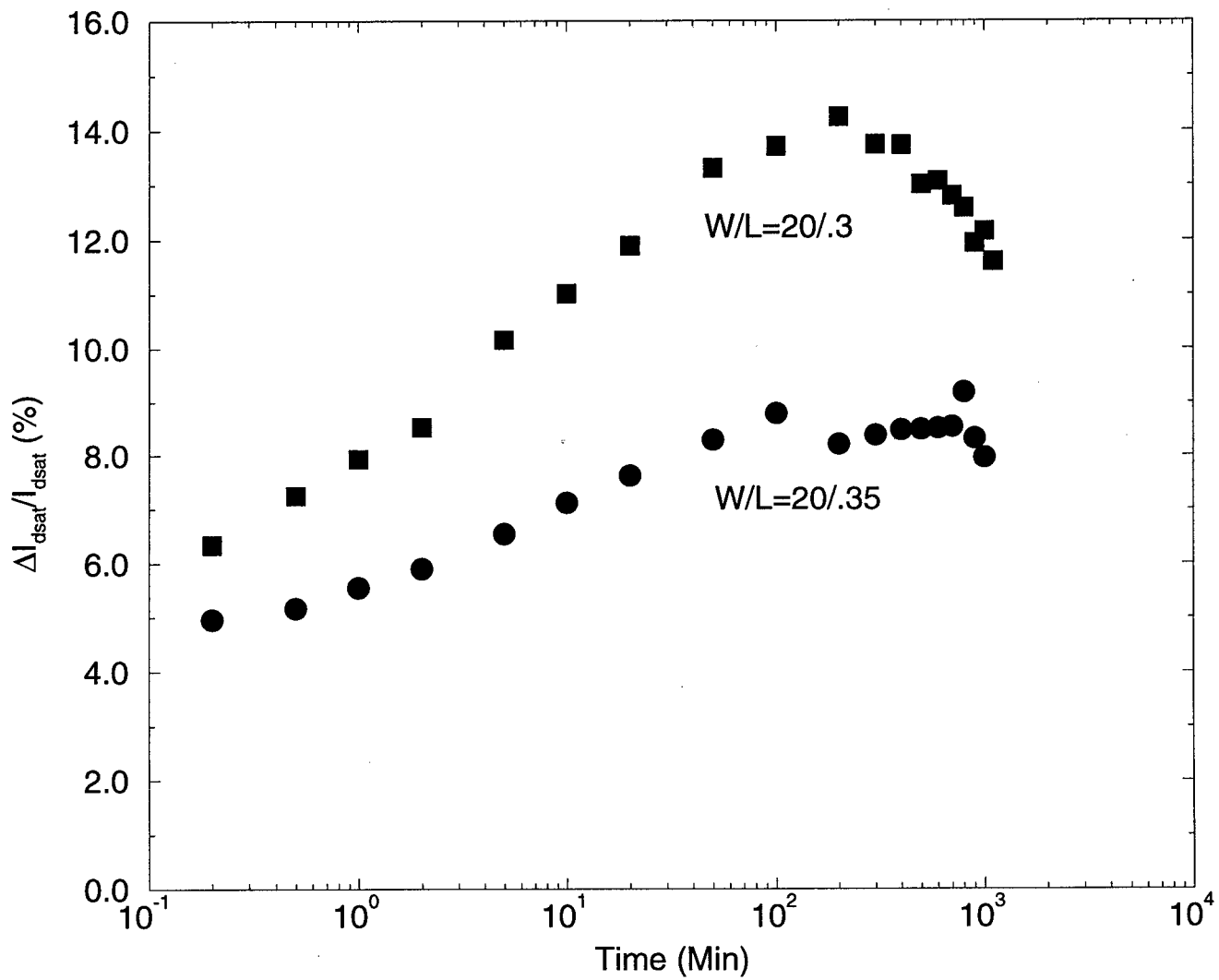


Figure 5.55: Degradation in devices biased at the maximum gate current condition.

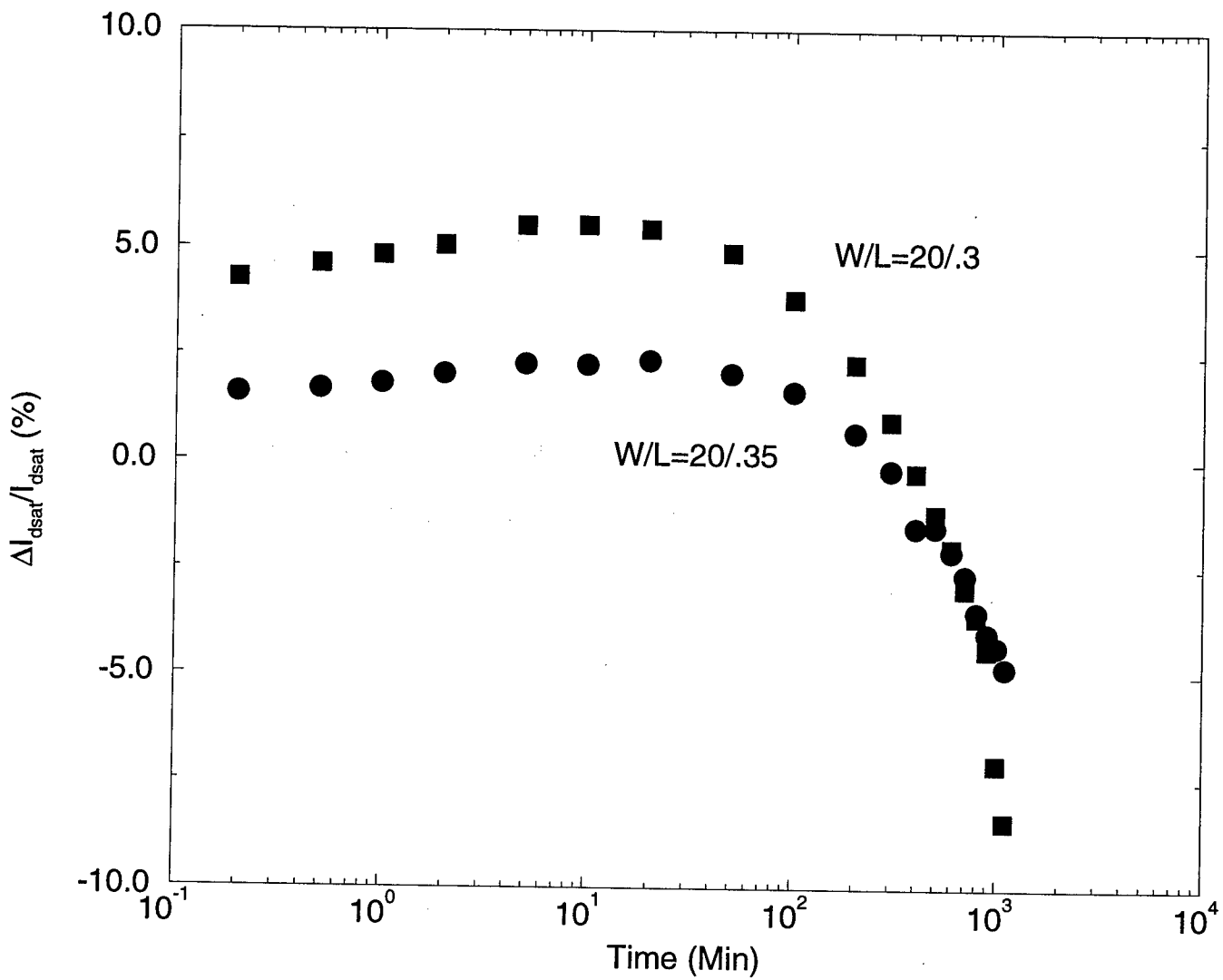


Figure 5.56: Degradation in devices biased at the maximum substrate current condition.

### 5.3 Results

Figures 5.55 and 5.56 show the drain current degradation as a function of stress time for devices biased at peak  $I_g$  and peak  $I_{sub}$ , respectively. The drain voltage ( $V_{ds}$ ) was set to -4.6V during the stress. Under both stress conditions, a “turn-around” is observed; that is, first the drain current increases (indicative of electron trapping in the oxide) and later it decreases. The turn-around occurs sooner in the devices biased at peak  $I_{sub}$ .

Consider the results for the  $0.3\mu\text{m}$  devices. The device stressed at peak  $I_g$  reaches the end of lifetime first ( $|\frac{\Delta I_{dsat}}{I_{dsat}}| = 0.1$  at  $T_{stress} = 50$  minutes). However, at this point, the drain current has *increased* by 10%. The device which is stressed at peak  $I_{sub}$  shows a reduction in its drain current much sooner than the device which is biased at peak  $I_g$  does and this change in drain current is far more likely to affect circuit performance. Thus, these results strongly suggest that the definition of transistor lifetime should be changed. Note that we also monitored threshold voltage shifts and linear drain current shifts and these showed the same turn-around phenomenon. Thus, if any of these quantities are used as the lifetime monitor, the same conclusion will be reached.

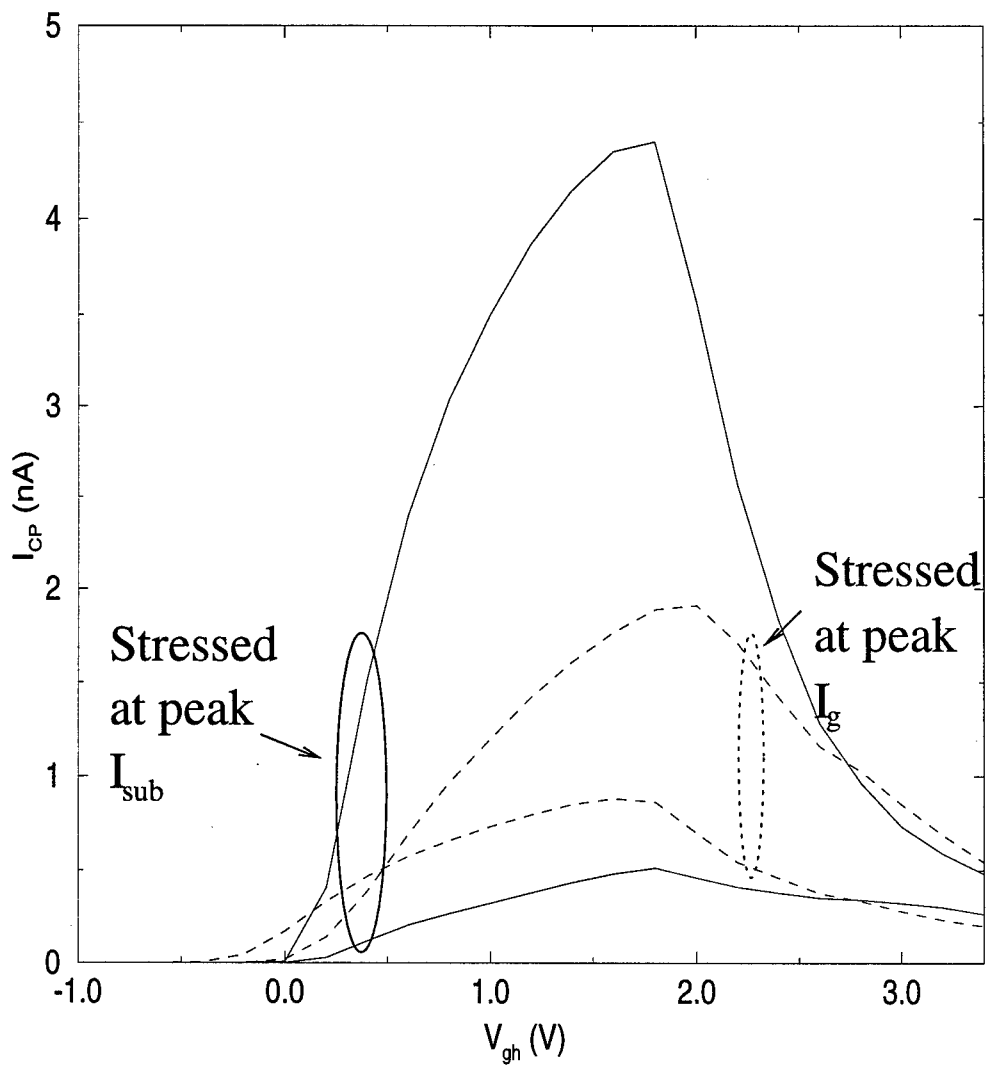


Figure 5.57: Charge pumping current characteristics before and after stressing for 1000 minutes.  $W/L=20/0.35$ .  $V_{gh}$  denotes the upper level of the gate voltage waveform; magnitude is fixed at 2V.

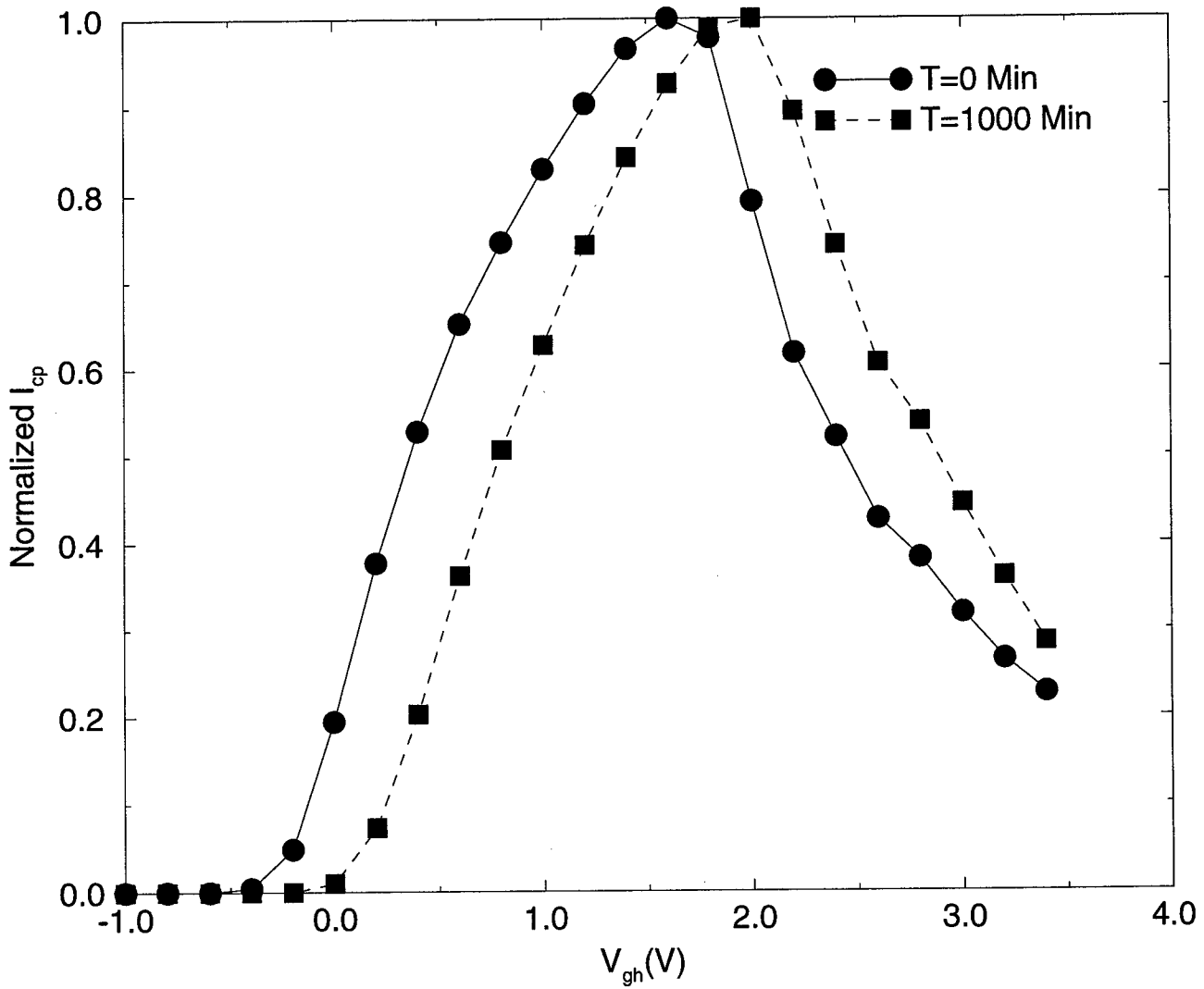


Figure 5.58: Normalized charge pumping current for the device stressed at peak  $I_g$ .

Figure 5.57 shows charge pumping characteristics before and after stressing at peak  $I_g$  and peak  $I_{sub}$ . Clearly, more interface states are generated in the device which is stressed at peak  $I_{sub}$ . This presumably accounts for the drain current reduction which follows a relatively short stress (see Fig. 5.56). The turn-around effect is attributed to competing mechanisms: drain current enhancement due to trapped electrons and drain current reduction due to interface states. That electron trapping occurs may be seen from the normalized charge pumping characteristics (Fig. 5.58). The shift of the curve to higher gate voltages indicates electron trapping in the oxide.

## 5.4 Conclusions

Interface state generation is at least as important as charge trapping in the gate oxide as a cause of hot carrier induced degradation in deep-submicron PMOSFETS. Degradation models on circuit reliability simulators should be updated and simulation studies performed to explore the impact of the turn-around effect on circuit reliability. Following these works, new and practical procedures for accelerated lifetime testing can be proposed.

At present, we are extracting degradation parameters for PMOSFETS in a  $0.25\mu\text{m}$  CMOS technology. We shall perform the simulation studies and determine an appropriate criterion for device lifetime.

## References

- [1] K. Fukahori and P. R. Gray, "Computer simulation of integrated circuits in the presence of electrothermal interaction," *IEEE Journal of Solid-State Circuits*, vol. 11, pp. 834–846, December 1976.
- [2] S. S. Lee and D. J. Allstot, "Electrothermal simulation of integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 28, pp. 1283–1293, December 1993.
- [3] C. H. Diaz, S. M. Kang, and C. Duvvury, "Circuit-level electrothermal simulation of electrical overstress failures in advanced MOS I/O protection devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, pp. 482–493, April 1994.
- [4] C. P. Wan and B. J. Sheu, "Temperature dependence modeling for MOS VLSI circuit simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 8, pp. 1065–1073, October 1989.
- [5] A. M. Hill, *Switching Density Analysis for Power and Reliability in VLSI Circuits*. Ph.D. thesis, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1996.
- [6] A. Dharchoudhury, S. M. Kang, K. H. Kim, and S. H. Lee, "Fast and accurate timing simulation with regionwise quadratic models of MOS I-V characteristics," in *Proceedings of the ACM/IEEE International Conference on Computer-Aided Design*, San Jose, CA, November 1994, pp. 190-194.
- [7] M. S. Lin, "A better understanding of the channel mobility of Si MOSFET's based on the physics of quantized subbands," *IEEE Transactions on Electron Devices*, vol. 35, pp. 2406–2411, 1988.

- [8] M. S. Liang, J. Y. Choi, P. K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFET's," *IEEE Transactions on Electron Devices*, vol. 33, pp. 409–412, 1986.
- [9] K. Hess, *Advanced Theory of Semiconductor Devices*. Prentice-Hall, 1988.
- [10] C. T. Sah, T. H. Ning, and L. L. Tschopp, "The scattering of electrons by surface oxide charges and by the lattice vibrations at the Si-SiO<sub>2</sub> interface," *Surface Science*, vol. 32, pp. 561–575, 1972.
- [11] H. Ezawa, S. Kawaji, and K. Nakamura, "Surfons and the electron mobility in silicon inversion layers," *Japanese Journal of Applied Physics*, vol. 13, pp. 126–155, September 1974.
- [12] A. Hartstein, T. H. Ning, and A. B. Fowler, "Electron scattering in silicon inversion layers by oxide and surface roughness," *Surface Science*, vol. 58, pp. 181–190, 1976.
- [13] D. W. Marquart, *Journal of the Society for Industrial and Applied Math.*, vol. 11, pp. 431–441, 1963.
- [14] M. N. Ozisik, *Boundary Value Problems of Heat Conduction*. New York, NY: Dover, 1968.
- [15] V. Koval, I. W. Farmaga, A. J. Strojwas, and S. W. Director, "MONSTR: A complete thermal simulator of electronic systems," in *Proceedings of the ACM/IEEE Design Automation Conference*, June 1994, pp. 570–575.
- [16] J. F. Thompson, Z. Warsi, and C. W. Mastin, *Numerical Grid Generation*. New York, NY: North-Holland, 1985.
- [17] Y. C. Lee, H. T. Ghaffari, and J. M. Segelken, "Internal thermal resistance of a multi-chip packaging design for VLSI based systems," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 12, pp. 163–169, June 1989.

- [18] J. R. Black, "Electromigration failure modes in aluminum metallization for semiconductor devices," in *Proceedings of the IEEE*, vol. 57, September 1969, pp. 1587-1594.
- [19] C. C. Teng, Y. K. Cheng, E. Rosenbaum, and S. M. Kang, "iTEM: A new electromigration (EM) reliability diagnosis tool using electrothermal timing simulation," in *Proceedings of the IEEE International Reliability Physics Symposium*, April 1996, pp. 172-179.
- [20] C. C. Teng, *Hierarchical Electromigration Reliability Diagnosis for ULSI Interconnects*. Ph.D. thesis, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1996.
- [21] M. G. Xakellis and F. N. Najm, "Statistical estimation of the switching activity in digital circuits," in *Proceedings of the ACM/IEEE Design Automation Conference*, June 1994, pp. 728-733.
- [22] I. Miller and J. E. Freund, *Probability and Statistics for Engineers*. Englewood Cliffs, NJ: 3rd ed. Prentice-Hall, 1985.
- [23] A. Dharchoudhury, *Advanced Techniques for Fast Timing Simulation of MOS VLSI Circuits*. Ph.D. thesis, Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1995.
- [24] S. Chowdhury and J. S. Barkatullah, "Estimation of maximum currents in MOS IC logic circuits," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 9, pp. 642-654, June 1990.
- [25] H. Kriplani, F. Najm, and I. N. Hajj, "Improved delay and current models for estimating maximum currents in CMOS VLSI circuits," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, May 1994, pp. 433-436.
- [26] H. Kriplani, F. Najm, and I. N. Hajj, "Pattern-independent maximum current estimation in power and ground buses of CMOS VLSI circuits: algorithms, signal correlations,

- and their resolution,” in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, pp. 998–1012, August 1995.
- [27] R. Burch, F. N. Najm, P. Yang, and T. Trick, “A Monte Carlo approach for power estimation,” *IEEE Transactions on VLSI Systems*, vol. 1, no. 1, pp. 63–71, 1993.
- [28] L.-P. Yuan, C.-C. Teng, and S.-M. Kang, “Statistical estimation of average power dissipation in CMOS VLSI circuits using nonparametric techniques,” *IEEE/ACM International Symposium on Low Power Electronics and Design*, Monterey, CA, 1996, pp. 73–78.
- [29] C.-Y. Tsui, M. Pedram, and A. M. Despain, “Exact and approximate methods for calculating signal and transition probabilities in FSMs” *31st ACM/IEEE Design Automation Conference*, San Diego, CA, 1994, pp. 18–23.
- [30] T.-L. Chou and K. Roy, “Statistical estimation of sequential circuit activity,” *IEEE/ACM International Conference Computer-Aided Design*, 1995, pp. 34–37.
- [31] A. Amerasekera, S. Ramaswamy, M. Chang, and C. Duvvury, “Modeling MOS Snap-back and Parasitic Bipolar Action for Circuit-Level ESD and High Current Simulations,” *IRPS*, pp. 318–326, 1996.
- [32] S. Ramaswamy, E. Li, E. Rosenbaum, and S. M. Kang, “Circuit-level simulation of CDM-ESD and EOS in submicron MOS devices,” *EOS/ESD Symposium*, Orlando, FL, 1996, pp. 316–321.
- [33] J. Huang, Z. Liu, M. Jeng, K. Hui, M. Chan, P. Ko, and C. Hu, “BSIM3 Manual (version 2),” University of California, Berkeley, 1994.
- [34] V. Dwyer, A. Franklin, and D. Campbell, “Thermal Failure in Semiconductor Devices,” *Solid State Electronics*, vol. 33, no. 5, pp. 553–560, 1990.

- [35] S. Ramaswamy, *Modeling, Simulation and Design Guidelines For EOS/ESD Protection Circuits in CMOS Technologies*, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 1996.
- [36] E. Li and E. Rosenbaum, *BPE Version 1.0 User's Manual*, University of Illinois at Urbana-Champaign, 1996.
- [37] C. Duvvury and R. Rountree, "Output ESD protection techniques for advanced CMOS processes," *EOS/ESD Symposium*, 1988, pp. 206-211.
- [38] T. Polgreen and A. Chatterjee, "Improving the ESD failure threshold of silicided nMOS output transistors by ensuring uniform current flow," *EOS/ESD Symposium*, 1989, pp. 167-174.
- [39] L. Avery, "Using SCRs as transient protection structures in integrated circuits," *EOS/ESD Symposium Proceedings*, 1983, pp. 177-180.
- [40] A. Chatterjee and T. Polgreen, "A low-voltage triggering SCR for on-chip ESD protection at output and input pads," *IEEE Electron Device Letters.*, vol. 12, no. 1, pp. 21-22, 1991.
- [41] M. Ker, C. Wu, and C. Lee, "A novel CMOS ESD/EOS protection circuit with full-SCR structures," *EOS/ESD Symposium Proceedings*, 1992, pp. 258-264.
- [42] B. G. Carbajal III, R. A. Cline, and B. H. Andresen, "A successful HBM ESD protection circuit for micron and sub-micron level CMOS," *EOS/ESD Symposium Proceedings*, 1992, pp 234-242.
- [43] G. Krieger and P. Niles, "Diffused resistors characteristics at high current density levels - analysis and applications," *IEEE Transactions on Electron Devices*, vol. 36, pp. 416-423, February, 1989.
- [44] T. Maloney and N. Khurana, "Transmission line pulsing techniques for circuit modeling of ESD phenomena," *EOS/ESD Symposium Proceedings*, 1985, pp. 49-54.

- [45] A. Amerasekera, L. Roozendaal, J. Bruines, and F. Kuper, "Characterization and modeling of second breakdown in NMOST's for the extraction of ESD-related process and design parameters," *IEEE Transactions on Electronic Devices*, vol. 38, no. 9, pp. 2161–2168, 1991.
- [46] Berkeley Technology Associates, *BSIMPro for Windows*, BTA Technology Inc., 1993.
- [47] S. Ramaswamy and S. M. Kang, *iETSIM Version 2.0 User's Manual*, University of Illinois at Urbana-Champaign, 1996.
- [48] A. Amerasekera, M. Chang, J. A. Seitchik, A. Chatterjee, K. Mayaram, and J. Chern, "Self-heating effects in basic semiconductor structure," *IEEE Transactions on Electron Devices*, vol. 40, pp. 1836–1844, October, 1993.
- [49] D. Suh and J. G. Fossum, "The effect of body resistance on the breakdown characteristics of SOI MOSFET's," *IEEE Transactions on Electronic Devices*, vol. 41, no. 6, pp. 1063–1066, 1994.
- [50] "Apache Run-time Configuration Directives," in *Apache User's Guide*, <http://www.apache.org/docs/>, 1994.
- [51] L. Wall, T. Christiansen, and R. L. Schwartz, *Programming Perl*, 2nd Ed. Sebastopol, CA: O'Reilly & Associates, Inc., 1996.
- [52] J. Deep and P. Holfelder, *Developing CGI Applications with Perl*. New York, NY: Wiley Computer Publishing, 1996.
- [53] C.-C. Teng and S.-M. Kang, *iRULE Version 1.0 User's Manual*, Coordinated Science Laboratory, University of Illinois, Urbana, IL, 1996.
- [54] C.-C. Teng, E. Rosenbaum and S.-M. Kang, *iTEM Version 1.0 User's Manual*, Coordinated Science Laboratory, University of Illinois, Urbana, IL, 1996.

- [55] C. Mead and L. Conway, *Introduction to VLSI Systems*. Reading, MA: Addison-Wesley, 1980.
- [56] J.-I. Pi and S.-L. Lu, *Magic Technology Manual for MOSIS' Scalable CMOS*, Department of Electrical & Computer Engineering, Oregon State University, 1994.
- [57] R. N. Mayo, M. II Arnold, W. S. Scott, D. Stark, and G. T. Ilamachi, *1990 DECWRL/Livermore Magic Release*, Western Research Laboratory (WRL) Research Report, July 1990.
- [58] C. Hu, S. C. Tan, F.-C. Hsu, P.-K. Ko, T.-Y. Chang, and K. W. Terrill, "Hot-electron-induced mosfet degradation – model, monitor, and improvement," *IEEE Trans. Electron Devices*, pp. 375–385, February 1985.
- [59] M. Brox, A. Schwerin, Q. Wang, and W. Weber, "A model for the time- and bias-dependence fo p-mosfet degradation," *IEEE Transactions Electron Devices*, pp. 1184–1196, July 1994.
- [60] P. Heremans, R. Bellens, G. Groeseneken, and H. E. Maes, "Consistent model for the hot-carrier degradation in n-channel and p-channel mosfet's," *IEEE Transactions Electron Devices*, pp. 2194–2209, December 1988.
- [61] R. Woltjer, A. Hanada, and E. Takeda, "Time dependent of p-mosfet hot-carrier degradation measured and interpreted consistently over ten orders of magnitude," *IEEE Transactions Electron Devices*, pp. 392–401, February 1991.
- [62] B. S. Doyle and K. R. Mistry, "A lifetime prediction method for hot-carrier degradation in surface-channel pmos devices," *IEEE Transactions Electron Devices*, pp. 1301–1307, May 1990.
- [63] T. Tsuchiya, Y. Okazaki, M. Miyake, and T. Kobayashi, "New hot-carrier degradation mode and lifetime prediction method in quarter-micrometer pmosfet," *IEEE Transactions Electron Devices*, pp. 404–408, February 1992.

- [64] R. Woltjer, G. M. Paulzen, H. C. Pomp, H. Lifka, and P. H. Woerlee, "Three hot-carrier degradation mechanisms in deep-submicron pmosfet," *IEEE Transactions Electron Devices*, pp. 109–115, January 1995.
- [65] P. Heremans, J. Witters, G. Groeseneken, and H. E. Maes, "Analysis of the charge pumping technique and its application for the evaluation of mosfet degradation," *IEEE Transactions Electron Devices*, pp. 1318–1335, July 1989.
- [66] G. Groeseneken, H. E. Maes, N. Beltran, and R. F. D. Keersmaecker, "A reliable approach to charge-pumping measurements in mos transistors," *IEEE Transactions Electron Devices*, pp. 42–53, January 1984.