ANALYSIS AND DESIGN OF RELIABLE MIXED-SIGNAL CMOS CIRCUITS

A Dissertation
Presented to
The Academic Faculty

By

Xiangdong Xuan

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in
School of Electrical and Computer Engineering

Georgia Institute of Technology

December 2004

Copyright © Xiangdong Xuan 2004

ANALYSIS AND DESIGN OF RELIABLE MIXED-SIGNAL CMOS CIRCUITS

Approved by:

Dr. Abhijit Chatterjee, Advisor

Dr. David C. Keezer

Dr. Adit D. Singh

Dr. Gary S. May

Dr. Madhavan Swaminathan

Date Approved: July 26, 2004

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Abhijit Chatterjee, for his precious support and advice guiding me through my whole Ph.D. career. I will always be grateful that he gave me such a wonderful opportunity working on this research project in an encouraging and vigorous atmosphere, from which I benefit for the whole life.

I would like to thank Dr. Gary May, Dr. David Keezer, and Dr. Madhavan Swaminathan, for being on my thesis committee, and for providing valuable feedback and suggestions on my thesis and research. I would like to extend my sincere appreciation to Dr. Adit Singh, who has given me important advices on some key research issues and served as my thesis committee member.

I would like to acknowledge the project sponsorship of the U. S. Air Force Research Laboratory, the Northrop Grumman Corporation, and NSF, as well as the great help on the experiments from The Boeing Company through collaborative work at Seattle, WA.

I also thank my group-mates in the mixed-signal testing group at Georgia Tech, for their help and friendship through all these years.

I am grateful to my parents, Yimin Zhao and Zian Xuan, for their love and encouragement from the other side of the earth, and to my sisters, Zhengnan and Zhanbei Xuan, for their emotional support and motivation.

Most of all, I would like to express the gratitude deep in my heart to my dear wife, Ang, for her infinite love and understanding, and for accompanying me through ups and downs in our life during these unforgettable years.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES ix

SUMMARY xiii

CHAPTER 1. INTRODUCTION 1

1.1. IC reliability 1

1.2. Reliability degradation and simulation 4

1.3. Design for reliability 8

1.4. Research objectives 10

1.5. Organization of contents 11

CHAPTER 2. MODELING OF FAILURE MECHANISMS 13

2.1. Electromigration 13
2.1.1. Mechanism overview 13
2.1.2. Failure physics 14
2.1.3. Physics-of-failure modeling 17
2.1.4. Incorporation of physical defects 22

2.2. Hot-carrier 24
2.2.1. Mechanism overview 24
2.2.2. Failure physics 25
2.2.3. Physics-of-failure modeling 28

2.3. Gate oxide wear-out 34
2.3.1. Mechanism overview 34
2.3.2. Physical models 34

CHAPTER 3. CIRCUIT LEVEL RELIABILITY SIMULATION 36

3.1. Hierarchical reliability evaluation 36
3.1.1. Algorithm description 36
3.1.2. Simulation examples 38

3.2. EM degradation modeling using defect statistics 41
3.2.1. Component level post-fab EM reliability 42
3.2.2. Circuit level post-fab EM reliability 43
3.2.3. Lifetime prediction under post-fab EM degradation 46

CHAPTER 4. ARET – ASIC RELIABILITY EVALUATION TOOL 56

4.1. Tool overview 56

4.2. Reliability simulation function 58

4.3. Reliability hotspot identification function 63

CHAPTER 5. ARET CALIBRATION 65

5.1. Test structures 65
5.1.1. EM test structures 66
5.1.2. HC test structures 68
5.1.3. Test structures for circuit level simulation algorithms 68

5.2. Stress tests 70
5.2.1. Tests for EM test structures 72
5.2.2. Tests for HC test structures 74
5.2.3. Tests for circuit level test structures 74

5.3. Calibration of ARET 76

CHAPTER 6. DESIGN FOR RELIABILITY 83

6.1. Reliability hotspot identification 83
6.1.1. Hotspot under interconnect failure 84
6.1.2. Hotspot under device degradation 84

6.2. Basic DFR approach 88

CHAPTER 7. DFR WITH INTERCONNECT FAILURES 94

7.1. Basic approach 94

7.2. Algorithm implementation 95

7.3. Experimental results 97

CHAPTER 8. DFR FOR CMOS DIGITAL CIRCUITS 102

8.1. DFR by dimension modulation 102
8.1.1. Algorithm based on inverter network 102
8.1.2. Algorithm with technology scaling 108
8.1.3. Complete algorithm for CMOS digital logic 109
8.1.4. Circuit area involved 115
8.1.5. Algorithm implementation 116
8.1.6. Discussions and trade-offs 119

8.2. DFR by signal modulation 120
8.2.1. Algorithm based on inverter network 120
8.2.2. Algorithm feasibility 123
8.2.3. Algorithm with technology scaling 124
8.2.4. Complete algorithm for CMOS Digital Family 125
8.2.5. Circuit area involved 129
8.2.6. Algorithm implementation 130
8.2.7. Discussions and trade-offs 131

8.3. Other DFR approaches 133

8.4. Experiments 135
8.4.1. SPICE check 135
8.4.2. CMOS inverter chain 136
8.4.3. ISCAS benchmark circuits 140

CHAPTER 9. DFR FOR ANALOG CIRCUITS 146

9.1. Basic DFR approach 146

9.2. Implementation 148

9.3. Experiments 150

CHAPTER 10. CONCLUSIONS 152

APPENDIX A. ARET OPERATION GUIDE 155

A.1. System requirements 155

A.2. General operations 155
A.2.1. Open netlists 155
A.2.2. Select failure mechanisms 155
A.2.3. Select evaluation level 156
A.2.4. Sensitivity analysis 156
A.2.5. Re-design 157

A.2.6. Help 157

A.3. How to create netlist file 157
A.3.1. Device model parameters 157
A.3.2. Device parameters 158
A.3.3. Interconnect parameters 158
A.3.4. Signal parameters 159
A.3.5. Analysis specifications 159

A.4. How to add interconnect information into data file 160

APPENDIX B. ISCAS85 BENCHMARK CIRCUITS 163

REFERENCES 168

VITA 176

LIST OF TABLES

Table 1. Predicted interconnect lifetimes for different defect conditions........................ 44

Table 2. Predicted lifetimes of interconnect r2 in op-amp circuit. 45

Table 3. Interconnect lifetime prediction of op-amp. 55

Table 4. Simulation results for Al-5%Cu traces compared with measured data. 78

Table 5. Interconnect hotspot identification. 99

Table 6. Op-amp lifetime prediction after local DFR.. 100

Table 7. Results of DFR by dimension modulation on benchmark circuits. 141

Table 8. Results of DFR by signal modulation on benchmark circuits. 142

Table 9. Op-amp deigns with different sizing. 150

LIST OF FIGURES

Figure 1. IC reliability bathtub curve. 2

Figure 2. Feature size trend of Intel processor. 5

Figure 3. Interconnect current density trend of Intel chips. 6

Figure 4. Thesis organization. 12

Figure 5. Schematic illustration of metallurgical statistical properties of interconnect. 15

Figure 6. Two-dimensional grain texture generated by Voronoi approach. 18

Figure 7. Modeling of electromigration. 20

Figure 8. EM degradation of an Al interconnect trace by ARET. 21

Figure 9. Interconnect trace with a physical defect. 23

Figure 10. Charges and their locations in Si-SiO_2 system. 26

Figure 11. Measured nMOS interface trap distribution. 29

Figure 12. Triangular charge distribution profile. 29

Figure 13. Modeling of hot-carrier. 32

Figure 14. Drain current degradation of nMOS transistor under hot-carrier. 33

Figure 15. Hierarchical circuit reliability simulation. 36

Figure 16. Degradation of two-stage op-amp. 39

Figure 17. Degradation of CMOS mixer. 40

Figure 18. Degradation of CMOS logic path. 41

Figure 19. EM degradations of pure Al traces with different defects. 42

Figure 20. EM degradations of op-amp specs. 46

Figure 21. Defect probability distribution. 48

Figure 22. Two-stage analog op-amp. 53
Figure 23. ARET functions. 56
Figure 24. ARET at a glance. 57
Figure 25. Hierarchical reliability simulation with ARET. 59
Figure 26. Equivalent stress. 61
Figure 27. Lifetime prediction for post-fab IC interconnect under EM. 63
Figure 28. Reliability hotspot identification in ARET 64
Figure 29. Final layout and package of test structures 66
Figure 30. Straight traces. 66
Figure 31. Corner structure. 67
Figure 32. Spiral structure. 67
Figure 33. nMOS transistors by AMI C5N process. 68
Figure 34. CMOS inverter. 69
Figure 35. Two-stage op-amp. 69
Figure 36. Environmental chamber and temperature profiles. 70
Figure 37. Stress contactor board. 71
Figure 38. Test interface board. 71
Figure 39. EM measurement. 72
Figure 40. Schematic of measurement. 73
Figure 41. Measurement instrumentation. 73
Figure 42. HC measurement. 74
Figure 43. Inverter measurement. 75
Figure 44. Op-amp measurement. 75

Figure 45. EM model calibration with test data... 77

Figure 46. ID vs. V_D of test structure S5a (W/L=6μm/0.6μm). 79

Figure 47. ID vs. V_D of test structure S5c (W/L=9μm/0.6μm). 80

Figure 48. Noise margin degradation of inverter. 81

Figure 49. VTC shift of inverter. 82

Figure 50. Locating RCP and hotspot. 87

Figure 51. Basic local DFR strategy. 89

Figure 52. Hot-carrier degradation during signal transition in CMOS inverter. 91

Figure 53. Degradations of interconnect lines with different widths in 100 hours........... 95

Figure 54. Local DFR algorithm for interconnect under EM. 96

Figure 55. CMOS inverter network with effective capacitances. 103

Figure 56. Overall change in propagation delay of RCP after local DFR. 107

Figure 57. Available range of K_L in local DFR for different feature sizes. 109

Figure 58. Structure of CMOS static logic. 110

Figure 59. General digital logic network – dimension modulation. 111

Figure 60. $\Delta\tau$ after DFR by dimension modulation for different gate complexities. 114

Figure 61. Local DFR by dimension modulation. 116

Figure 62. Local redesign process in DFR... 118

Figure 63. CMOS inverter network with effective capacitances. 120

Figure 64. Change of delay in DFR by signal modulation for different feature sizes.... 123

Figure 65. $\Delta\tau$ vs. channel length in DFR by signal modulation. 125

Figure 66. General CMOS logic network. 127

Figure 67. Gain of speed after local DFR for circuits with different complexities. 128

Figure 68. DFR algorithm by signal modulation. 132

Figure 69. Schematic for power supply modulation. 133

Figure 70. Inverter network for SPICE simulation. 135

Figure 71. Verification of delay model used in local DFR. 136

Figure 72. CMOS Inverter chain. 137

Figure 73. CMOS inverter layout. 137

Figure 74. Degradation of propagation delay over the inverter chain. 139

Figure 75. Fanout distributions of ISCAS benchmark circuits. 143

Figure 76. Comparison between dimension modulation and signal modulation. 144

Figure 77. Mapping between circuit performance and device parameter. 147

Figure 78. DFR for analog circuits. 149

Figure 79. Op-amp Spectre netlist. 151

Figure 80. ISCAS benchmark circuit C432. 163

Figure 81. ISCAS benchmark circuit C880. 164

Figure 82. ISCAS benchmark circuit C1908. 164

Figure 83. ISCAS benchmark circuit C2670. 165

Figure 84. ISCAS benchmark circuit C3540. 166

Figure 85. ISCAS benchmark circuit C6288. 167

SUMMARY

Facing the constantly increasing reliability challenges under technology scaling, the topics in IC reliability technique have been receiving serious attention during recent years. In this work, based on the understanding of existing physical failure models that have been concentrating on the pre-fab circuits, a set of revised models for major failure mechanisms such as electromigration, hot-carrier, and gate oxide wear-out are created. Besides the modeling of degradation behaviors for circuits in design phase, these models tend to deal with the post-fab device characteristics with the presence of physical defects. In addition, the simulation work has been taken from device level to circuit level hierarchically, presenting the evaluation of circuit level reliability such as degradations of circuit level specs and circuit lifetime prediction. For post-fab ICs under electromigration, the expected circuit lifetime is calculated based on statistical processes and the probability theory.

By incorporating all physics-of-failure models and applying circuit level simulation approaches, an IC reliability simulator called ARET (ASIC reliability evaluation tool) has been developed. Besides the reliability evaluation, the reliability hotspot identification function is developed in ARET, which is a key step for conducting IC local design-for-reliability approaches. ARET has been calibrated with a series of stress tests conducted at The Boeing Company.

Design-for-reliability (DFR) is a very immature technical area, which has been becoming critical with the continuously shrinking reliability safety margin. A novel concept, local design-for-reliability is proposed in this work. This DFR technique is

closely based on reliability simulation and hotspot identification. By redesigning the circuit locally around reliability hotspots, this DFR approach offers the overall reliability improvement with the maintained circuit performance. Various DFR algorithms are developed for different circuit situations. The experiments on designed and benchmark circuits have shown that significant circuit reliability improvements can be obtained without compromising performance by applying these DFR algorithms.

CHAPTER 1

INTRODUCTION

Reliability of a product describes the probability that it functions as it is supposed to during a given period of time. For an integrated circuit (IC), as a critical product specification under today's aggressive technology scaling, reliability has always been very difficult and costly to measure, and to achieve in leading-edge technology. This work was motivated by the considerable benefit associated with efficient reliability evaluation and reliable circuit design.

1.1. IC reliability

Reliability is the ability of an item to perform a required function, under stated conditions, for a stated period of time, by the International Electrotechnical Commission [I.E.C., 1974]. The term reliability is also used as a reliability characteristic denoting a probability of success or a success ratio [1]. In IC manufacturing practice generally the reliability is specified either by the lifetime during that IC is expected to perform its designed functions, or by the failure rate that is the instantaneous probability that IC fails performing its functions at the given time. IC reliability failures could be due to both material wear-out and defects, and they occur after the ICs are delivered to the customers. An overview is given by Figure 1, which shows the typical IC reliability bathtub curve expressing the failure rate as a function of product lifetime.

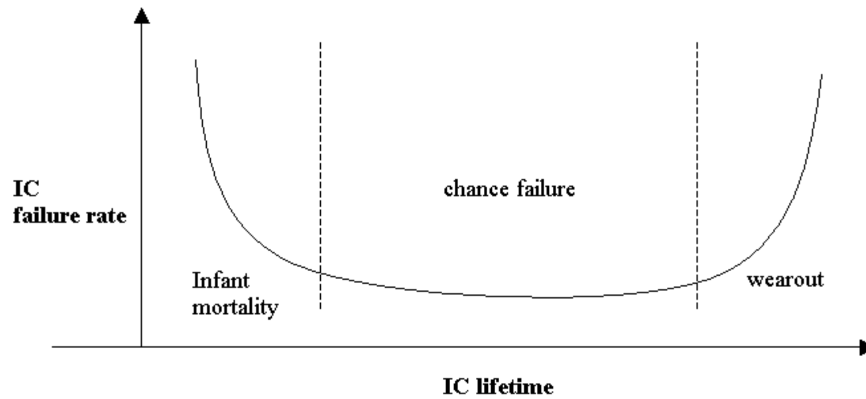


Figure 1. IC reliability bathtub curve.

In Figure 1, the early infant mortality period is attributed to defective material [2]. In this stage the failure rate is quite high and usually the highly expensive burn-in test is performed before product delivery to screen out the severely defective parts. The next region is the chance failure, in which the failure rate is low and nearly constant. This is the useful IC lifetime. The failures are mainly due to a low level of residual defects or electrical overstress/electrostatic discharge events. The qualification test is performed by IC reliability engineers to predict the failure rate and the IC lifetime. The final increase in failure rate occurs as the result of intrinsic material wear-out. For a mature process this region may not actually show up before the IC product is replaced by a new one.

Various models are used in reliability analysis. Among them, the cumulative failure function $F(t)$ is the most common entry point. It is defined as the cumulative probability that an IC fails at time t, or the fraction of the total number of ICs that have failed [2]. If we focus on a specific time interval, the failure probability density $f(t)$ is obtained by taking the derivative of $F(t)$. This is the so-called empirical hazard function and describes the probability of failure within a small time instant dt. Following this, the reliability

function *R(t)* is defined as the fraction of the surviving good parts at any time and is expressed as

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

In IC reliability engineering the more frequently used functions are the lifetime and the failure rate. The product lifetime is defined by the mean time to failure (MTTF). The failure rate *h(t)* is the expected instantaneous fraction of failures per unit time and is given as

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{2}$$

This is the function plotted in the bathtub curve, although in practice the model parameters are used more often than the curve. *h(t)* is sometime simplified to a constant in practice, which corresponds to an exponential distribution function for *f(t)*. For metals and oxides, the lognormal and Weibull distributions are used extensively for *f(t)* in experimental studies.

There are certain links between these reliability indices. Depending on different applications, one reliability function may be obtained from another by simplified theoretical conversion. For example, if the expected value of failure time is designated as the mean time to failure (MTTF), by recognizing

$$\int_0^\infty f(t)dt = 1 \tag{3}$$

the MTTF μ is thus given by

$$\mu = \int_0^\infty tf(t)dt \tag{4}$$

In the simplest case of exponential distribution, *f(t)* is expressed in the form

$$f(t) = \lambda e^{-\lambda t} \tag{5}$$

which generates

$$\lambda = \frac{1}{\mu} \tag{6}$$

where the failure rate *h(t)* is simply the constant λ by equation (2). This simple example shows how the other reliability measures, such as the failure rate *h(t)* can be derived from MTTF, at least from a statistical point of view.

1.2. Reliability degradation and simulation

For IC design, there is always a trade-off between reliability margins and the performance. In order to be faster and smaller, feature size has been dramatically shrinking. As an example shown in Figure 2 [3], the feature size of Intel processor has decreased from 3.0 μm to 0.09 μm in the past 25 years, and a significant increase of power supply for Intel CPUs was reported from the same source due to the increased operating frequency and transistor count. Other key device dimension such as oxide thickness and interconnect width have also decreased accordingly. The overall result is, by accepting the shrinking device dimension and subsequent high operating temperature, the IC has become much more vulnerable to failure mechanisms. Serious reliability challenges has been generated by aggressive technology scaling.

ICs are degraded by various failure mechanisms. In terms of interconnect failure, the ruling mechanism is electromigration (EM). Under elevated current density and temperature, EM can generate voids on the interconnect traces, which finally break the interconnect off. For device degradation, hot-carrier and oxide wear-out are two major mechanisms. The former is initiated by channel electric field and causes permanent oxide damage resulting in parameter degradations such as threshold voltage shift, while the

latter is due to oxide electric field and can generate defects inside the oxide that could induce the catastrophic oxide breakdown.

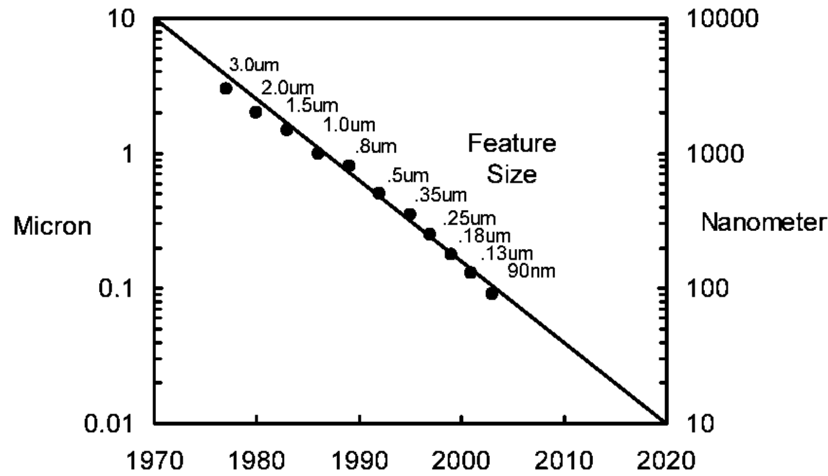


Figure 2. Feature size trend of Intel processor.

All three failure mechanisms are major causes of IC failures, and are becoming more serious with technology scaling down. This can be seen from Figure 2. With such a rapid dimension shrinking, if the power supply does not scale proportionally, almost every aspect of the circuit becomes more fragile. Unfortunately, this is exactly what has been happening. Not only the device is scaling, interconnect suffers too from EM damage due to increased driving current density. Figure 3 shows the current density inside interconnect for Intel processor chips [3], which clearly indicates that the interconnect current density increases with a rate of ×1.5 per generation for current and future technology. This serious fact has undoubtedly pointed out that further EM improvement is definitely needed for our technologies even for Cu interconnect, as stated in [3]. All these failure mechanisms are discussed in this work.

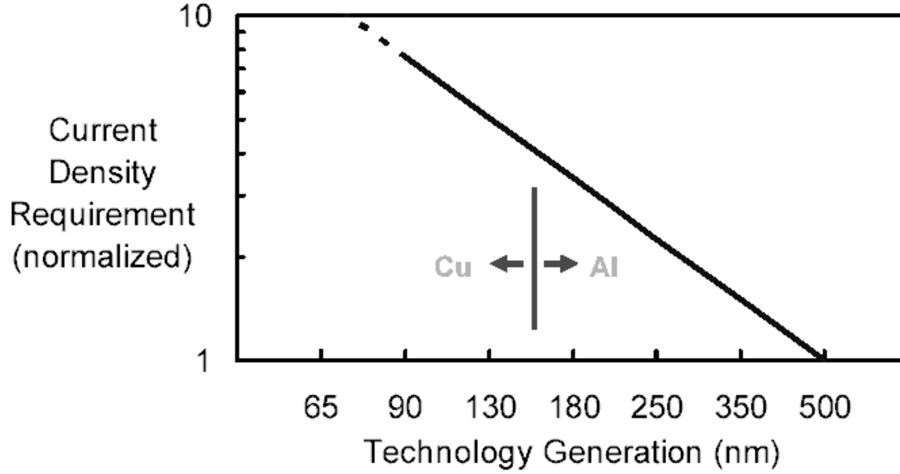


Figure 3. Interconnect current density trend of Intel chips.

Reliability tests are accelerated stress tests, and extremely expensive and time-consuming. The temperature acceleration factor, A_T, is given by Arrhenius as follow,

$$A_T = e^{E_a/k(1/T_u - 1/T_s)} \tag{7}$$

where E_a is activation energy, k is the Boltzmann's constant, T_u is use temperature, and T_s is stress temperature, and the voltage acceleration factor, A_V, is given as

$$A_V = e^{(E_s - E_u)/E_{eff}} \tag{8}$$

where E_s is stress field, E_u is use field, and E_{eff} is a factor dependent on temperature and oxide quality [4]. Two major processes that usually conducted by reliability engineers are burn-in test in the infant mortality, and the qualification test in chance failure region, as shown in Figure 1. Both use elevated temperature and voltage to make ICs fail sooner, and could take days even months to finish. In addition, the accuracy of accelerated stress test may be doubtful, since the parameters such as E_a and E_{eff} are usually determined at stress conditions, which is completely different from the use conditions. To deal with this

situation, reliability simulation technique is introduced as a critical supplemental means to stress tests in IC reliability evaluation process.

Reliability simulation technique is based on physics-of-failure modeling of IC failure mechanisms to simulate device and interconnect degradations. These physical degradations are then propagated to circuit and system levels by certain simulation approaches. Using reliability simulation, IC reliability can be evaluated as soon as the layout is done. Usually it only takes minutes to simulate a circuit design containing thousands of gates with acceptable accuracy. Moreover, it is much cheaper than the industrial stress tests such as burn-in and qualification tests. With the supplemental screening using reliability simulation tool, the stress test becomes less expensive with an enhanced confidence, and the failure analyses can be effectively facilitated as well.

Several works in reliability simulation area have been accomplished during the past years. Among them, RELY [5] at the University of Southern California, BERT [6], at the University of California, Berkeley, and ARET (ASIC Reliability Evaluation Tool) [7], developed at Georgia Institute of Technology have supplied the complete simulators for reliability evaluation under major IC failure mechanisms such as electromigration and hot-carrier. Some other results have also been reported [8][9]. However, most of reliability simulation tools so far have focused on 1) device level degradations, and 2) pre-fabrication ICs. These problems with existing reliability simulators cause the simulations unable to give a clear indication of when the circuit starts to malfunction. Even the simulation results at device level become less trustable due to the defective changes happening during the fabrication. Compared with the other reliability simulators, ARET successfully handles some reliability issues for post-fabrication ICs [10], with an

emphasis in circuit-level simulation. This is an essential step towards the accurate reliability simulation. It has been know for long that due to the uncontrollable processes during fabrication, various physical defects are to be generated inside the IC [11]. While the exact causes for these defects remain unclear, they could have crucial impact on circuit reliability. For example, a defect on an interconnect trace can significantly raise the current density giving the circuit a very short lifetime even the infant mortality. This is one of the main reasons that sometimes the reliability simulation shows a quite different result than the actual qualification test. In ARET, by handling the interconnect defect situation statistically based on probability theory, the expected lifetimes of post-fabrication circuits under electromigration can be obtained.

1.3. Design for reliability

While some solid progress have been made in reliability simulation in recent years, design approaches to reduce the circuit degradation and optimize the circuit design for reliability have not received enough attention in previous research. With the significantly narrowed reliability safety margin resulting from the aggressive feature size scaling of contemporary VLSI circuits into sub-micron range, there has been a rapid growing need for both topology-based and geometry-based design approaches that can be readily applied in the early-stage of IC development. This is especially true in the leading-edge technology generations, where the immature process has triggered serious reliability problems.

Design-for-reliability (DFR) was first introduced in IC development as a direct application of reliability simulation [12][13]. A simulation phase is added into IC design immediately after circuit design. The circuit reliability is thus evaluated before it is

fabricated. For a simulation result that does not meet reliability expectation, further investigations are taken in order to revise the original design followed by another reliability simulation. This design-simulation-redesign cycle is repeated until the reliability requirement is met. Reliability simulation technique has always been playing a fundamental role in DFR.

DFR is a tough challenge to both reliability and design engineers. This is due to the complex failure mechanisms, fast technology scaling, as well as lengthy and costly experimental support. By now, only a few works in this area have been presented in literature. Part of these research results is about the improvement of design rules for reliability [14], which basically present circuit level design guidelines such as reducing signal transient period. These guidelines are very useful in reliable sub-micron IC design. However, how to implement these guidelines in actual design is a critical issue and can be extremely challenging. There are several other works by M. A. Styblinski *et al*, proposing the drift reliability optimization based on the maximum income approach [15][16]. The algorithm presented is supposed to optimize the circuit reliability by a global design revision. It worked well for two small example CMOS circuits under hot-carrier only. However, the algorithm is very sophisticated and a large amount of computation time is required when working on large circuits, which generally limits the algorithm only to some applications. It can become even more limited as the ICs are getting more complex in the continuous technology scaling.

In this work, a new DFR approach – local design-for-reliability is proposed and implemented. This approach is based on reliability simulation technique. It takes advantage of the reliability hotspot identification function, which is another distinct

function developed for DFR in this work. The proposed local DFR technique greatly improves circuit overall reliability without compromising performance by only updating the design around the reliability hotspot. The design work involved is thus much simplified and reduced. Several DFR algorithms are developed for different circuits. A series of experimental results on various circuits have shown very promising reliability improvements to CMOS ICs by using this DFR technique.

1.4. Research objectives

The first major research objective of this work is to accomplish effective IC reliability simulation by developing a reliability simulation tool. To achieve this goal, the following tasks are set.

- Understand device level physics-of-failure modeling of IC major failure mechanisms for electromigration, hot-carrier, and gate oxide wear-out
- Develop circuit level reliability simulation algorithms
- Incorporate some post-fabrication defect effects in both device level and circuit level simulations
- Develop the reliability simulator, and calibrate the simulator with stress tests

The second major objective is to implement effective design-for-reliability for VLSI circuits based on reliability simulation by proposing the local design-for-reliability approach and developing the corresponding DFR algorithms. This includes the following tasks.

- Identify reliability hotspots
- Develop DFR algorithms for CMOS digital circuits
- Develop DFR algorithms for analog circuits

- Verify DFR approaches with experimental data

1.5. Organization of contents

This thesis is organized as shown by Figure 4, where the highlighted items indicate the major contributions of this work. Chapter 2 to Chapter 5 talk about IC reliability simulation technique. In Chapter 2 the physics-of-failure modeling process is discussed, and the impact of post-fabrication defects on IC interconnect reliability is evaluated. The major failure mechanisms involved are electromigration for IC interconnect failures and hot-carrier for device degradations, as well as gate oxide wear-out for that only a review of mechanism is presented. Based on these component level failure models, circuit level hierarchical simulation is discussed in Chapter 3. For post-fabrication ICs, a reliability model based on statistical process and probability theory is presented for circuit level interconnect lifetime prediction under electromigration. In Chapter 4, ASIC reliability evaluation tool (ARET) is developed as the final outcome of the reliability simulation work. Several critical issues in reliability simulation such as time stepping are discussed. The tool needs to be calibrated before use. This is done in Chapter 5 with a series of stress tests conducted at The Boeing Company. Some experimental details are also revealed such as test structure design, tester design, and data analyses.

With the foundation built by reliability simulation work, from Chapter 6 until Chapter 9, the simulation-based local design-for-reliability approach is presented and discussed. Chapter 6 provides the fundamentals of the proposed local DFR technique, where the basic approach is described and reliability hotspot identification function is developed. The chapters following Chapter 6 discuss the developments of various DFR algorithms. In Chapter 7, DFR for interconnect failures is presented. In Chapter 8, DFR

algorithms for CMOS digital circuits are developed, including dimension modulation, signal modulation, and etc. Experiments on designed and benchmark circuits are conducted to evaluate the algorithms. In Chapter 9, a high level local DFR algorithm is discussed for analog circuits based on design synthesis. The local design-for-reliability approach as well as its implementation closely depend on the reliability simulation technique and the understanding of major failure mechanisms.

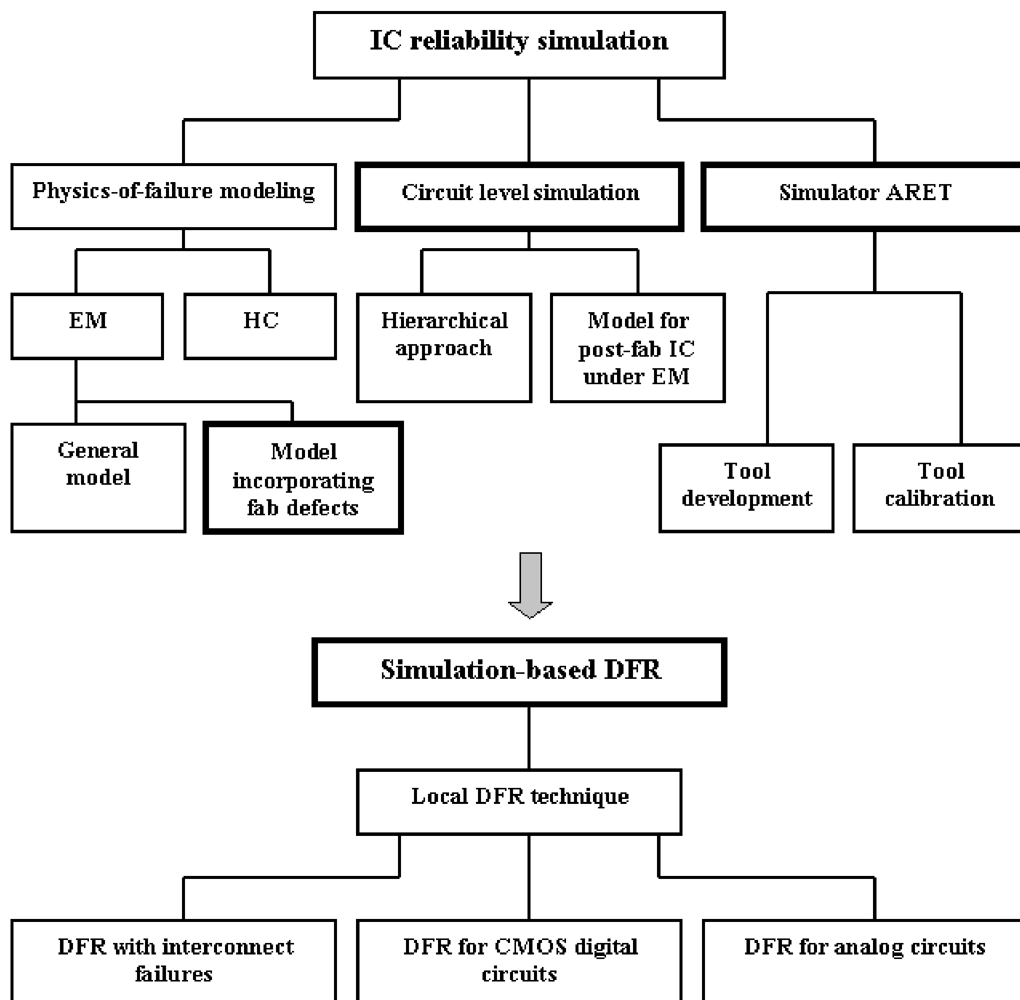


Figure 4. Thesis organization.

CHAPTER 2

MODELING OF FAILURE MECHANISMS

Failure mechanisms are the physical processes inside circuit components that are responsible for the characteristics degradation. The most active failure mechanisms vary from technology generation to generation. For contemporary VLSI circuits with dramatically shrunk feature sizes and dimensions, the major failure mechanisms are electromigration for interconnect, and hot-carrier and gate oxide wear-out for devices. The modeling of these major failure mechanisms is the foundation of any reliability simulation work.

2.1. Electromigration

2.1.1. Mechanism overview

Electromigration (EM) has been a major failure mechanism in discrete solid state devices and integrated circuits since 1970. Its classical definition refers to the structural damage caused by ion transport in metal thin films as a result of high current densities. EM damages are in forms of voids and hillocks on interconnect traces, where the void is the major concern due to the increased current density. In addition to current density, temperature and material properties also play a critical role. As a major failure mechanism that has been known by IC industry for long, EM is still with us today, and has been becoming a serious concern in terms of interconnect reliability with continuous technology scaling down [3]. Physics based models for electromigration are based on the magnitude of the electric field, grain boundary diffusivity, and grain boundary structural factors that determine the atomic flux distribution and the distribution of flux divergence.

2.1.2. Failure physics

Electromigration is due to mass transport in a diffusion-controlled process under certain driving forces. However, the driving force here is due not only to the concentration gradient in a pure diffusion process, also to the applied electric field. This includes the so-called “electron wind force” and the electric field force. The electron wind force refers to the effect of kinetic energy exchange between moving electrons and metal ion atoms when a current is applied to the IC interconnection. If the current density in the interconnection is high enough, the energy exchange can be significant resulting in noticeable mass transport and generate EM damage. At the same time, the positively charged ions also tend to move in the direction of the applied electric field, which is opposite to the direction of electron wind force. Thus, the balance of these two forces determines the movement of the ions. For example, in gold and aluminum, electron wind force dominates the ion movement and therefore the net driving force is in the direction of electron movement. In the temperature range commonly concerned ($<0.5T_{melt}$), the diffusion is mainly through grain boundaries.

Three predominant mechanisms exist in the EM failure process. They are 1) the metallurgical statistical properties of the conductor, 2) the thermal acceleration process, and 3) the so-called healing effects [17].

The metallurgical statistical properties refer to the microstructure parameters of the conductor, such as the grain size. These parameters can only be dealt with statistically since they are totally random. Generally the most meaningful parameters in this category are the misorientation angle θ, inclination angle ϕ, and the grain size distribution as shown in Figure 5.

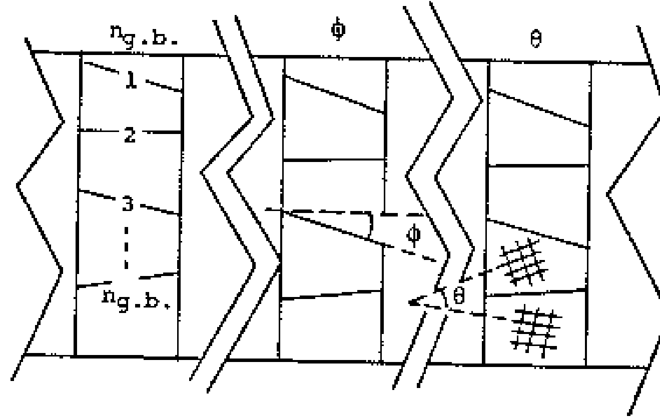


Figure 5. Schematic illustration of metallurgical statistical properties of interconnect.

The misorientation angle θ is the angle between two grain boundaries. It determines the mobility of the atoms at that boundary. The inclination angle ϕ is the angle between the grain boundary and the applied field. It determines the effectiveness of the electrical field at that boundary. And the grain size distribution determines the change in the number of the atomic paths across a cross section of the conductor. The variations of all these parameters can cause a non-uniform distribution of atomic flow rate resulting in a nonzero atomic flux divergence, which is

$$\nabla \bullet J = \sum_{i=1}^{ngb} J_i \tag{9}$$

where J_i is the atomic flux at ith grain boundary, n_{gb} is the number of grain boundaries defining an intersection that is most likely to be the failure site. It should be noted that the grain boundary intersections often represent the locations where the mass flux has the maximum divergence. At such areas there can be an abrupt change in grain size. This can produce a change in the number of paths for mass movement. There can also be some other microstructure changes affecting the atomic diffusivity.

The thermal acceleration process refers to the accelerating EM damage due to a rise in the local temperature. Once a void is initiated, the current density in the void area increases due to the reduced cross section area of the conductor. This is referred to as the current crowding effect. Since the joule heat is proportional to the square of the current density, this current crowding effect leads to a local temperature rise around the void area, which in turn accelerates the void growth. It is shown in the following sections that this acceleration can be dramatic as the temperature is in an exponential term of EM equation.

One of the approaches to obtain the temperature distribution is to solve the thermal equation assuming constant boundary conditions, that is, constant ambient temperature at the ends of the two-dimensional lines in x-y plane [18]

$$\frac{\partial}{\partial x}(\tau \frac{\partial T}{\partial x}) + \frac{\partial}{\partial y}(\tau \frac{\partial T}{\partial y}) + j^2 \rho_0 (1 + \alpha \Delta T) = 0 \tag{10}$$

where $\tau = \tau(x, y)$ is the thermal conductivity coefficient, ρ_0 is the resistivity, j represents the current density, and α is the temperature coefficient of the resistivity. In most experiments, the substrate of the conductor is kept either in a hot stage or in a constant-temperature chamber. Under such circumstances the thermal equation becomes [19][20]

$$-\tau \frac{\partial^2 T}{\partial x^2} - \tau \frac{\partial^2 T}{\partial y^2} = Q - \frac{\lambda}{h}(T - T_s) \tag{11}$$

where λ is the heat transfer coefficient between the film and the substrate, and Q is the Joule heat generated per unit volume per unit time.

The healing effect is caused by the atomic flow in the direction opposite to the electron wind force. This backflow can happen during or after electromigration. It is mainly because of inhomogeneities, such as temperature and concentration gradient,

resulting directly from EM damage. The healing effect tends to reduce the failure rate during electromigration and heal the damage after the applied current is taken away. So there exists a threshold current density for electromigration to become effective as the result of this healing effect. The value of this threshold depends on the minimum energy barrier that the atoms have to overcome to balance off the backflow driving force. This can be obtained by the following equation [21]

$$(jL)_{th} = \left| \frac{\Omega_0 \sigma_{\max}}{Z^* q'} \right| \tag{12}$$

where $(jL)_{th}$ defines the threshold value of the product of line length and current density, Ω_0 is atomic volume, σ_{max} is the maximum stress along the line, and Z^*q' represents the effective charge of the ions.

2.1.3. Physics-of-failure modeling

The essential work in EM modeling is generating the grain boundary texture including all these metallurgical statistical properties. In most applications, triple grain boundary junctions are the majority of grain boundary intersections where nonzero flux divergence usually happens, and a two-dimensional junction network of the material grain texture can be used to model EM process.

The most commonly accepted method for generating such a grain texture is the "Voronoi polygon" approach. In this approach, polygons are generated in a random fashion to represent the grains in the film [18]. First, the conductor stripe is discretized into a grid-like network with the cells being rectangular in shape. All cells are equal in size representing the average grain size. Then the crystal seed points are randomly laid down into the cells according to prescribed cell density (number of points per cell). These seed points are nucleating centers of grain boundary junctions. The edges of the polygons

are formed by constructing the perpendicular bisectors of rays connecting a given seed point and its neighboring seed points. Figure 6 shows a typical grain network generated using the Voronoi approach.

Figure 6. Two-dimensional grain texture generated by Voronoi approach.

When triple junctions are considered as the only areas where flux divergence exists, the procedure of grain texture generation can be simplified so that only triple grain boundary junctions are generated. This is called the triple-junction-lattice method [22]. In this approach, after the conductor line is discretized and the seeds are laid down, the seeds represent the triple junctions and the values of the parameters, such as θs and ϕs, are then assigned to each grain boundary randomly. The random assignment of the microstructural parameters is consistent with the randomness of the grain distribution generated by the Voronoi approach. Once the network is generated, all microstructural parameters can be extracted.

To begin the modeling process, a structural factor, ΔY, is used and defined as

$$\Delta Y = \sum_{i=1}^{ngb} \Theta_i \cos\phi_i \tag{13}$$

where ϕ is the inclination angle and the parameter Θ_i is defined by

$$\Theta_i = e^{-\Delta Qi(\theta i)/kT} \tag{14}$$

The structural factor at each cell is extracted from the grain texture generated. The flux divergence can finally be expressed in the form

$$\nabla \bullet J = \frac{N_{gb} D_0}{kT} Z^* q \rho_0 (1 + \alpha \Delta T)(j - j_c) \Delta Y e^{-Q0/kT} \tag{15}$$

where N_{gb} is the grain boundary concentration, ρ_0 is resistivity of the conductor and α is the temperature coefficient of the conductor resistivity. Here j_c represents the threshold current density due to the healing effect. Thus, the growth rate of the volume V of the mass depleted (or accumulated) at the grain boundary intersection becomes

$$\frac{\partial V}{\partial t} = \delta h \Omega_0 \nabla \bullet J \tag{16}$$

where δ is the grain boundary width, h is the thickness of the conductor film and Ω_0 is the atomic volume as previously defined. With this geometrical expression of the void, assuming a cylindrical void shape, the elemental fractional resistance change of the cell on the ith row and jth column, $(\Delta R/R0)_{ij}$, can easily be shown as

$$(\frac{\Delta R}{R_0})_{ij} = \frac{w}{l}[\frac{2}{\sqrt{1-x^2}} \tan^{-1}(\frac{\sqrt{1+x}}{\sqrt{1-x}})] - x - \frac{\pi}{2} \tag{17}$$

where w and l are the width and length of the cell, x is the normalized diameter of the cylinder d by the cell width, $x=d/w$. And the diameter of the cylindrical void can be obtained from the void volume given in Equation (16).

To calculate the total resistance of the conductor line, all cells have to be connected in an appropriate manner. There are two possible styles of connecting these cells: the parallel of series (PS) mode and series of parallel (SP) mode [17]. In SP mode, the resistance of each cell column is first calculated as if the cells are connected in parallel, and the total resistance of the line is then obtained by considering all cell columns in

series and adding up their resistances. The PS mode is constructed in the similar way with the series resistance calculated first. In the case the length of the conductor line is much larger than the width, the SP mode should be employed, computing the total resistance of the conductor line in the following form.

$$R_T(t) = \frac{n_w R_T(0)}{n_l} \sum_{i=1}^{nl} \{ \sum_{j=1}^{nw} [1 + (\frac{\Delta R}{R_0})_{ij}]^{-1} \}^{-1} \tag{18}$$

where $R_T(0)$ is the initial resistance of the conductor line, *nl* and *nw* denote the number of cells along the length and across the width, respectively. Thus, the physics model for electromigration degradation of interconnect line has been created. The model implementation is described by the flow chart shown in Figure 7.

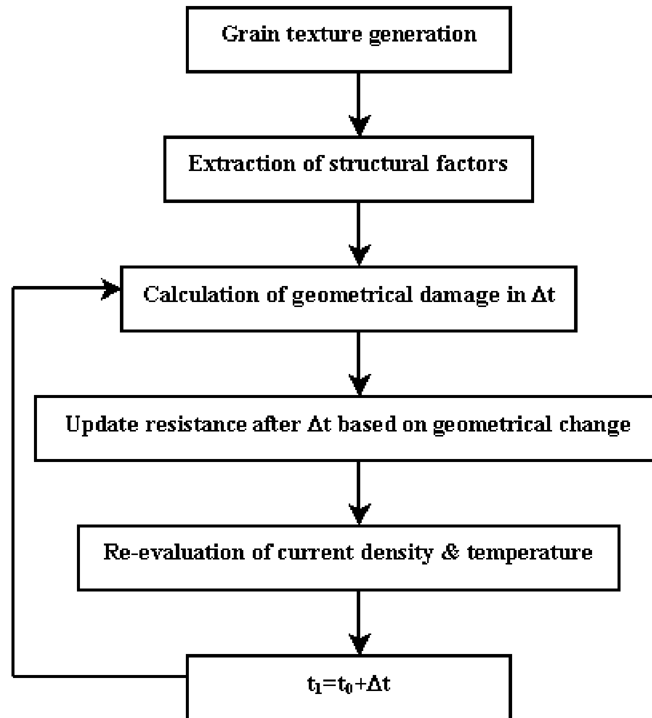


Figure 7. Modeling of electromigration.

In the chart, the generation of the grain texture is essential for the whole modeling process, based on which the structural factors are extracted. The geometrical change is evaluated using Equation (16) resulting in a resistance change updated by Equation (18). This process is repeated at every time step to generate the plot of resistance vs. time. Another issue is the thermal modeling. In practical cases the heat is not from the internal current only. In some special applications, such as for eddy currents in spiral conductors at a high operating frequency, the external heat source also needs to be considered.

A simulation result using the above-discussed EM model is shown in Figure 8. The interconnect under simulation is a pure aluminum trace with a dimension of 118×2×1 μm under 200°C and 0.1A DC current. The result is the resistance percentage change vs. time from the output of the reliability simulator ARET, which is discussed in details in Chapter 4.

Figure 8. EM degradation of an Al interconnect trace by ARET.

It can be seen that in 50 hours the resistance of the trace was increased by 24% due to the EM voids. The fast degradation is mainly caused by the high current density as well as the elevated temperature, where the temperature actually plays the critical role with an almost exponential acceleration to EM void growth.

2.1.4. Incorporation of physical defects

Like other EM models that have been reported in the literature, the physics model of EM degradation created thus far in this work is completely for the pre-fabrication IC interconnect, which assumes a defect-free circuit. Unfortunately, this has not been the case in practice for a fabricated IC, because of the existence of minor deformations and defects in all the fabricated interconnect lines (conductor traces) due to variance in the manufacturing process.

These physical defects generally fall into two major categories: global defects, those that affect multiple ICs across a relatively large area of the wafer and local defects, those that affect a relatively small area of the IC. Global defects include line dislocations and fabrication process control errors, which are usually called “systematic defects”. For example, the width variations of interconnect traces are systematic defects. Such defects can be easily detected early in the manufacturing process. Furthermore, for a mature fabrication process, these defects are due to process control errors, which can be minimized through careful cause-effect analysis. Unlike global defects, local defects originate from distinct, usually complicated and uncontrollable processes in the fabrication and thus can be considered random. It includes silicon substrate inhomogeneities, local surface contaminations, and photolithographic point defects. This type of defect is the primary target in terms of interconnect EM process evaluation.

To better understand the physical defect on IC interconnect, a schematic in Figure 9 shows a random defect crack presented on an interconnect trace. Due to the defect, the width of the trace is reduced to d′ from d and so is the cross-section area. This causes an increase in the current density, so-called the current crowding. During the EM process, the metal ions obtain the kinetic energy transferred from moving electrons to form mass flow. Thus, with the increased current density, clearly more energy is transferred to the ions and the EM degradation can be directly accelerated.

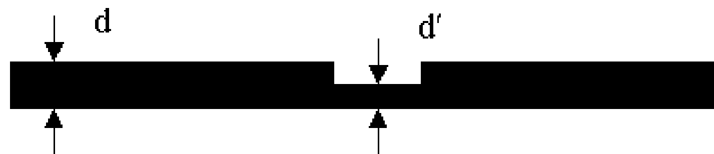


Figure 9. Interconnect trace with a physical defect.

From the energy perspective, as the current density increases the local temperature rises in the defect area causing a more unstable and disturbed status. The activation energy for metal ions to run off their original equilibrium positions is much reduced. This change basically follows an exponential function to accelerate EM damage. For the above reasons, the analysis of physical defects on IC interconnects is crucial for EM reliability evaluation.

In order to create an accurate thermal profile at the defect area, the thermal equations, Equation (10) and (11), are solved with the relationship between the Joule heat, current density and the temperature change described below.

$$Q = j^2 \frac{Wh}{\lambda_0} R_{th} \tag{19}$$

$$\Delta T = \frac{j^2 \rho}{h} \tag{20}$$

where ρ is the resistivity at temperature T, λ_0 is the average grain size, and R_{th} represents the thermal resistance per unit volume. For simplicity of analysis, the temperature gradient in the non-defective part of the interconnect trace is ignored.

Thus, the existing EM model is finally upgraded with both current density and temperature change at the defect area re-evaluated based on a partition process. The incorporation of physical defects in EM degradation basically proceeds in four major steps:

- Partition the interconnect trace into “defect-free” and “defective” segments based on the location(s) where defects are introduced.
- Calculate the current density in each interconnect segment.
- Modify the structural factors at the grain boundaries of the perfect segment(s) and defective segment(s).
- Determine the thermal profile at the defect site(s).

2.2. Hot-carrier

2.2.1. Mechanism overview

Hot-carrier (HC) induced degradation of MOS transistors is one of the primary mechanisms affecting the long-term reliability of VLSI circuits. It has been aggravated due the downward scaling of transistor dimensions without proportional scaling of the operating voltage [23]. Since the early 1980s, there has been an enormous increase in the

amount of research and literature in the area of VLSI hot-carrier reliability and the related technology has become much more mature.

During the operation of a transistor, due to the reduction in transistor dimensions, the electric fields along the channel are significantly increased in both horizontal and vertical directions. Those electrons and holes that gain enough kinetic energy under the electric fields can be injected into the gate oxide, causing permanent changes to the charge distribution at the oxide-interface. Therefore, the current-voltage characteristics of the MOSFET are degraded. These involved electrons and holes are referred to as hot-carriers, simply because that the particles having the same energy can be very "hot" if measured by their effective temperatures.

Under the same electric field, holes require much higher drain voltage to activate the hot-carrier effect due to the lower charge mobility. Experimental evidences have indicated that hot-carrier damages in nMOS transistors are more severe than in pMOS, which is why the hot-electron mechanism has been the major objective in most of research results over hot-holes. In addition, because of the very similar mechanisms, the modeling process based on hot-electron effect in nMOS transistors can be readily applied to pMOS transistors with minor modifications. Therefore, in this work, the physics-of-failure modeling is conducted only for the hot-electron in nMOS transistors although it has been suggested by some research results that the hot-carrier effect in pMOS is getting more significant in submicron technology.

2.2.2. Failure physics

The physical properties of the silicon-oxide interface and the gate oxide layer, and the gradual changes in these properties under operating conditions ultimately determine

the long-term hot-carrier reliability of the MOS transistor. Figure 10 shows the typical charge distribution at the MOSFET oxide-silicon interface [23].

Four types of charges exist at the oxide-silicon interface. They are the fixed oxide charge, mobile oxide charge, oxide trapped charge, and interface trapped charge. The fixed oxide charge is due to structural defects and it is not influenced by the electrical operating conditions of the MOS transistor. The mobile charge is primarily due to ionic impurities in the oxide, such as Na^+, K^+. These two types of charges basically do not contribute to the hot-carrier degradation. However, the oxide trapped charge and the interface trapped charge play an important role in the gradual degradation of oxide characteristics.

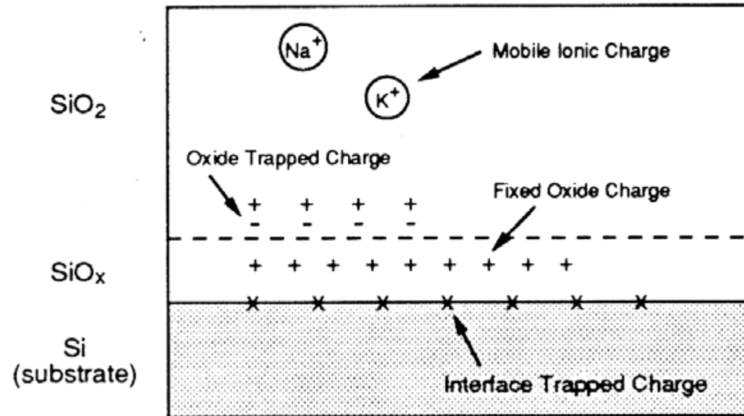


Figure 10. Charges and their locations in Si-SiO_2 system.

The hot-carrier damage is created when electrons and holes with high kinetic energies overcome the silicon-oxide potential barrier and enter the gate oxide, resulting in a change of the charge distribution. The charge distribution of the gate oxide is changed when excess electrons or holes are captured by the traps in the oxide, or by impact release

of the trapped electrons or holes by a hot-carrier. The probability of these injected carriers being captured by an empty trap depends on the available trap density and the trapping cross-section. Early efforts to model hot-carrier induced degradation have focused on localized charge trapping as the main cause [24][25]. However, recently it has been recognized that both charge trapping and interface trap generation contribute to the degradation of the device characteristics.

New interface traps are generated in nMOS transistors by hot-carriers, which upon injection into the Si-SiO_2 interface break the electron-pair bonds. Several atomic mechanisms for the creation of interface traps have been postulated by Sah [26]. It must be recognized that the hot-carrier induced interface traps are localized in a narrow region near the drain of the transistor (about 0.1 μm), since the lateral electrical field accelerating the electrons and holes in the channel attains its maximum near or in the drain area. Let $\Phi_{it,e}$ and $\Phi_{it,h}$ be the critical energies for electrons and holes, respectively, to form fast interface traps upon injection. The portion of the channel current density that consists of electrons with kinetic energies higher than $\Phi_{it,e}$ can be expressed as the bond-breaking current in form of [23]

$$I_{BB,e} = \frac{C_1}{W} I_{DS} \exp(-\frac{\Phi_{it,e}}{q\lambda_e E_m}) \tag{21}$$

where C_1 is an experiment-determined coefficient, λ_e represents the mean-free path for electrons. E_m represents the maximum lateral electrical field along the channel, which is defined by the following equation

$$E_m = \frac{V_{DS} - V_{DSAT}}{\sqrt{3t_{ox}x_j}} \tag{22}$$

where V_{DS} and V_{DSAT} are drain-source voltage and saturation voltage at drain, respectively, t_{ox} is the oxide thickness, and x_j represents the junction depth. Heremans *et al*. have indicated that C_1 is between 1.9 and 2 [27]. The bond-breaking current for holes can be obtained in a similar way. The net rate of interface trap generation is expressed as

$$\frac{dN_{it}}{dt} = KI_{BB,e} - B_p N_{it} n_H(0) \tag{23}$$

where the coefficient K is proportional to the density of the silicon-hydrogen bonds at the interface, B_p is a process-dependent constant, and $n_H(0)$ is the concentration of H at the interface. Once again, the interface trap generation rate for holes can be expressed similarly. Thus, by trapping the charges in the transistor channel, the trap generation simply means the change of interface charge distribution Q_{it}.

The localized oxide charge trapping and/or interface trap generation, as described above, gradually build up and permanently change the transistor oxide-interface charge distribution as the result of high-energy hot-carrier injection [28]. This causes the degradation in critical transistor parameters, such as the flat-band voltage, drain current, transconductance, and threshold voltage.

2.2.3. Physics-of-failure modeling

The degradation due to the hot-carrier effect is produced by the localized physical damage represented by the disturbed charge distribution along the channel. Therefore, in order to model the hot-carrier degradation, the first step is to generate the charge distribution profile to simulate the real situation. According to the previous research and published experimental data, the interface trap distribution is triangle-like in shape with a very sharp and localized peak near the drain area [29], as shown in Figure 11.

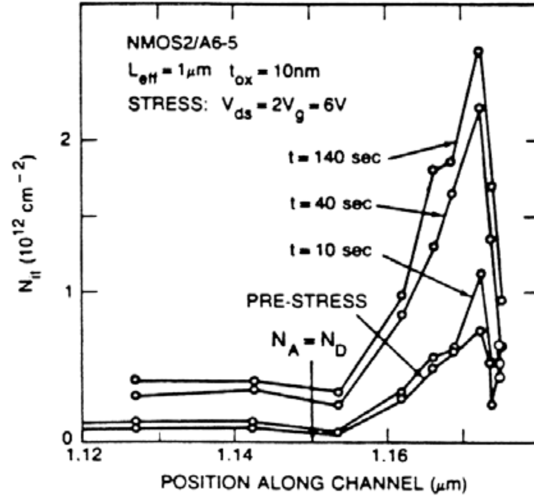


Figure 11. Measured nMOS interface trap distribution.

Thus, based on the experimental evidence, a simple triangular charge density distribution profile is used for the derivation of model equations, as shown in Figure 12. If the channel is designated as y axis, the oxide-interface charge density Q_{it} is then expressed in Equation (24), with the damaged region denoted by L_2 and undamaged region denoted by L_1.

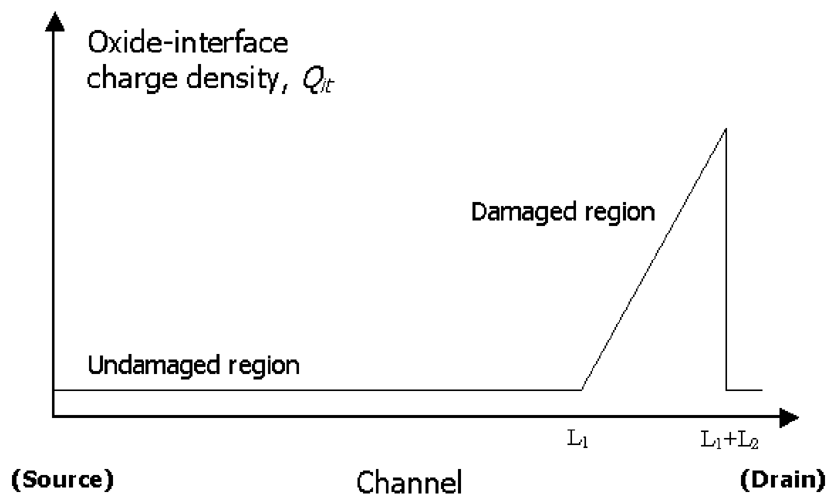


Figure 12. Triangular charge distribution profile.

$$
Q_{it}(y) = \begin{cases} 0 & 0 < y < L_1 \\ \dfrac{Q_{peak}}{L_2}(y - L_1) & L_1 < y < L_1 + L_2 \end{cases} \tag{24}
$$

With the created charge distribution profile, the degradations of key transistor parameters under hot-carrier can be modeled. Among these parameters, the drain current is a very important one to describe the characteristics a MOS transistor. It also has, in turn, the direct impact on the induced damage, since the bond-breaking current in Equation (21) is part of the channel drain current. To model the transistor drain current, the two operation regions, linear region and saturation region, have to be considered separately. Also, the drain current model derivation is based on the assumption that the gradual-channel approximation is valid for the damaged nMOS transistor. This means that the electric field in the direction of current flow is much smaller than the field perpendicular to the silicon surface allowing the one-dimensional analysis of the drain current.

In linear region, the effective channel length is the whole channel between source and drain. The drain current in undamaged region L_1 can be expressed as

$$
\begin{aligned} I_D = \frac{W}{L_1}\mu_1 C_{ox}\{(V_G - V_{FB1} - 2|\Phi_P|)V_P - \frac{{V_P}^2}{2} \\ - \frac{2\sqrt{2\varepsilon_S q N_a}}{3C_{ox}}[(2|\Phi_P| - V_B + V_P)^{3/2} - (2|\Phi_P| - V_B)^{3/2}]\} \end{aligned} \tag{25}
$$

where V_G, V_D are the voltages at gate and drain, V_P is the voltage at $y=L_1$, V_{FB1} is the flat-band voltage in region L_1, Φ_P represents the potential energy, V_B is the bulk voltage. In the damaged region L_2, it is changed to

$$I_D = \frac{W}{L_2}\mu_2 C_{ox}\{[(V_G - V_P) - V_{FB2} - 2|\Phi_P|](V_D - V_P) - \frac{(V_D - V_P)^2}{2}$$
$$- \frac{2\sqrt{2\varepsilon_S q N_a}}{3C_{ox}}[(2|\Phi_P| - (V_B - V_P) + (V_D - V_P))^{3/2} - (2|\Phi_P| - (V_B - V_P))^{3/2}]\} \quad (26)$$

with the similar definitions of parameters used in the equation. Here the flat-band voltage is defined as

$$V_{FB}(y) = \Phi_{MS} - \frac{Q_{ox}}{C_{ox}} - \frac{Q_{it}(y)}{C_{ox}} \quad (27)$$

where Φ_{MS} is the work function difference, Q_{ox} represents the constant positive oxide-interface charge density, and the interface charge Q_{it} under hot-carrier effect is defined in Equation (24). The flat-band voltages used in Equation (25) and (26) are the average values.

When the transistor is working in the saturation region, the effective operating channel shrinks due to channel length modulation. In this case, it can be approximated as the same transistor discussed in linear region situation, except that the damaged channel length L_2 has to be modified by the following equation due to channel length modulation.

$$\frac{1}{\Delta L} = \sqrt{\frac{qN_a}{2\varepsilon_S[V_D - V_{CE}(\Delta L)]}} + \frac{\varepsilon_0}{\varepsilon_S t_{ox}} \frac{\alpha[V_D - V_G^{'}] + \beta[V_G^{*}(\Delta L) - V_{CE}(\Delta L)]}{V_D - V_{CE}(\Delta L)} \quad (28)$$

However, in this equation V_{CE} is function of ΔL and the flat-band voltage, which is a function of the oxide-interface charge density. So here the Newton-Raphson iterations are needed to solve the equation. Once ΔL is obtained, the transistor in the saturation region can then be treated as a transistor with a damaged region equal to L_2-ΔL, instead of L_2. In the case that ΔL is larger than L_2, the transistor can simply be treated as an undamaged one in terms of hot-carrier effect.

Other basic transistor properties are also obtained from the oxide-interface charge density, the drain current, and the flat-band voltage. For example, the threshold voltage is expressed in the form

$$V_{Th} = V_{FB} + 2|\Phi_P| + \frac{\sqrt{2\varepsilon_S q N_a}}{C_{ox}} \sqrt{2|\Phi_P| - V_B} \tag{29}$$

Since the change of the charge distribution under hot-carrier is a function of time, the degradations of all parameters can be described as time functions. With a given failure criterion, this allows the prediction of device lifetime. The whole process is demonstrated in Figure 13.

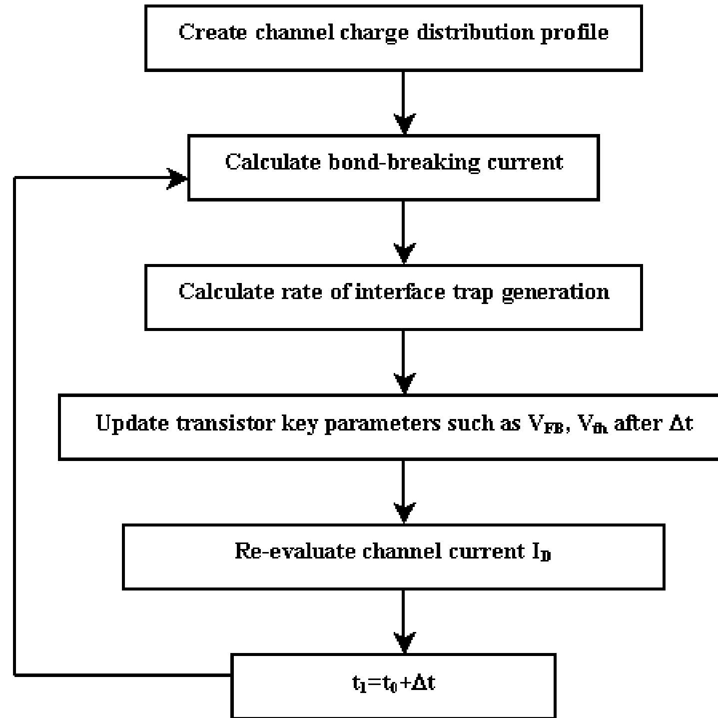


Figure 13. Modeling of hot-carrier.

The bond-breaking current as well as the channel charge density are calculated using Equation (22) to (24). The key transistor parameters such as flat-band voltage and threshold voltage are then calculated. The channel current is evaluated for different operation situations, in linear or saturation regions. Again, the transistor parameters are re-evaluated at every time step.

Figure 14 shows the simulated transistor drain current driving ability under hot-carrier effect with 1.7μm channel length and 8v drain voltage for about 14 hours. The continuous lines are simulation results using ARET, and the discrete points are the measured data published in literature [12].

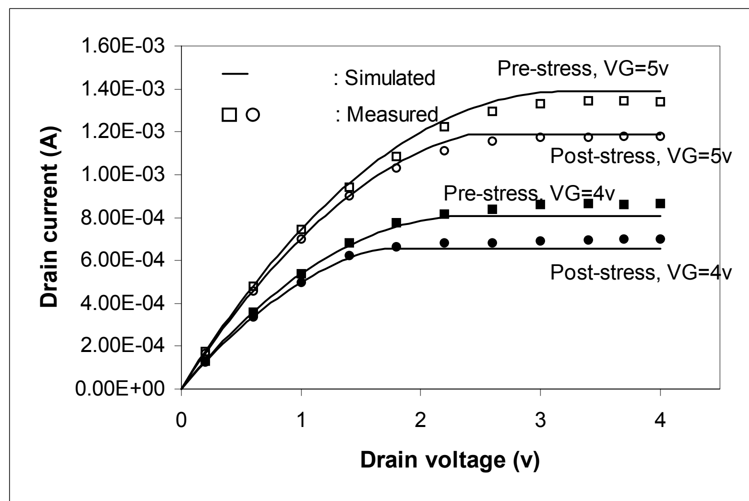


Figure 14. Drain current degradation of nMOS transistor under hot-carrier.

It can be observed that for both data sets under V_G of 4v and 5v, respectively, the apparent degradations in drain current are shown. This is due to the hot-carrier effect under high channel electric field. In addition, an excellent agreement between simulation and measurement is observed.

2.3. Gate oxide wear-out

2.3.1. Mechanism overview

As the feature size shrinks to submicron region with ultrathin gate oxide (t_{ox}<10 nm), the gate oxide wear-out has become a crucial reliability issue. The gate oxide wear-out, as well as the resulting time-dependent dielectric breakdown (TDDB) are the intrinsic reliability problems. It was first observed over three decades ago [30][31]. With the exact physics behind the mechanism remaining incompletely known, the basic mechanism is believed to be that the defects are created inside the oxide under certain driving forces, such as the oxide electric field, or the tunneling electrons through the ultrathin oxide. As the defects accumulate to a critical density, the defect paths are created through the oxide, which causes a sudden loss of the oxide dielectric property. The current surge is typically observed during this process leading to a permanent damage of the device [32].

2.3.2. Physical models

Two major modeling processes have been proposed in previous research for gate oxide reliability. The first model is known as the thermochemical model, or the E model [33]. It generally describes the electric filed dependence of the oxide wear-out. Based on this model, the weak Si-Si bonds are eventually broken by the oxide electric field creating charge traps in the oxide. As more electrons are trapped through the oxide, the final breakdown will happen. The time-to-breakdown of TDDB is proportional to the electric field in the form $t_{BD} \sim exp(-\gamma E)$, where E is the electric field and γ is the electric field acceleration factor. The second model is the so-called $1/E$ model, which was proposed based on anode hole injection and tunneling effect [34][35]. Based on this model, the Fowler-Nordheim tunneling takes place over the ultrathin oxide. The

tunneling electrons transfer the energy to the holes at the anode. These energized holes are then injected into the oxide and create the breakdown paths as the process continues. In this case, the time-to-breakdown is proportional to the reciprocal of the applied electric field, i.e., $t_{BD} \sim exp(\beta/E)$, where again β is the electric field acceleration factor.

Both E model and $1/E$ model have been under debates for the past years. The extrapolated data show that both models are very consistent under high electric fields (>10 Mv/cm) [32], while a significant discrepancy is observed at low electric field. This discrepancy can be a serious issue since the real ICs actually work in low electric fields. One of the reasons that this discrepancy has not been clarified is that it is very difficult to obtain the stress test data under low stress condition due to the cost and time involved.

In addition to the “hard” oxide breakdown, where an abrupt current surge is observed as the clear sign of device failure, the “soft” breakdown has been reported in ultrathin films [36]. Instead of the complete loss of dielectric property, what has been observed in soft breakdown is the slight change of voltage and current, accompanied by signal fluctuations. It is rather a degradation than a failure process. A variety of explanations exist trying to explain this phenomenon. In [36] it is explained by multiple tunneling events, while it is said to be the result of trap-trap transport of electrons in [37] and the process of dynamic trapping and detrapping in [38].

CHAPTER 3

CIRCUIT LEVEL RELIABILITY SIMULATION

Compared with component level parameters, circuit level specs are far more concerned in terms of product reliability. Circuit/system level reliability is the final reliability index of the product that interests customers, although it is the result of the component level degradation. Developing the proper circuit level algorithms for reliability simulations is critical for a circuit level reliability simulator.

3.1. Hierarchical reliability evaluation

3.1.1. Algorithm description

A hierarchical algorithm is proposed for simulating circuit level performance degradation [7], as demonstrated in Figure 15 where the circuit design is presented by various functional modules at different hierarchies.

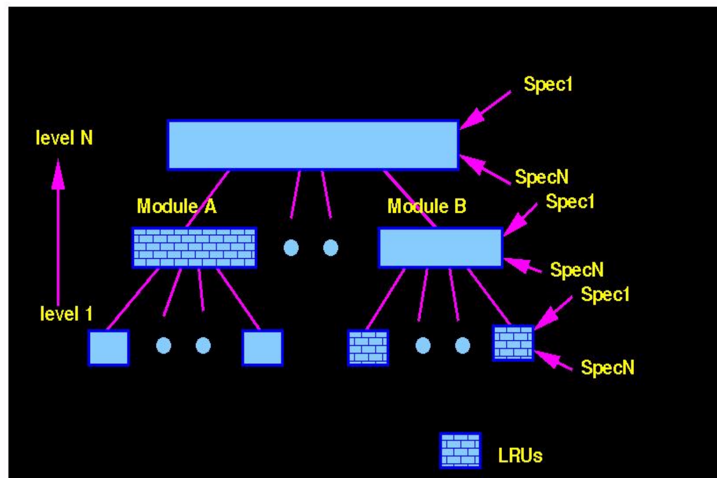


Figure 15. Hierarchical circuit reliability simulation.

At the highest level (level N), the complete circuit is described in terms of its sub-modules (A, B, etc. in the figure). The behavioral model for level N "calls" behavioral models for its sub-modules during simulation. The lowest level (level 1) consists of behavioral models of the circuit building-block components (interconnect traces, active devices, resistors, capacitors).

Given descriptions of signals that are applied to the input terminals of the highest-level modules (level N) during normal circuit operation or under stress condition, we determine the signals at the inputs to all the modules at the next level (level N-1), by circuit simulation using Spectre. This procedure is repeated in a "top-down" fashion to compute the current densities in each trace and voltage waveforms at every circuit node. From this information, the change in resistance of every interconnect in the circuit due to electromigration is computed. Similarly is the change in threshold voltage, etc., due to hot-carrier degradation in every transistor. From the basic physics-of-failure analyses, the changes in the corresponding electrical model parameters of modules at level 1 in Figure 15 are obtained as functions of time. As an example, the result of this analysis could be a set of mathematical functions that describes how the resistance of an interconnect trace or the transconductance/threshold voltage of a transistor changes with time due to electromigration and hot carrier degradation, respectively. If it is the case of an op-amp, by simulation it is possible to determine (from the knowledge of the way the electrical parameters of the interconnect and the transistor change with time) how the specifications of the op-amp change with time. Note that the above computation must take into account the fact that relative electrical stress values in different parts of the circuit change with time due to changing component performance. Hence, during

simulation, the behavioral simulation models must be updated periodically with new (degraded) values to maintain accuracy. A particular time-interval of simulation is selected so that the errors in the values of the node voltages and branch currents at the end of the time interval are less than a specified bound (this is similar to time-step selection in circuit simulation so that integration errors during transient simulation are minimized or bounded).

As discussed earlier, using simulation it is possible to determine how the behavioral model parameters of the embedding modules at level 2 (shown as spec1, spec2, … specN) are related to the behavioral model parameters of the modules at level 1. From the time-dependence of the behavioral model parameters of all the modules at level 1, the same is extracted for all the modules at level 2 of Figure 15 using hierarchical simulation. The analysis is performed hierarchically in a "bottom-up" manner to minimize overall simulation effort. Eventually, functions that describe how the high-level circuit specifications change with time are obtained. These functions are used to predict as accurately as possible the expected time at which the circuit is likely to fail due to electromigration and hot carrier degradations, where failure is defined as a condition in which the circuit no longer meets its original specifications.

3.1.2. Simulation examples

3.1.2.1. Two-stage op-amp

A simple two-stage op-amp is designed and laid out using AMI C5N process with a feature size of 0.5 μm with an Al interconnect width of 5 μm assumed. The stress condition is V_{DD}/V_{SS}=±3.5 v with T=300 °C for 100 hours. All components are subject to both EM and HC degradations. The schematic (left) and the simulation results (right) are

shown in Figure 16. The degradation of open-loop gain is simulated using the hierarchical simulation algorithm.

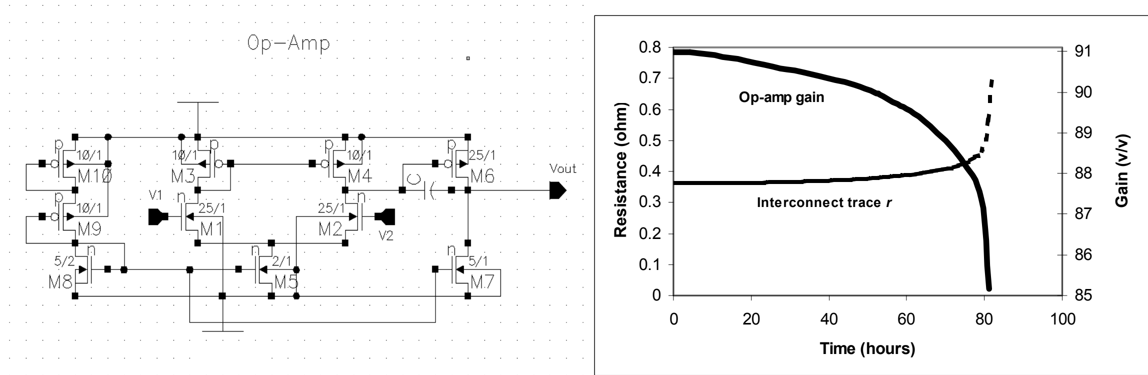


Figure 16. Degradation of two-stage op-amp.

Based on the reliability simulation, the open-loop gain of the op-amp drops from about 91 v/v at pre-stress condition to 88 v/v in 100 hours' stress under hot-carrier degradation, then completely fails due to the catastrophic break-off of interconnect wire r, which is the first wire broken under electromigration.

3.1.2.2. CMOS mixer

Hierarchical simulation on a CMOS mixer is demonstrated in Figure 17 (left – circuit schematic, right – simulation result), and again, the layout is assumed to be done using AMI C5N technology with 0.5μm feature size. The stress condition is V_G=3v, V_D=7v for 168 hours at room temperature, which is propagated down to every node involved by Spectre simulation. The degradation of the correlated gain is simulated using the hierarchical approach and all nMOS transistors are assumed to be exposed to hot-carrier degradation. From the simulation a clear change in the gain of the mixed under a sine

wave with a 10 mV magnitude and a 50 MHz frequency is observed after periodic stress, indicating the circuit level performance degradation under hot-carrier.

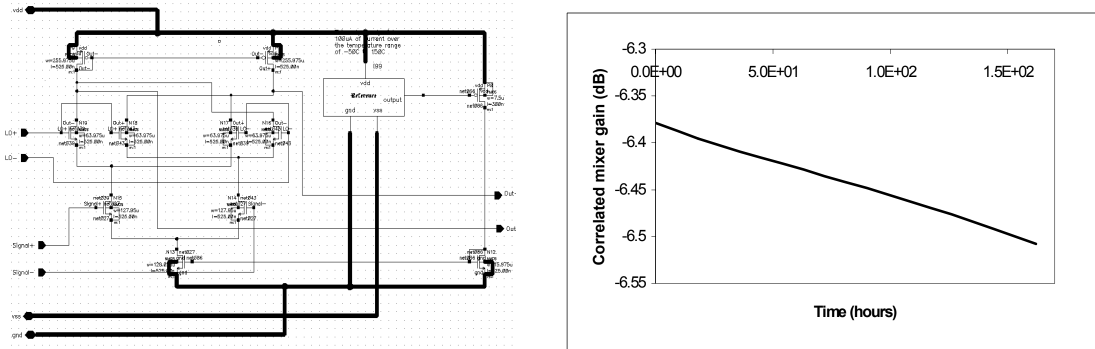


Figure 17. Degradation of CMOS mixer.

3.1.2.3. CMOS digital path

Another example is a CMOS digital logic path shown in Figure 18, which consists of a CMOS NAND gate and two inverters in series (left – circuit schematic, right – simulation result) and laid out using the same AMI technology used in previous examples. The major impact of failure mechanisms on digital circuit is the increasing switching delays to the point where the circuit fails performance specifications. Thus, the propagation delay becomes the critical spec being modeled in simulation.

In this example, all nMOS transistors were stressed with V_G=3v, V_D=7.3v under room temperature for 168 hours under hot-carrier degradation, and all interconnect traces were assumed to degrade under electromigration at the same time. While most of the device parameters were taken from AMI technology, the interconnect layer was assumed to be made in pure Al with half micron wide traces.

A 10ns clock interval was selected and it was found by simulation that the initial path delay was 6 ns. Figure 18 shows the simulation result for this path. The degradation of the path delay was simulated versus time and the circuit was predicted to fail in about 110 hours of stress. It was also shown by analyzing the simulation result that, the degradation of the interconnect did not make contribution to the overall circuit failure. This is because that before any catastrophic interconnect open failure happens, the quantitative change of interconnect is too small compared with total resistance of the circuit to make meaningful difference in overall performance. More discussion about interconnect degradation is given in the next section.

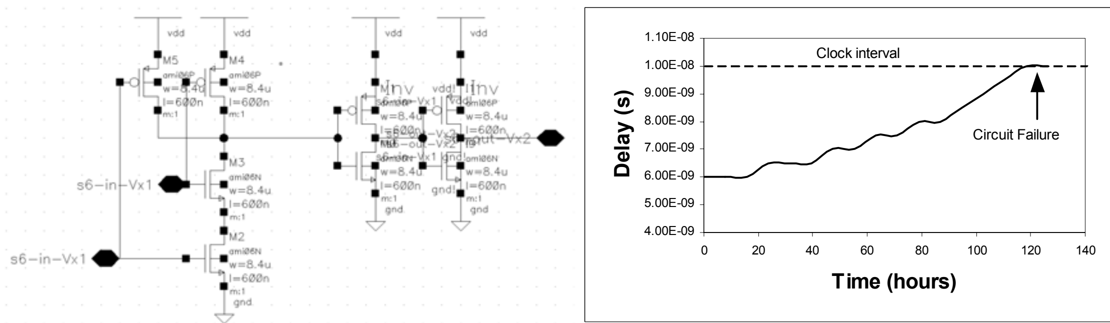


Figure 18. Degradation of CMOS logic path.

3.2. EM degradation modeling using defect statistics

For circuit level EM degradation, the defects generated during fabrication affect the overall interconnect reliability significantly. Due to the randomness of defect generation, the overall post-fab interconnect reliability (expected lifetime) at circuit level is modeled based on statistical process and probability theory.

3.2.1. Component level post-fab EM reliability

As discussed in previous sections, the EM physics-of-failure model has been extended to incorporate the post-fab physical defects. Using this upgraded EM model at component level, a pure aluminum interconnect trace is evaluated as following with a 1μm mean grain size assumed. The trace is 118μm long, 5μm wide and 2μm thick. Totally four cases are simulated: defect-free, 20%-defect, 50%-defect, and 80%-defect, where the 20%-defect means the defect size is 20% of the line width and so on. The base temperature of the simulation is 200°C and a constant 300mA current is being conducted through the interconnect trace. As the result of EM degradation, the percentage resistance change is simulated as a function of time. The simulation results are shown in Figure 19.

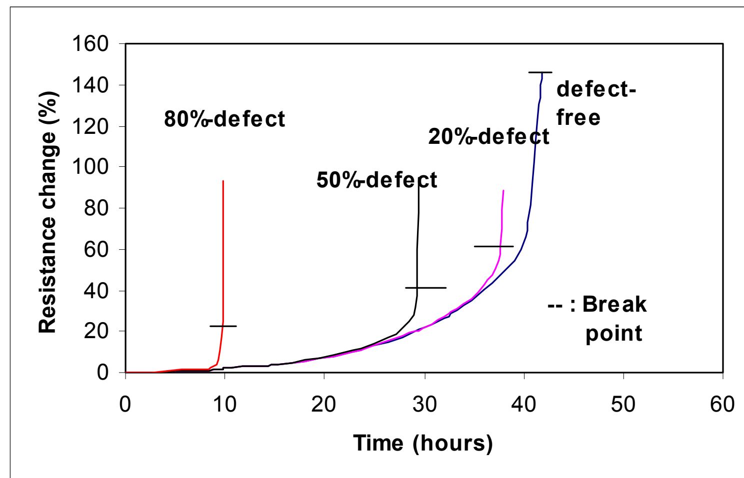


Figure 19. EM degradations of pure Al traces with different defects.

As it can be observed from Figure 19, the bigger the defect size, from 0% to 80%, the shorter the interconnect lifetime, and the smaller resistance change is observed before the final break-up. This phenomenon is due to the larger current density and the higher

local temperature at the defect site. This result is fully consistent with our analyses in previous sections.

Although very few experimental data in the scope of this topic has been reported due to the difficulty to obtain the defective traces and set up the stress tests, a set of data on three groups of defective metal stripes that clearly demonstrate this situation was found in literature [39]. The results prove that, if the size of physical defect is comparable to the line width, it will often be the major cause for final break-up. The detailed data are shown in Chapter 5.

3.2.2. Circuit level post-fab EM reliability

According to the simulation results using component level model, the resistance degradation of single interconnect line before the line is completely broken is about 10% ~ 50%. However, because the interconnect resistance itself is very small, this change is not supposed to make any meaningful contribution to overall circuit-level properties until the line is open. For a 300μm long and 5μm wide aluminum trace with a thickness of 2 μm, if it is defect-free the total resistance is around 0.81 ohm and the degradation can be as much as 0.4 ohm based on simulation, which makes the degraded total resistance 1.21 ohm, while in a circuit the resistances of passive components and devices are usually at least thousands of ohms. Thus, the interconnect resistance degradation due to EM will not be able to make any meaningful difference on the electrical stress distribution, such as the currents flowing through the interconnect lines. This leads to the conclusion that, before a interconnect line is completely open, the EM degradation process on the line is basically independent of the EM processes on other lines in the circuit, and the growth of one EM void is independent of other degradation sites on the same interconnect line as well. For

the same reason the circuit performance, such as the gain of an op-amp, will not be meaningfully affected until the final interconnect break-up comes out.

Based on the simulations in the previous section shown in Figure 19, the above conclusion indicates that an interconnect line will be most likely to break off due to the damage at the worst-case defect site, usually the biggest physical defect. The contributions of the other defects on the same line with sizes less than or same as the biggest one, and those on other interconnect lines, can simply be ignored. Further more, under the interconnect degradation the circuit will fail at the time that the first interconnect break-up happens. To prove and demonstrate the above conclusions, further simulations are conducted on a 118μm long, 5μm wide, and 2μm thick aluminum trace. The base temperature is 300 °C and the current is 50 mA DC. The results are given in Table 1, in which different defect combinations and corresponding lifetimes are listed.

Table 1. Predicted interconnect lifetimes for different defect conditions.

Defect condition	Predicted lifetime (hours)
One 1μm defect	22.3
One 3μm defect	10.05
One 1μm defect + one 3μm defect	10.10
Two 3μm defect	10.27
One 2.5μm defect + two 3μm defect	10.94

In Table 1, as the defect size increases from 1 μm to 3 μm, the lifetime (time to open) decreases by more than 50%. However, as long as the 3μm defect stays as the

biggest defect, any combination with other defect sizes does not contribute to the overall lifetime significantly. It can be concluded that the final break-off is due to the 3μm defect growing almost independently.

For the case that various interconnect lines are involved, an op-amp circuit is simulated. Three 5μm wide interconnect traces in the circuit, r1 with a 3μm defect, r2 with a 4μm defect, and r3 with a 3.5μm defect are involved. The simulation results for interconnect r2 with different involvement situations are shown in Table 2.

Table 2. Predicted lifetimes of interconnect r2 in op-amp circuit.

Interconnect lines involved	Predicted lifetime of r2 (hours)
r2 only	81.53
r2 and r1	81.54
r2, r1 and r3	81.54

It can be seen that the EM degradations of the interconnect traces other than r2 does not have impact on the lifetime of r2, because the absolute values of any interconnect resistance degradation is too small to affect the electrical stress conditions at other interconnect lines in the circuit.

From the circuit performance perspective under interconnect degradation, the circuit will fail when any of its interconnect lines becomes open due to EM. This is demonstrated by the simulation results shown in Figure 20. In the figure, interconnects r1, r2, and r3 are simulated separately first. The op-amp (gain) is then simulated with all three interconnects degrading simultaneously. It can be seen that the op-amp fails at

almost the same time that interconnect r2 breaks off, although r1 and r3 have not reached their failure points. The simulations also show that, before the line is open and the op-amp fails, the degradation of the gain due to the interconnect degradations is so small, from 90.971 to 90.968, that it can be completely ignored.

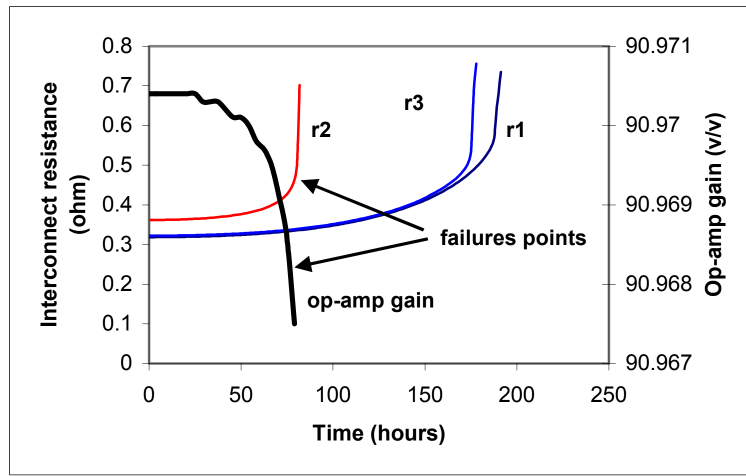


Figure 20. EM degradations of op-amp specs.

3.2.3. Lifetime prediction under post-fab EM degradation

3.2.3.1. Defect size distribution and relative probability

Due to the close dependence on uncontrollable process parameters, it is very difficult to generate a complete physical model for local defects such as the photolithographic defects. However, in this work it has been shown that modeling can be accomplished based on a statistical process.

Due to the significantly different roles that defects with different sizes play in the interconnect EM degradation process, the defect size distribution and relative occurrence must be determined. A number of people at IBM have made experimental effort to

determine this distribution. In G. F. Guhman's work at IBM, Burlington, defects in memory chips were counted using optical microscope and the relative occurrences were recorded. A mathematical function then must be generated based on these data describing the defect size distribution *D(x)*, where *x* represents the defect size. This distribution can be related to the relative probability density function *pdf(x)* by

$$D(x) = \overline{D} pdf(x) \tag{30}$$

where $\overline{D}$ is the average defect density. In C. H. Stapper's work [40], a normalized defect probability distribution was given by

$$pdf(x) = \frac{2(n-1)x}{(n+1)x_0^{\ 2}} \qquad \text{for} \qquad 0 \le x \le x_0$$

$$pdf(x) = \frac{2(n-1)x_0^{\ n-1}}{(n+1)x^n} \qquad \text{for} \qquad x_0 \le x \le \infty \tag{31}$$

where x_0 is a process-dependent fitting parameter and can be determined from measurement as [41]

$$x_0 = \frac{3M(x)}{4} \tag{32}$$

where *M(x)* is the mean of the measured defect size distribution. Thus the relative probability distribution can be described as in Figure 21.

This distribution function has been frequently used in IC yield models. Test results have shown that n=3 gives an excellent fit to the measurements, especially for metal defects [40]. This value will be used throughout this work giving the probability distribution as

$$pdf(x) = x / x_0^2 \qquad \text{for} \qquad 0 \le x \le x_0$$

$$pdf(x) = x_0^2 / x^3 \qquad \text{for} \qquad x_0 \le x \le \infty \tag{33}$$

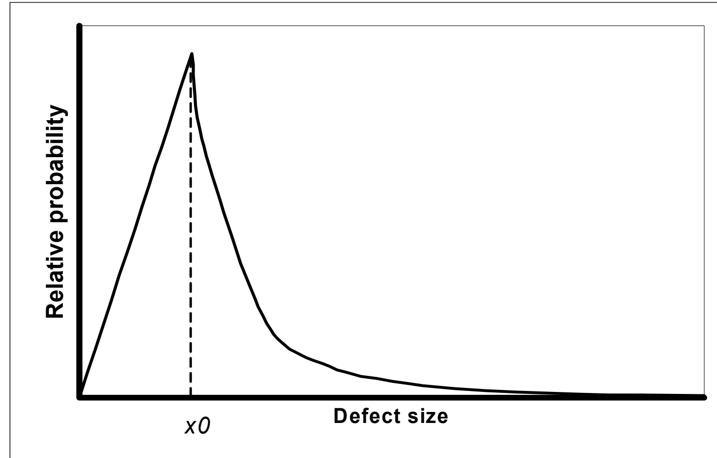


Figure 21. Defect probability distribution.

3.2.3.2. Probability of presence of defect on interconnect

In order to determine the probabilities that defects with different sizes are presented on certain interconnect, the continuous relative probability distribution has to be quantized at some key defect sizes, which result in significantly different interconnect lifetimes. The ICs under reliability evaluation are supposed to pass the wafer test and burn-in test, in which the global and serious local defects are detected and screened out. Therefore the defects in this stage are all potential failure seeds with relatively smaller sizes. Also, since the so-called "bamboo structure" has totally different failure mechanism from the other defect sizes, this case has to be considered separately. Thus a series of *n* key defect sizes are generated as following

$$0 < x_1 < x_2 < x_3 < \cdots < x_n < x_{BB} \tag{34}$$

where x_{BB} represents the defect size causing “bamboo structure” and x_n is the biggest defect size possibly presented under certain situation. It must then be normalized giving the relative probability of a key defect i to be presented on the interconnect line as following.

$$\Pr'\{x_i\} = \frac{\int_{x_{i-1}}^{x_i} pdf\,(x)dx}{\int_0^{x_n} pdf\,(x)dx} \qquad 1 \le i \le n \tag{35}$$

By substituting Equation (33) into (35), the relative probability becomes

$$\Pr'\{x_i\} = \frac{\int_{x_{i-1}}^{x_i} pdf\,(x)dx}{1 - \frac{{x_0}^2}{2{x_n}^2}} \tag{36}$$

Here x_0 is the process-dependent parameter and can be determined from Equation (32). If we know the defect line density D_L, then the true probability that at least one defect i is presented on the interconnect line is

$$\Pr\{x_i\} = 1 - [1 - \Pr'\{x_i\}]^{D_L L} \tag{37}$$

where L is the interconnect length and $D_L L$ is a positive integer [10]. For the case that $D_L L$ is not an integer, the following discussion is employed to prove that Equation (37) still stands correct.

For the case $0 < D_L L < 1$, from the perspective of probability theory, the probability of defect i presented on this line L can be described as $D_L L \Pr'\{x_i\}$. In Taylor’s expansion, the following series exists

$$(1-x)^m = 1 - mx + \frac{m(m-1)}{2!}x^2 - \frac{m(m-1)(m-2)}{3!}x^3 + \cdots \tag{38}$$

By using Taylor's series and ignoring the higher-order items, Equation (37) can then be re-written as

$$\Pr\{x_i\} \approx 1 - [1 - D_L L \Pr'\{x_i\}] = D_L L \Pr'\{x_i\} \tag{39}$$

which is simply the probability derived from probability theory when $0 < D_L L < 1$.

3.2.3.3. Model implementation based on defect probability

A probability model for circuit-level reliability evaluation can thus be developed based on the following conclusions drawn from the analyses and simulation results in previous sections:

- The fabrication-induced physical defects are responsible for an absolute majority of the final EM failures of interconnect. Without considering the special cases such as the "bamboo structure", as the on line defect size increases the statistically expected interconnect lifetime decreases.
- The growth of interconnect defect due to EM can be approximately considered to be independent of other possible defects on the same line and those on other interconnect lines in the circuit.
- Since the interconnect resistance change caused by EM degradation is negligible compared with other resistances in the circuit, the circuit-level specifications do not show noticeable degradations until the time that the first interconnect failure happens.

For any interconnect line subject to EM damage in a given circuit, first the expected interconnect lifetime (time to open) under circuit operating condition is evaluated for every possible presence of *n* key size defects, respectively, including the defect-free case. These key defect sizes are selected based on their significant impacts on interconnect

lifetime. Then the corresponding defect probability is calculated using Equation (36)&(37) with given process information such as mean defect line density. Given the fact that with the presence of the bigger defect the smaller ones do not have meaningful contribution to the trace lifetime, the next step is, starting from the biggest defect, i=n, to calculate the trace lifetime t_i when defect i is the biggest defect presented, as well as the corresponding probability $\Pr\{t_i\}$. Taking the same process to i=n-1 until i=0, which means the interconnect line is defect-free, all possible lifetimes of this specific interconnect line as well as their probabilities can thus be obtained. NOTE, since the defect i only works when defect i+1 is NOT presented, the probability that the line shows a lifetime t_i has to be scaled giving the following equation.

$$\Pr\{t_i\} = (1 - \Pr\{x_n\})(1 - \Pr\{x_{n-1}\})...(1 - \Pr\{x_{i+1}\})\Pr\{x_i\} \tag{40}$$

By assuming totally m interconnect lines are involved, repeating this process at every line in the circuit, all possible circuit lifetimes due to interconnect failure and their probabilities are obtained in the form of t_{ij} and $\Pr\{t_{ij}\}$ with i=1 … m and j=1 … n, where j=n represents the biggest quantized defect.

From the circuit perspective, if only interconnect failures are concerned the circuit will fail upon the first line broken and the lifetime of the first-broken line indeed represents the lifetime of the circuit due to interconnect failure. It also indicates that the circuit will end up with the lifetime of a interconnect line only when all other cases that could show a shorter interconnect lifetime do not happen. In other words, a interconnect line with certain defect combination can make the circuit fail only when it is the first line to open. Thus, the same argument and scaling process for single interconnect line applies here as well.

To proceed, all possible interconnect lifetimes t_{ij} are ranked from the shortest to the longest, in form of t_L_k with k=1 … $m \times n$, along with their corresponding probabilities $\Pr\{t_L_k\}$. Besides, a nominal circuit lifetime is defined here specifically for the defect-free case at circuit-level

$$t_L_0 = \text{min (defect-free interconnect trace lifetimes)} \tag{41}$$

which actually represents the longest circuit interconnect lifetime that we can expect. The expected circuit lifetime due to interconnect failure is then described as follow.

$$t_C = (1-\Pr\{t_L_1\})(1-\Pr\{t_L_2\})...(1-\Pr\{t_L_{m\times n}\})t_L_0 + \sum_{k=1}^{m\times n}\Pr\{t_C_k\}t_L_k \tag{42}$$

with scaled probability at circuit-level

$$\Pr\{t_C_k\} = (1-\Pr\{t_L_1\})(1-\Pr\{t_L_2\})...(1-\Pr\{t_L_{k-1}\})\Pr\{t_L_k\} \tag{43}$$

and the circuit reliability function is obtained similarly as

$$R(t) = (1-\Pr\{t_L_1\})(1-\Pr\{t_L_2\})...(1-\Pr\{t_L_t\}) \tag{44}$$

where $\Pr\{t_L_t\}$ represents the probability of the defect/line combination that shows a lifetime around $t_L=t$.

This model is developed based on probability theory and statistical data, due to the uncertainty and randomness of the post-fab defects. All results are expected values over a large amount of data.

The model has been fully realized and integrated in ARET [10]. The following is an example to demonstrate the application of this model and all results are obtained from simulations using ARET.

3.2.3.4. An example of post-fab circuit lifetime prediction

In this example an analog op-amp is used as the circuit under EM reliability evaluation just to demonstrate the probability EM model. The schematic is shown in Figure 22. For simplicity, three of its interconnect lines, r1, r2, and r3, are selected as the group subject to EM damage. Among these interconnect traces, r1, r2 are at output and r3 is at input of the circuit, and the rest of the circuit is assumed to be EM damage-free.

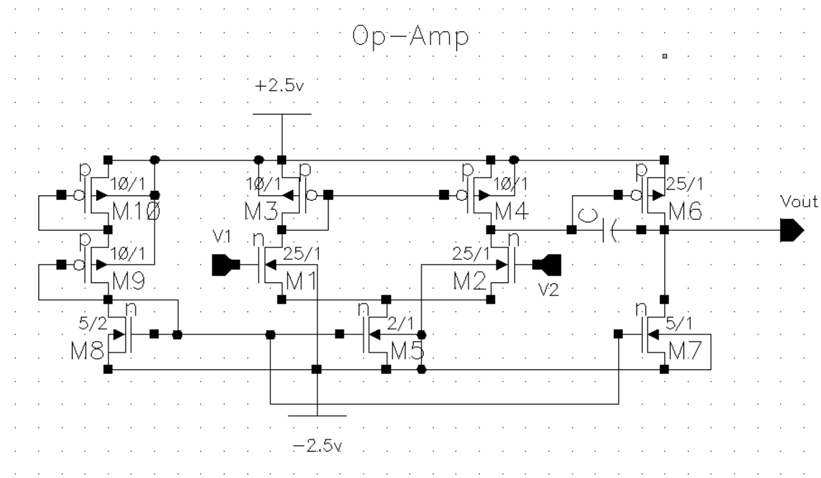


Figure 22. Two-stage analog op-amp.

Based on the experiments conducted by Z. Stamenkovic *et al* [41], a value of 0.01/μm is taken as the mean defect line density. All three interconnects are assumed to be pure aluminum traces having same geometries: 118μm long, 1μm wide, and 1μm thick. The temperature is 300 °C. Also, based on pre-simulations using ARET, defect sizes 0.3μm, 0.6μm, and 0.8μm are shown as the most significant key sizes in terms of interconnect lifetime.

The evaluation results are listed in Table 3. All involved interconnect lines with possible defects are listed in 1st and 2nd column. The corresponding defect probabilities are calculated using Equation (37) and the interconnect lifetimes are obtained from simulations. The overall interconnect lifetime is then predicted by Equation (42).

This result shows that at the time around 144.77 hours, the circuit will be most likely to fail due to interconnect break-off. It is also noticed that the nominal circuit lifetime is 168.68 hours, which is shorter than the lifetime of line r1 with 0.3μm defect presented (171.59 hours). This means that such a defect condition will almost never be responsible for a circuit failure and thus can be excluded from the evaluation process.

This op-amp circuit design has been fabricated as one of the test structures for calibrating ARET with accelerated stress tests in Chapter 5.

Table 3. Interconnect lifetime prediction of op-amp.

Defect type		Probability	Lifetime (hours)
r1	0.3µm	0.9364	171.59
	0.6µm	0.0887	108.68
	0.8µm	0.0248	56.176
	Defect-free	0.0564	172.84
r2	0.3µm	0.9364	167.01
	0.6µm	0.0887	104.92
	0.8µm	0.0248	54.093
	Defect-free	0.0564	168.68
r3	0.3µm	0.9364	166.59
	0.6µm	0.0887	104.92
	0.8µm	0.0248	54.093
	Defect-free	0.0564	168.68
Expected circuit lifetime (hours)	144.77		

CHAPTER 4

ARET – ASIC RELIABILITY EVALUATION TOOL

With all these physics-of-failure models of major failure mechanisms and the circuit level simulation algorithms created, a CAD tool has become the logic and technical follow-up to manage the modules accomplishing various reliability evaluations, and to supply a friendly user operation interface.

4.1. Tool overview

ARET is an IC reliability simulation tool. It was developed to integrate all the physics-of-failure models of major failure mechanisms, as well as the circuit level simulation modules discussed thus far. Compared to other reliability simulators, ARET focuses on circuit level reliability simulation, and makes effort to work with post-fab ICs. In addition, ARET is able to identify the reliability hotspot(s), which is a crucial step in circuit local design for reliability. A diagram in Figure 23 clearly shows the major functions currently contained in ARET tool.

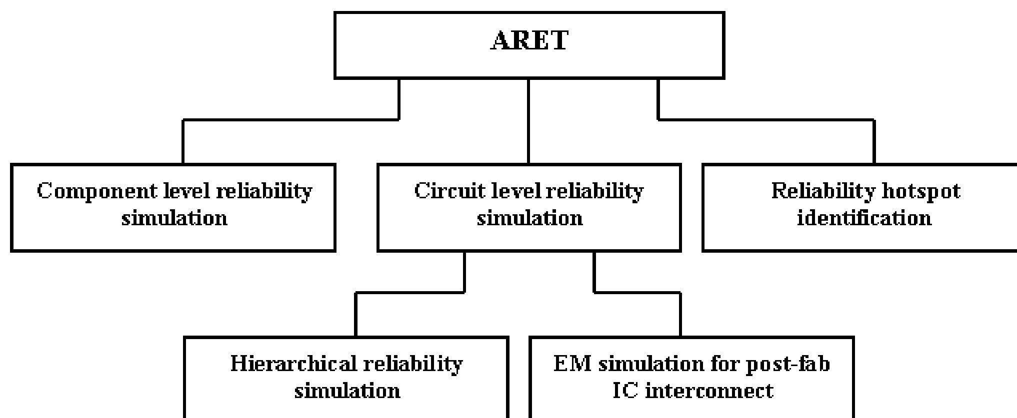


Figure 23. ARET functions.

The models for component and circuit level simulation functions have been discussed in detail in previous two chapters. In this section, the algorithms and the execution flows for these functions are presented and discussed, respectively. All circuit reliability simulation functions are explained in the next section, and the hotspot identification function is explained in this chapter as well as in Chapter 6.

The function modules in ARET were written in C while the GUI was developed using Tcl/TK. A high level view of the tool is given in Figure 24.

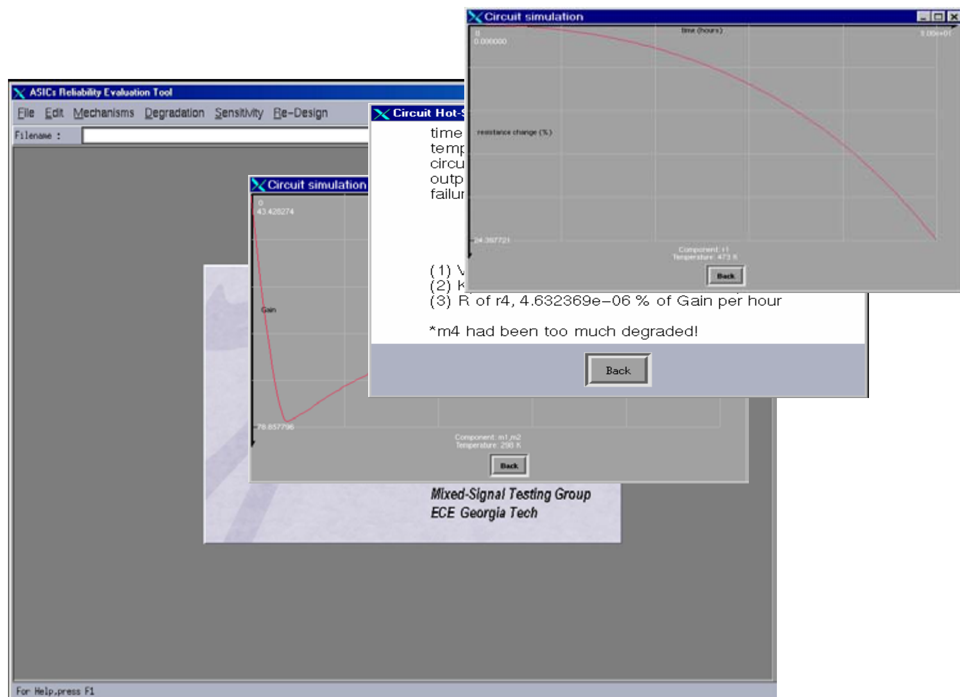


Figure 24. ARET at a glance.

For now, the tool can simulate IC degradations at component level, such as interconnect and transistors, as well as at circuit level, under hot-carrier and electromigration. Besides, currently the degradations of post-fab ICs under EM are

handled. The reliability hotspot identification function is integrated in the tool giving a list of hotspot components. The development of the tool was supported by the U.S. Air Force Research Lab and the Northrop Grumman Corp., and the tool has been well calibrated by a series of stress tests conducted at The Boeing Company at Seattle. A complete operation guide is given in Appendix A.

4.2. Reliability simulation function

Based on physics-of-failure models, differential equations are set up at circuit components to calculate the degradations at those nodes. At the beginning of simulation, circuit components that are subject to degradation are identified, and the stress factors at these components are produced hierarchically by circuit simulations using Cadence Spectre. This is done under either normal use or accelerated stress conditions. The obtained stress factors are then used as the input parameters of the degradation differential equations, of which the outputs are the performance degradations of the components, such as threshold voltage increases of transistors, at the end of that time step. Updating the key parameters for these components, the original circuit description is thus modified. A following Spectre simulation cycle then takes all these changes to circuit level specs, level by level, completing a single reliability simulation cycle for current time step. This basic process is demonstrated in the flow chart shown in Figure 25. The process is repeated until the specified simulation time is reached. In case a failure criterion is given in form of selected circuit specs, the circuit time to failure, TTF, can then be predicted.

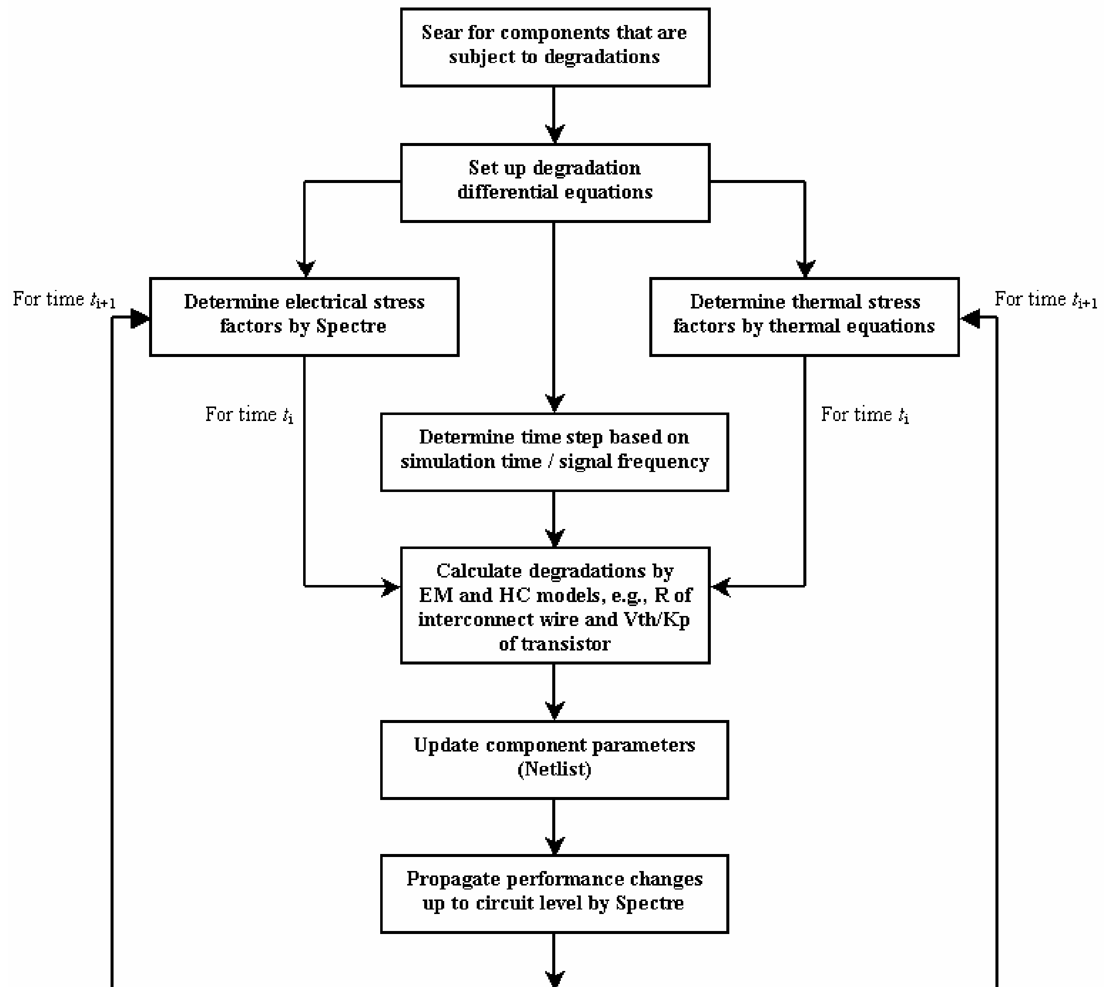


Figure 25. Hierarchical reliability simulation with ARET.

The biggest issue in reliability simulation algorithm is the lengthy simulation time. Unlike circuit simulation, reliability simulation often simulates the circuit in a long period of time, even in the lifetime. Furthermore, ARET is designed to be able to monitor circuit specs even under normal operating condition, which gives much longer circuit lifetime than under stress condition. Therefore, the selection of simulation time step used in regular circuit simulator will actually kill the whole process due to the amount of computation involved in reliability simulation, especially when the signal frequency goes high. In ARET, two levels of time stepping are set up.

At the first level, the whole simulation period is divided to sub-periods. These sub-periods can be quite large giving the fact that the circuit performance is degrading very slowly in a long term. Two options determining the sub-periods are given. The first option is to specify the fixed period manually in the database before the simulation begins. This is a flexible and very time-efficient approach for experienced users who have a good understanding of the circuit and have the control to select the proper stress condition. However, since the stress condition at every node keeps changing during the reliability simulation, the potential risk with this approach could be the poor accuracy if the periods are not small enough. The second approach is to obtain the sub-periods dynamically. This is similar to the time step control in circuit simulators like SPICE. For an existing period T with overall stress condition S_0, the performance is updated to P_1 at the end of the period, with the stress condition updated to S_1 too. An averaged stress $S_a = (S_0 + S_1)/2$ is then used for this period resulting in an updated performance P_1'. The difference between P_1 and P_1' serves as the measure of truncation error. If it is too large the existing period will be tightened.

At the second level of time stepping, an approach called “stress equivalency” is used within the sub-periods, explained as follows. On the contrary to circuit simulations, the detailed information about the signal such as frequency and trajectory is not important in most cases in reliability simulations. Instead, the accumulative stress level to the simulation point is the major concern. Thus, an equivalent stress of the actual stress signal with much lower frequency can be used for the reliability simulation of analog circuits, as demonstrated in Figure 26.

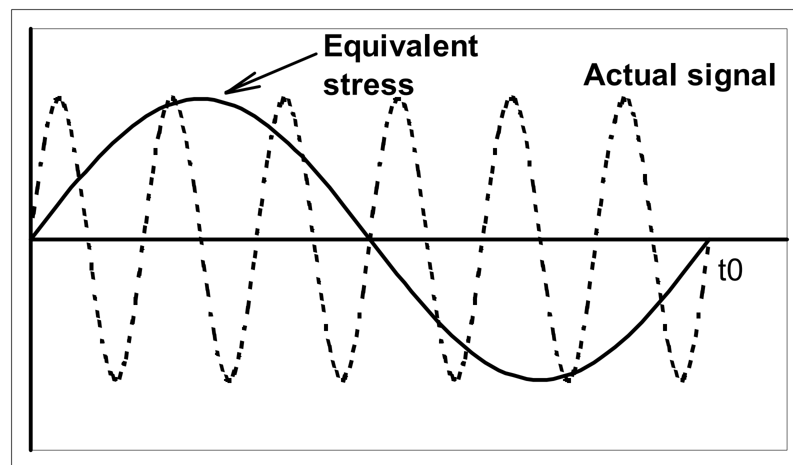


Figure 26. Equivalent stress.

In Figure 26 the first simulation point is at time t_0. Before t_0, the actual stress signal at the selected circuit component has a frequency $f_s=6/t_0$. Since for the reliability simulation of analog circuits it is the accumulative stress level driving the degradation process, an equivalent signal with the same amplitude as the actual signal but with the period of t_0 is used to approximate the actual stress condition. Thus, the frequency of this equivalent stress signal f_a is six times lower than the actual signal, which still gives the

same level of accumulative stress. Because that, based on sampling theory, to maintain the same simulation accuracy, the higher the signal frequency, the higher the sampling frequency will have to be, which requires more simulation points/cycles. By stress equivalency, the frequency of the stress signal used in simulation can be extremely low and the time step within a sub-period thus can be very large, resulting in a very small number of necessary simulation points/cycles. This significantly reduces the time required for reliability simulation over a long time span.

The stress equivalency approach works for periodic signals at the degradation nodes. As far as the stress level is concerned, it gives the best accuracy of approximation for square waveform in analog circuits. However, for CMOS digital circuits, since the degradations happen mainly during signal transitions, this approximation approach is no longer valid. In ARET, both time stepping schemes are used for different circuit situations.

When dealing with post-fab ICs under EM, a different simulation process is performed using the probability model discussed in Chapter 3. The possible effective defect sizes are quatized first based on previous results. Here it is assumed that a burn-in test is to be conducted before reliability evaluation. Thus, the maximum defect size is only about 50% of trace width. The nominal circuit lifetime is defined as the interconnect lifetime of the defect-free circuit, obtained only from simulations. The simulation flow for post-fab ICs under EM is given in Figure 27.

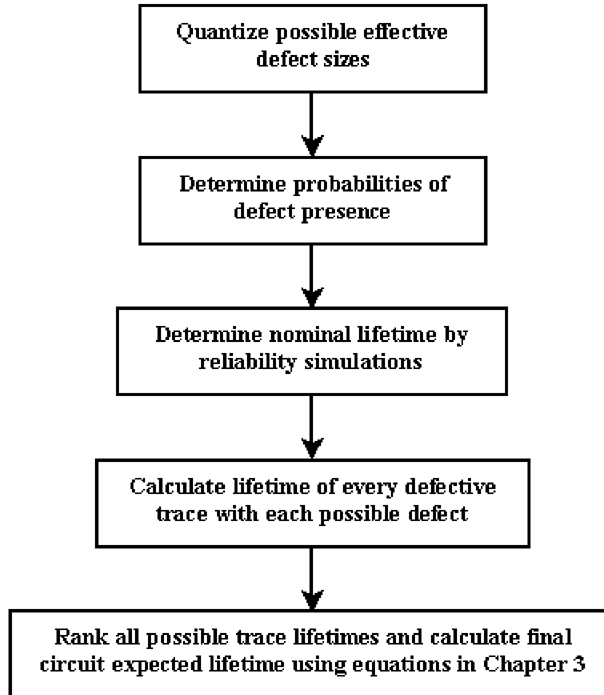


Figure 27. Lifetime prediction for post-fab IC interconnect under EM.

4.3. Reliability hotspot identification function

A reliability hotspot is defined as the circuit component that is most likely to cause the circuit fail under certain failure criterion, in other words, the component that is most likely to fail first failing the whole circuit. Identifying such reliability hotspots offers the opportunity to improve overall circuit reliability by locally redesigning the hotspot components [42].

ARET implements the hotspot identification in various ways for different failure/degradation situations by conducting reliability simulations. For CMOS digital circuits, the most critical specification is the speed, or the propagation delay. Under gradual degradation mechanisms such as hot-carrier, the reliability critical path that exceeds the design-specified maximum propagation delay first is located, and the hotspot

gate is identified as the gate that contributes the most to the delay increment. Under catastrophic failures such as gate oxide breakdown as well as the interconnect open failure, the first device/trace exhibiting breakdown is identified as a reliability hotspot. For analog/mixed-signal circuits, hotspot is identified similarly as the component contributing the most to the degradation of circuit key spec(s), which are chosen to evaluate circuit performance and judge circuit failure.

ARET identifies reliability hotspots by giving a list of up to three such components. These hotspots are extremely useful for conducting design-for-reliability, as it makes possible to zoom into the huge scale circuit and only focus on the areas that contains the reliability hotspots [42].

A typical output of reliability hotspot identification from ARET tool is given in Figure 28. Specific information about how to identify the reliability hotspot is presented in section 6.1.

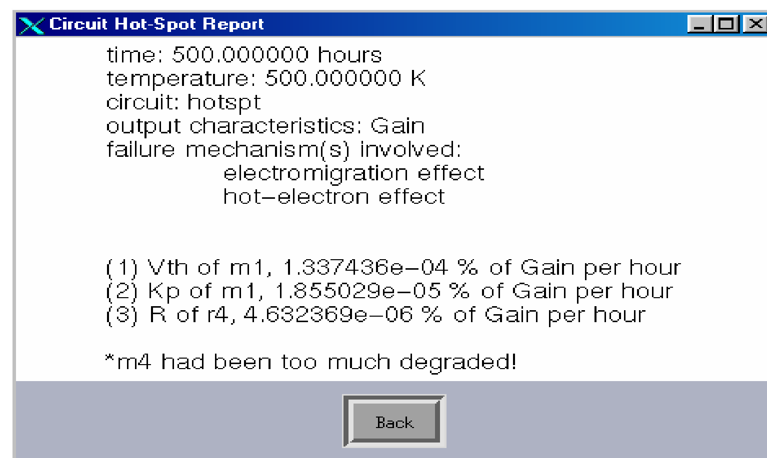


Figure 28. Reliability hotspot identification in ARET.

CHAPTER 5

ARET CALIBRATION

CAD tools need to be verified/calibrated before use. Generally, experimental data (not simulation data) are required to fulfill this requirement. However, for reliability simulators, this has been proved to be very difficult. Besides the test structure design and fabrication, the means to accelerate the experiments are extremely expensive including electrical and thermal accelerations in specially deigned environmental chambers and laboratories. In addition, even with dramatic accelerations, the so-called stress test is still lengthy, usually days to months. All these make reliability stress test a costly "must-do".

5.1. Test structures

Three categories of test structures were designed and fabricated with AMI C5N technology for calibrating EM component-level models, HC component-level models, and verifying the circuit-level simulation algorithms, respectively. The EM test structures are basically the metal traces with various geometries and shapes, which were fabricated in two metal layers. The HC structures are all single nMOS transistors with different dimensions, among which the shortest channel length was the feature size of AMI C5N process, 0.5 μm. The test structure circuits to verify the circuit-level simulation algorithms include an analog op-amp and a CMOS digital inverter. The detailed information about the test structures is given in the following sections. The final layout and the picture of the package (LCC) are shown in Figure 29. All circuits were fabricated using AMI C5N technology by MOSIS.

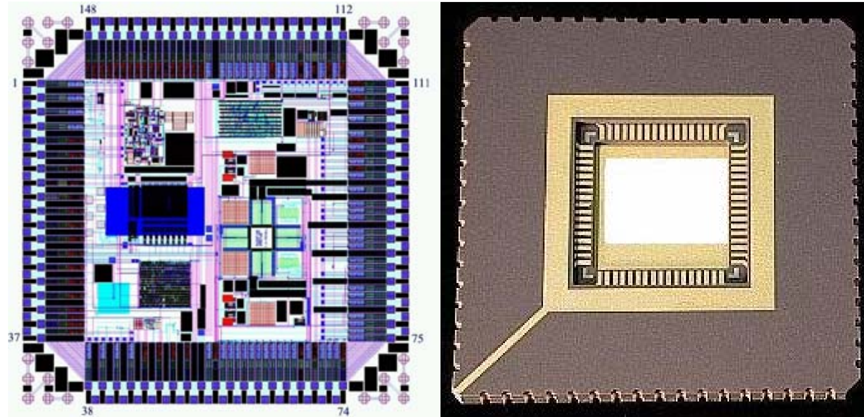


Figure 29. Final layout and package of test structures.

5.1.1. EM test structures

Figure 30 shows the straight metal traces designed to measure the mass depletion in the basic EM test models. The minimum trace width is 1 μm. The end structure geometry was assigned for Kelvin measurement. Those geometrical parameters can be modified accordingly based on the certain IC process. The structures were fabricated in layer metal 1 and the layer parameter can be modified to accommodate different process technology files. Te resistance of the trace was measured to demonstrate the EM degradation.

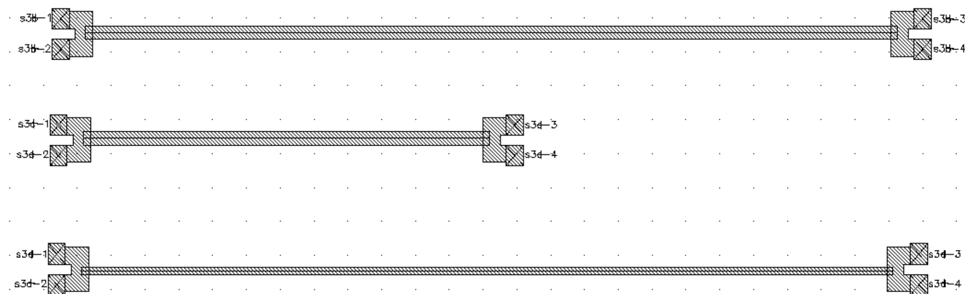


Figure 30. Straight traces.

The test structure shown in Figure 31 is used to measure the EM effect of a series of corners, which is supposed to make the EM degradation worse. The same layer and end structures as the straight traces were used.

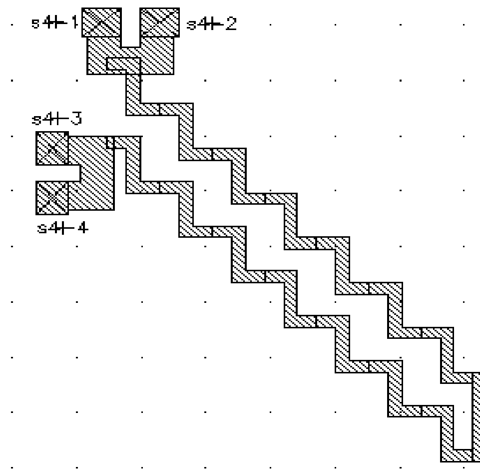


Figure 31. Corner structure.

Figure 32 is a spiral structure. Besides the corner effect, the eddy current under high frequency may take effect inside the structure.

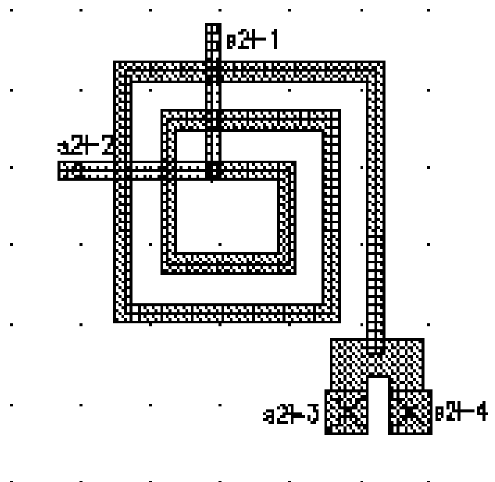


Figure 32. Spiral structure.

5.1.2. HC test structures

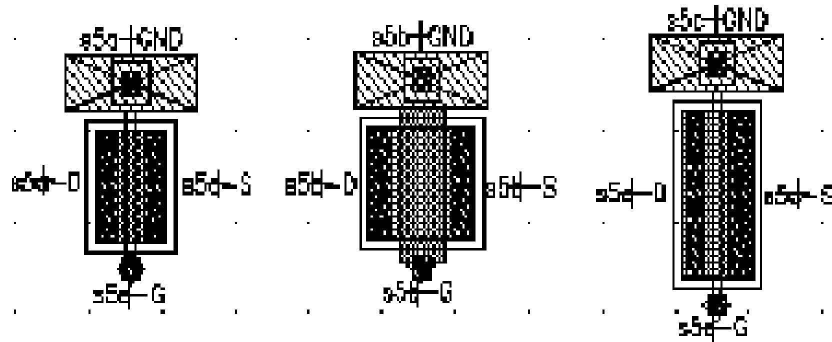


Figure 33. nMOS transistors by AMI C5N process.

The test structures shown in Figure 33 include a set of nMOS transistors with different channel lengths and widths by AMI C5N technology. The minimum channel length is 0.6μm. They are designed to show the transistor performance degradation due to the hot-carrier effect by the measurement of critical parameters. The saturated drain current and the threshold voltage were chosen to be measured. The electrical stress level is the key condition to control the experimental process. The determination of these experimental parameters is discussed later in this chapter.

5.1.3. Test structures for circuit level simulation algorithms

Two test structure circuits were designed to calibrate the circuit level simulation algorithms in ARET. They are a CMOS inverter and an analog op-amp, as in Figure 34 and Figure 35, respectively. Both were fabricated by AMI C5N process.

For the digital inverter, the logic 1 noise margin was to be measured and the hot-carrier degradation was the only failure mechanism considered. For the op-amp, the open-loop gain and the CMRR were to be measured under hot-carrier degradation.

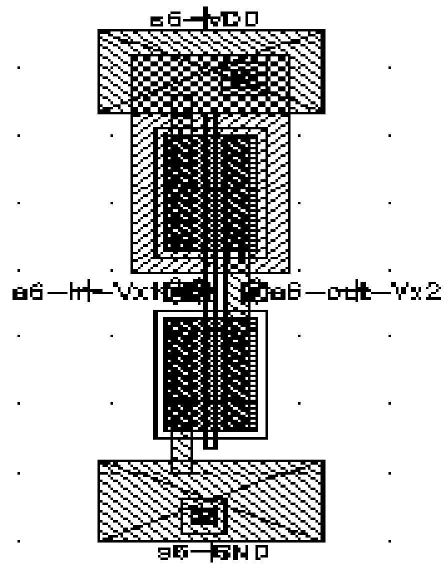


Figure 34. CMOS inverter.

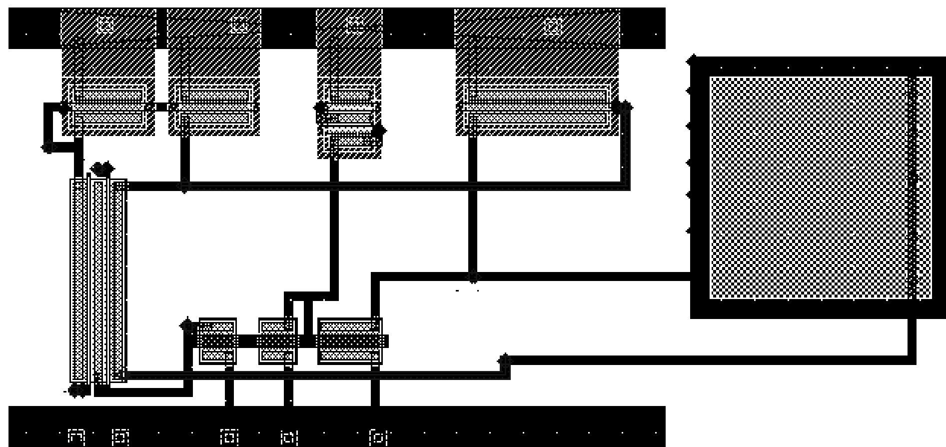


Figure 35. Two-stage op-amp.

5.2. Stress tests

In order to collect data from the test structures calibrating ARET, a set of stress tests were designed and conducted at The Boeing Company, Seattle, WA. Two environmental chambers were set up, one for EM structures and the other for HC/circuit level structures. The chambers were set to different temperatures. A picture of the chamber and the corresponding temperature profiles are displayed in Figure 36.

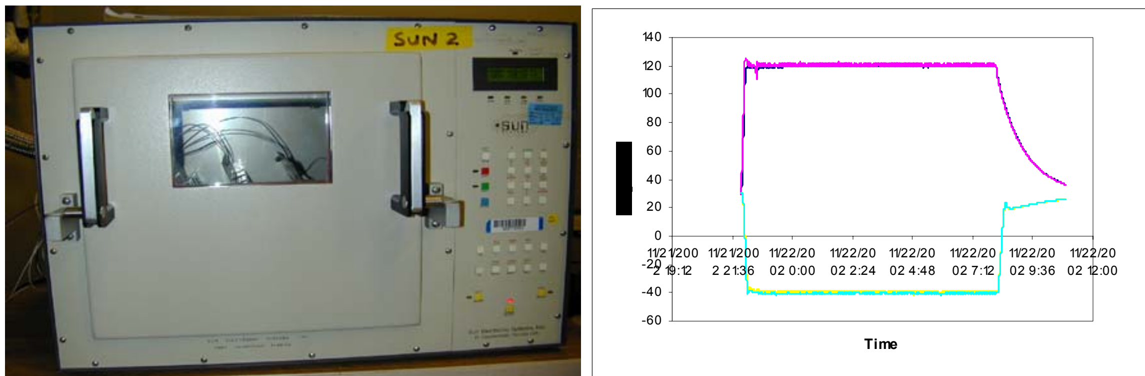


Figure 36. Environmental chamber and temperature profiles.

All packages were stressed at the same time in the chambers. To ensure reliable contacts for the packages under stress, a contactor board was designed with all burn-in sockets, as shown in Figure 37, to hold the packages securely and to supply the proper electrical contacts.

During each stress/test cycle, packages were loaded into the sockets on the contactor board. The boards were then installed in the slots inside the chambers. After the chambers were closed, the electrical connections were supplied through the special connectors on the chambers to the pads on the boards.

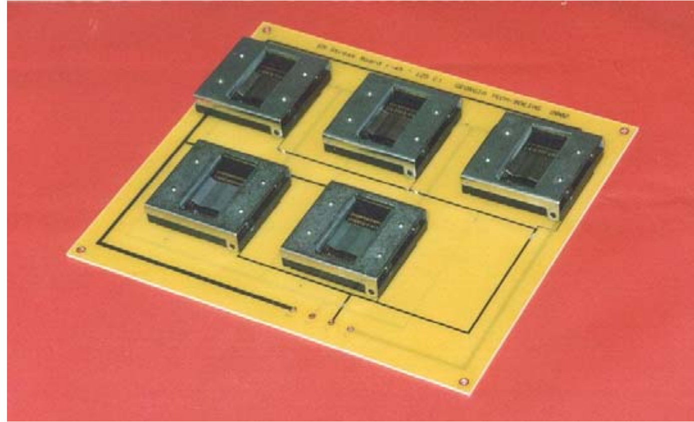


Figure 37. Stress contactor board.

At the end of each stress cycle, packages were retrieved and loaded into a test interface board, as shown in Figure 38. This board was designed to supply the proper electrical and mechanical interface between the tester and the device under test (DUT). The board also contains the test circuitry discussed later in this section. The board uses specially designed burn-in sockets for reliable contacts to the packages.

Figure 38. Test interface board.

To conduct the measurement, an automatic tester was developed using the National Instrument NI-6115 data acquisition card and LabView to ensure an accurate and reliable data collection. Besides the power supplies needed during the measurement, the data acquisition card NI-6115 was used with its signal connection terminal. This is a 12-bit data acquisition card with 10 MHz sampling rate. It has 16 DAC channels and 4 ADC channels. The card was installed in a PC. The testing program was written in NI LabView. The schematic of the measurement and the actual instrumentation are shown in Figure 40 and Figure 41, respectively. The stress conditions and the test circuitry are given as follows for different test structures.

5.2.1. Tests for EM test structures

Stress condition:

- Thermal stress: 120°C
- Electrical stress: 1.2v DC
- Stress duration: 350 hours

Measurement setup:

- Resistance measurement with Kelvin structure
- R_T=V/I, as in Figure 39

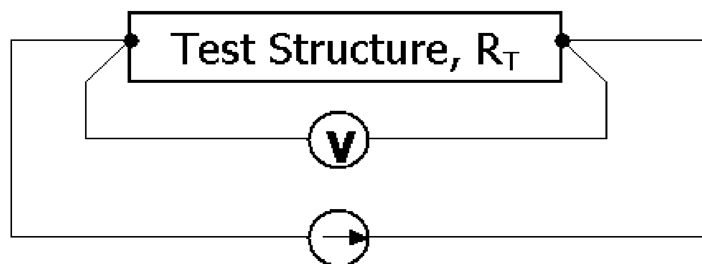


Figure 39. EM measurement.

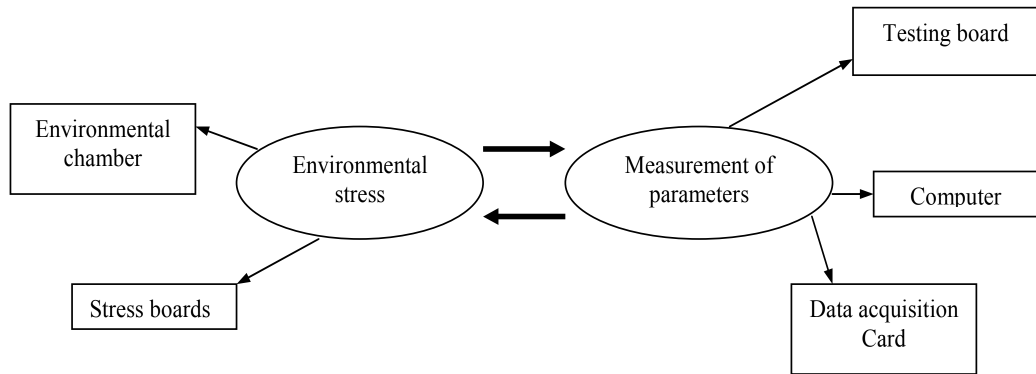


Figure 40. Schematic of measurement.

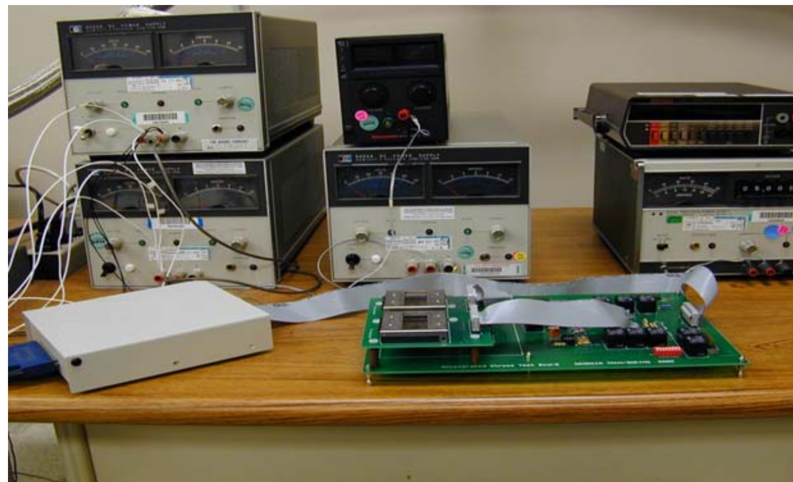


Figure 41. Measurement instrumentation.

5.2.2. Tests for HC test structures

Stress condition:

- Thermal stress: −40°C
- Electrical stress: V_D=7.15v, V_G=3v
- Stress duration: 350 hours

Measurement setup:

- Drain current measurement with different V_G and V_D, as in Figure 42

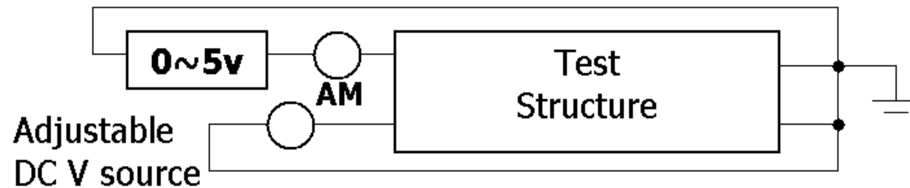


Figure 42. HC measurement.

5.2.3. Tests for circuit level test structures

Stress condition:

- Thermal stress: −40°C
- Electrical stress: V_D=7.15v and V_G=3v for differential nMOS pair
- Stress duration: 500 hours

Measurement setup for inverter:

- VTC and noise margin measurement at logic 1
- Binary search to find V_{IH}
- NMH=V_{OH}-V_{IH}, measured as in Figure 43

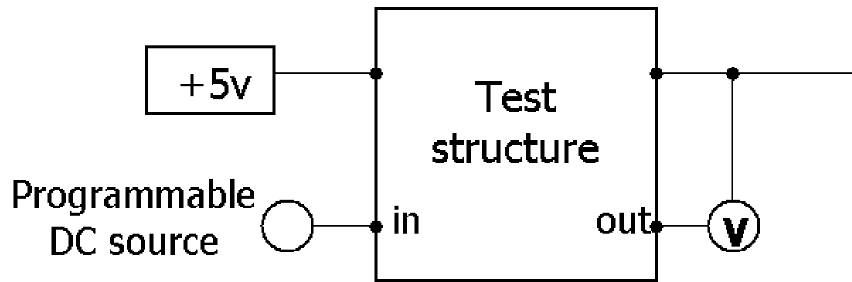


Figure 43. Inverter measurement.

Measurement setup for op-amp:

- Open-loop gain measurement
- A nulling amp is used to prevent DUT saturation
- $G = -\left(\frac{R_1 + R_2}{R_1}\right)\frac{\Delta V_{SRC1}}{\Delta V_{O-NULL}}$ as in Figure 44 [48]

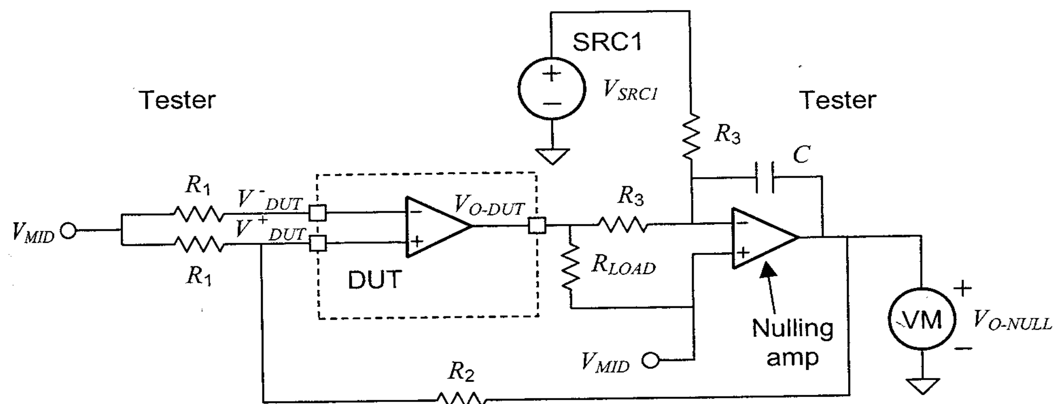


Figure 44. Op-amp measurement.

5.3. Calibration of ARET

For EM models, some of the material parameters such as the grain boundary activation energy were not supplied accurately. In order to validate and use the simulator, those key parameters, as well as the equations themselves, must be calibrated by the parameters extracted from the test data. The EM damage growth function can be written in a basic form

$$V = A_G A_C D e^{-A_B Q(t)} \quad (45)$$

where parameters A_G, A_C, and A_B have to be calibrated by using the curve fitting functions [43]. The general form of the fit is given by

$$F = a e^{\tau X} \quad (46)$$

where F is the output sequence, X is the input sequence, a is the amplitude, and τ is the damping constant. Among those parameters, A_C is specifically for modifying the structural factor in the corner areas. Thus, it was generated using the data from the corner test structure S4, as the rest of parameters were extracted from the data of structure S3d.

The EM stress test was conducted in an environmental chamber under DC 1.2v at 120 °C for 350 hours for all six types of metal traces with different geometries. The test data for two most significantly stressed test structures are used in this thesis. Structure S3d is a straight metal trace with a length of 120 μm and a width of 1.05 μm, and structure S4 is a metal trace with corners, 118μm long and 1.05μm wide. After calibration ARET was run repeatedly simulating the trace resistance change under the same stress condition as in the actual test. A comparison of simulated degradation curve with test data is shown in Figure 45 [44].

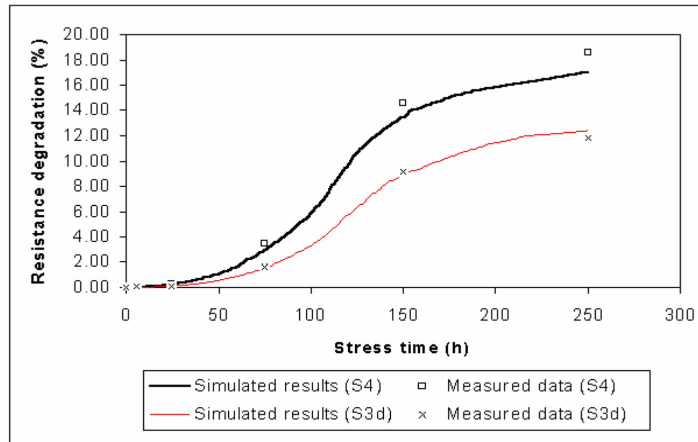


Figure 45. EM model calibration with test data.

From Figure 45, a good agreement between the stress test data and the simulated EM degradation using ARET can be observed, especially within the first 150 hours test period. The sharp degradation slope right before the 150-hour point was produced by the current crowding effect. When the test reached 150 hours or so, the size of the metal part left in the trace was approaching its material average grain size, which in turn started to limit the number of mass flow paths and caused a relatively slower increase of degradation. This is the so-called "bamboo structure". However, since the trace width left at this point has been very thin, and the current crowding and local temperature elevation have dominated the degradation process, the trace will soon be broken.

The corner structure S4 showed a faster overall degradation than straight trace S3d due to the modified structural factors at the corner areas. The increasing deviation between the simulation and the test data at 250 hours shown by structure S4 could be due to the different local temperature change assumed in thermal modeling.

The EM models for interconnect with post-fab physical defects was checked by the published data in Table 4, by J. R. Lloyd *et al* [39]. In that work the author was

conducting the experiments for a totally different purpose – to demonstrate the passivation defects in metal thin film. However, the data can be used as an excellent reference in our work.

Table 4. Simulation results for Al-5%Cu traces compared with measured data.

Sample group	MTTF (hours)		No. of total failures		No. of failures at defect site	
	S	M	S	M	S	M
A	96~114	**113	N/A		N/A	
B	15~22	22.3	24	23	24	20
C	95~102	105	24	20	24	20

*** S – Simulated results; M – Measured data, J. R. Lloyd et al, 1982.**
**** A deviation σ of 0.37, 0.99, and 0.92 exists for group A, B, and C.**

The simulation results are listed in the table compared with the measured data. Totally three groups of sample stripes with 20 ~ 24 stripes in each group were set up:

- Group A – nearly defect-free samples under a current density of 25 mA/μm^2 and 270 °C,
- Group B – samples with a 90%-defect under a current density of 25 mA/μm^2 and 312 °C and,
- Group C – samples with a 90%-defect under a current density of 10 mA/μm^2 and 250 °C,

where the 90%-defect means the defect size is 90% of the stripe width. The mean time-to-failure (MTTF) was simulated and the failure positions were recorded. The table

shows a very good consistency of the results from ARET simulation with the measured data. It also proves that, if the size of physical defect is comparable to the line dimension, it will often be the major cause for final break-offs.

The HC test structures had been in the chamber under V_D=7.15v, V_G=3v at –40 °C for 250 hours. After 250 hours stress, significant degradations in transistor drain current of test structures had been observed, which will lead to the degradations on many transistor major properties, such as the threshold voltage. Again, calibrated ARET was used to generate the transistor drain current vs. drain voltage characteristics, and compared with the stress test data before and after the stress cycle. The measurements were conducted under V_G=4v. The final test results for the test structure S5a and S5c are demonstrated in Figure 46 and Figure 47, compared to the simulation curves generated by ARET. Transistor S5a and S5c are both nMOS transistors with aspect ratios of W/L=6μm/0.6μm and W/L=9μm/0.6μm, respectively.

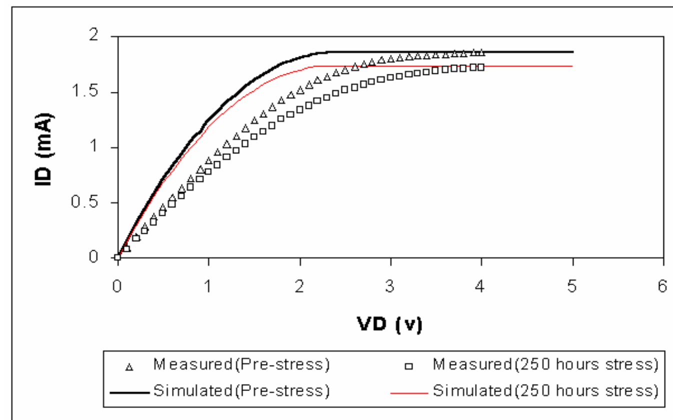


Figure 46. ID vs. V_D of test structure S5a (W/L=6μm/0.6μm).

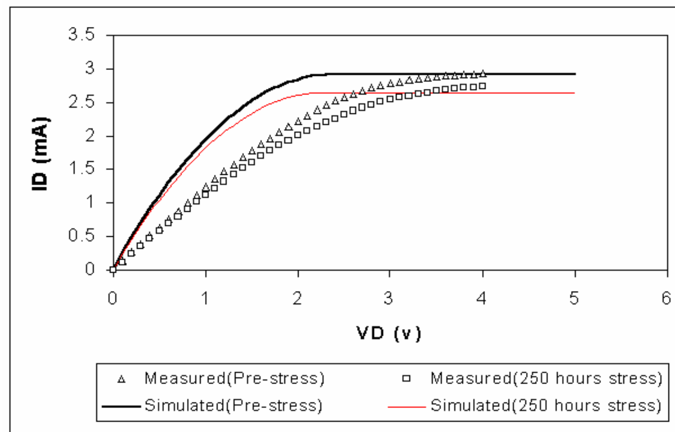


Figure 47. ID vs. V_D of test structure S5c (W/L=9μm/0.6μm).

Both results show that the simulations are very consistent with the stress test data. In Figure 46, the degradation of drain current in saturated region dropped by 7.4% for S5a after 250 hours stress, and in Figure 47 it dropped by 6.6% for transistor S5c. Notice there is a significant discrepancy between the simulations and the measurements in the linear regions of all four curves. This is because of the localization of hot-carrier damage at the drain area. After some amount of pre-stress under relatively low stress level before the stress test was actually started, the hot-carrier had shown its influence on charge distribution, although it was very minor. However, after the transistor pinch-off, the channel current is governed by the physical properties of the inverted channel between the source and the pinch-off point. With hot-carrier-induced damage mostly located in the area near the drain, its influence on the drain current thus becomes relatively less in saturation. This is the reason for this discrepancy to appear in linear region but not in the saturated region. Figure 47 also shows a relatively poor agreement between simulation and test compared with Figure 46. This indicates that the models in ARET handle the degradation evaluation better in low-channel current situation.

The op-amp and inverter circuits were designed to verify ARET simulation at circuit level. All test circuits have been under electrical/thermal stresses for more than 500 hours. For both circuits, selected transistor groups were stressed under V_D=7.15v, V_G=3v at –40 °C and the degradations on circuit circuit-level specs were recorded, compared with ARET simulations during the verification.

For the inverter circuit, the nMOS transistor (W/L=8.4μm/0.6μm) was under the applied electrical/thermal stress and the degradation of logic 1 noise margin (NMH) due to HC was monitored through the test, which is defined as

$$NMH = V_{OH} - V_{IH} \tag{47}$$

where $V_{OH} \approx V_{DD}$ is the output high voltage and V_{IH} is defined as the input voltage at that $dV_{out}/dV_{in} = -1$. For an inverter to perform normal functions the noise margin has to be greater at least than zero. Within 500 hours, the inverter had shown significant performance failures. Figure 48 shows the measured noise margin degradation versus stress time, compared with ARET simulation result, and Figure 49 shows the corresponding VTC shift.

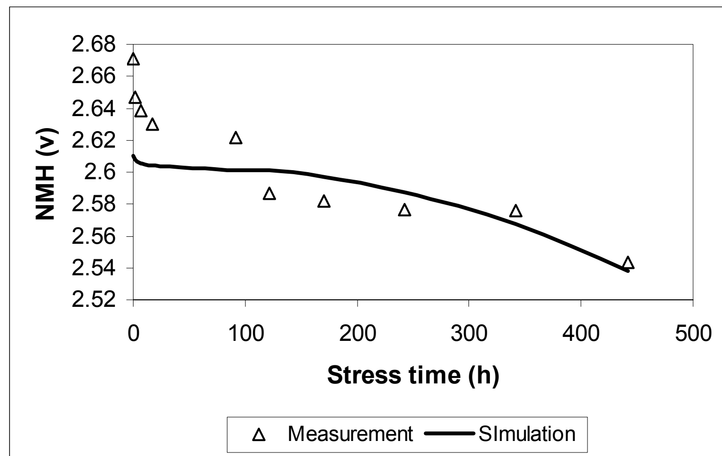


Figure 48. Noise margin degradation of inverter.

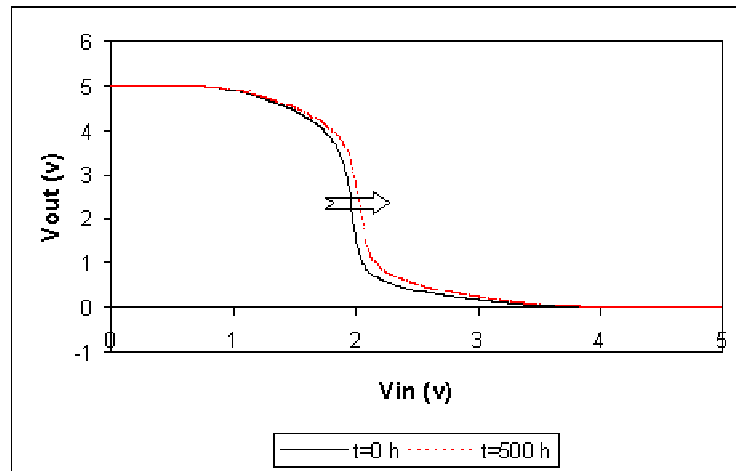


Figure 49. VTC shift of inverter.

From this comparison, a good overall consistency is observed between the simulation and the actual test, which well validates the ARET circuit-level simulation with great confidence. However, a potential problem with this verification is that, somehow a significant deviation in the data was observed and the data points used are based on 2 of 13 total packages. For the rest of the packages, six have been broken and five have not shown significant degradations. Since the total number of samples is relatively small, this actually indicates a poor confidence, although those packages showing inconsistent degradations with simulations could be due to the existing random defects and improper test setup. Other issue is that, the total degradation at 500 hours was fairly small (<0.2v). This is due to the partial stress applied only to a small part of the circuit during the test.

CHAPTER 6

DESIGN FOR RELIABILITY

Design-for-reliability (DFR) has not received as much attention as obtained by design-for-manufacturability and design-for-testability. It is partially because that the mature process and material technology used in past technology generations made the wear-out degradation extremely inactive. In fact, semiconductor industry had worried about the infant mortality much more than wear-out. However, as the technology continues to scale into submicron range, this is no longer the case. Due to the aggressively shrunk device sizes, the increased power for enhanced performance, as well as the resulting high operating temperature, the use lifetime (chance failure) of the bathtub curve has become much shorter, and not constant any more. The wear-out degradation and failure have started to show up in the use lifetime of semiconductor products, especially in leading-edge technologies. IC reliability has become an active issue to be worried by the industry, and design-for-reliability is absolutely an effective approach to address this issue.

6.1. Reliability hotspot identification

In case of circuit failure, most of time it is not that every component in the circuit fails. Instead, only a minor part of the circuit fails causing the whole circuit malfunction. This indicates a simple rule that, it is possible to improve circuit overall reliability by making part of circuit components more reliable. This is the local design-for-reliability approach proposed in this work, and this part of circuit components are defined as the circuit reliability hotspot(s).

Basically a reliability hotspot means the circuit component that is most likely to cause the circuit fail under certain failure criterion. In other words, the component that is most likely to fail first failing the whole circuit. Identifying such reliability hotspots offers the opportunity to improve circuit reliability by redesigning the hotspot components.

6.1.1. Hotspot under interconnect failure

Any interconnect open can immediately cause the circuit malfunction due to the loss of connection, and before it opens the resistance degradation does not affect circuit overall performance meaningfully simply because that the change is too small compared to the existing resistances in the circuit. Thus, the interconnect line that is most probable to open first in the circuit is taken as the reliability hotspot.

To find such a interconnect line, the total probability of breaking off first in the whole circuit with all possible defect conditions is calculated for each involved line, respectively. According to Equation (43), this can be written in following form.

$$\Pr\{F_i\} = \sum_{k=1}^{m\times n} \Pr\{t_C_k \mid k @ Line_i\} \tag{48}$$

where i is from 1 to m for all involved interconnect lines, and $\Pr\{F_i\}$ thus represents the probability that the circuit fails at interconnect line i. The hotspot is thus designated to the line with the maximum such probability.

6.1.2. Hotspot under device degradation

6.1.2.1. CMOS digital circuit

For a CMOS digital circuit, the most critical specification is the speed, or the propagation delay. Under gradual degradation mechanisms such as hot-carrier, the devices will slow down and the propagation delay over that path will increase as the

result of long-term wear-out. Once the propagation delay over a path in the circuit exceeds the designated maximum allowable delay the timing failure occurs. Thus the reliability hotspot is identified in the following way in ARET.

- First, the degradation of propagation delay as function of time is simulated path by path, and the first path in the circuit over that the propagation delay exceeds the allowable maximum delay is located, and defined as the reliability critical path (RCP). This is the path that gives the current circuit reliability (lifetime).
- Then, in the RCP the degradation of delay is simulated gate by gate, and the gate that generates the maximum delay increment by the time the circuit fails is finally identified as the reliability hotspot.

To find RCP, the following steps are employed.

1) Simulate propagation delay for every single gate in the circuit, and obtain path delays by adding gate delays together for all paths.
2) Calculate degradation factors for every gate at time t_i, based on component level reliability simulation.
3) Update path delays with degradation factors at time t_i, until the delay over one path reaches time constraint.

To find the hotspot gate in the RCP, in general, the increase of delay from beginning to failure time is simulated for every gate, and the gate that generates the maximum delay increment is identified as the final reliability hotspot gate. From the viewpoint of reliability, this is to find the gate that directly contributes the most to the current circuit failure. Therefore, improving the reliability of hotspot gate will benefit the overall circuit

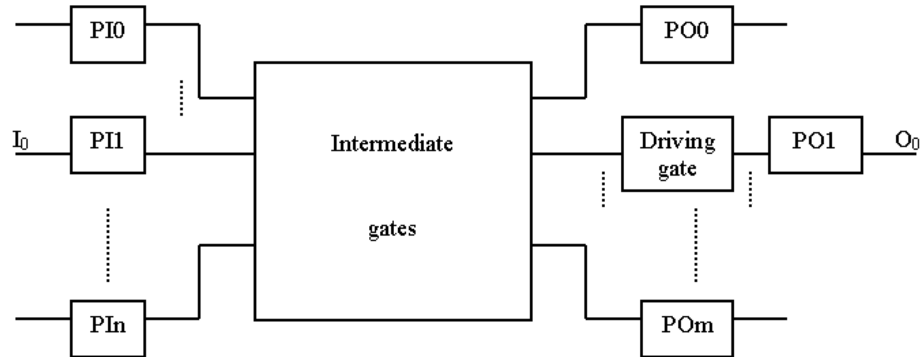
reliability the most. In terms of hot-carrier degradation, since the damage is mainly managed by transient current in devices, the gate with maximum $\eta_i C_{FOi}$ is identified as the hotspot gate, where η_i is the switching activity and C_{FOi} is the fanout capacitance. For simplicity, term $\eta_i FO_i$ is used instead of $\eta_i C_{FOi}$ in ARET. The pseudo-code of hotspot identification process for a schematic circuit is given in Figure 50.

Under catastrophic failures such as gate oxide breakdown, although the oxide wear-out is time-dependent, the transistor parameters don't change significantly until the final dielectric breakdown occurs, when extraordinarily large current is drawn and the device becomes very slow. Thus, the first device exhibiting breakdown is identified as a reliability hotspot. Similarly, improving the gate oxide reliability of the hotspot maximally improves the gain in the overall circuit lifetime.

6.1.2.2. Analog circuits

Due to the complexity to evaluate analog circuit performance, a group of *k* circuit-level key specifications are first created with corresponding acceptability ranges. The first failure in these key performance specifications fails the whole circuit. The reliability hotspot is thus identified as follow.

Assume there are *n* circuit components involved in circuit performance degradation. By reliability simulation, circuit level performance degradation is simulated until one of the key specifications fails the circuit. The degradation of this spec by the time of failure, ΔP_i with $1 \leq i \leq k$, can be obtained based on degradations of all *n* components, Δx_j with $1 \leq j \leq n$. The reliability hotspot is defined as the component that contributes the most to the overall degradation ΔP_i and identified to be the component with the maximum value of ΔPH, as described in the following equation.




```
Begin
{ do {
for (all gates) delay_update(gate_i)
for (all paths) delay_simulation(path_i)
for (all gates) reliability_simulation(gate_i)
save_path (min timing slack)
} while (timing slack ≥ 0)
Identify RCP with path saved
for (t_0 to t) {
reliability_simulation(gate_j in RCP)
}
gate i with max degradation → Hotspot
}
End
```

Figure 50. Locating RCP and hotspot.

$$\Delta PH_j = \left| \Delta P_i\left(\Delta x_1, \Delta x_2 ... \Delta x_n\right) - \Delta P_i\left(\Delta x_1, \Delta x_2 ... \Delta x_n\right)|_{\Delta x_j = 0} \right| \tag{49}$$

where j is an integer between 1 and n. The identification process is implemented by reliability simulation using ARET.

6.2. Basic DFR approach

The essential goal of the proposed DFR approach is performing redesign work only in local areas of the circuit to improve the overall circuit reliability with the original circuit performance maintained. While conducting the local DFR, this generally requires to meet two conditions: 1) the local redesign gives the most effective improvement possible in circuit overall reliability, and 2) the local redesign does not significantly change the overall circuit-level performance specified in original design.

Based on previous discussions, it is obvious that the reliability hotspots should be the components of which the reliability is improved by DFR to meet condition 1. However, such a redesign will most probably change the original circuit performance, which contradicts meeting condition 2. Therefore, extra redesign works have to be conducted maintaining the original circuit performance. These design works are taken in an extended area around the hotspot, so that the performance change (loss) due to the redesign at hotspot can be absorbed without spreading beyond this local area. This strategy is demonstrated in Figure 51, where component C1 the reliability hotspot, and the designs at C1, C2, C4 and C5 are all updated so that when they are considered together as a new component, the outer component parameters are still same as original ones. It is simply the same case as replacing a resistor with another resistor having the same resistance reading but higher reliability. While conducting the local DFR, generally a three-step process is followed: reliability hotspot identification → local redesign around

hotspot → reliability simulation. The last step is used to evaluate the effect of DFR as it is necessary to decide if further DFR cycles are required.

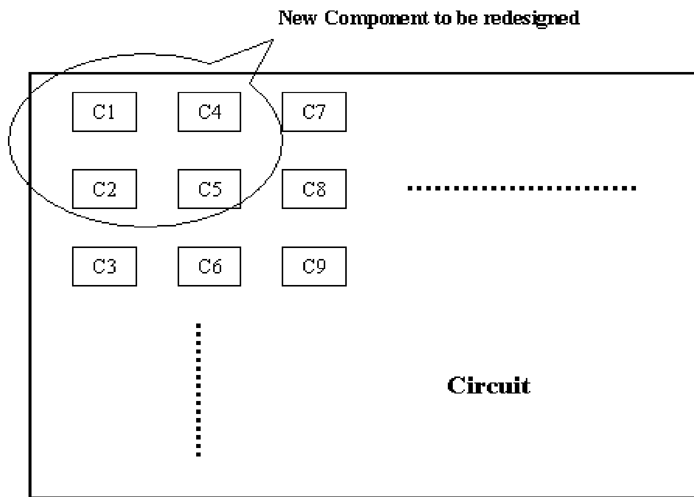


Figure 51. Basic local DFR strategy.

Interconnect open can be considered as a type of catastrophic failures. When an interconnect trace is broken open, the connection is gone and the designed circuit function is no longer maintained. Since it is most probable that, or in most of time, the circuit will fail due to interconnect open at the hotspot line, any design update that makes the hotspot line last longer will enhance the circuit overall interconnect reliability as well. For the major interconnect failure mechanism EM, the increase of the trace width can effectively enhance the EM resistance of the trace. Thus, by widening the hotspot trace, condition 1 of local DFR approach is satisfied. As discussed earlier in Chapter 3, the change of the dimensions of a minor part of circuit interconnect traces, and thus the change of resistances of these traces, do not change the circuit overall performance. Therefore, the second condition of DFR is automatically met too.

Device degradations are overwhelmingly caused by charge carriers that obtain high energy under large electric fields due to small feature sizes. Thus, in order to improve the device reliability, increasing channel length, or *dimension modulation*, has been the most direct and effective means. For example, in hot-carrier degradation, the change of channel length has an almost exponential impact on transistor damage. However, in a CMOS digital circuit, by doing this to the hotspot gate the gate will slow down, which violates the required condition 2. To solve this, the channel width of the hotspot gate is increased too to speed up the gate, while basically increasing transistor channel width can also release the device degradation such as HC damage [45]. Moreover, due to the increase of load capacitance, an additional speed loss at the preceding gate of hotspot in the reliability critical path (RCP) will also be observed. Thus the speed-up by the increase of channel width will have to compensate the speed loss at the preceding gate as well, so that the total delay over RCP is same as in the original design. For complex logic circuits, in which more paths are affected by the redesign along RCP, necessary redesign processes will also have to be performed accordingly to maintain the speeds along those affected non-RCPs.

In CMOS digital logic almost all types of device degradations happen intensively during the signal transient period when a switching current is flowing through the transistor. This can be easily understood from the fact that most device degradations are caused and accelerated by energized carriers in the channel. Figure 52 shows the plot of bond-breaking current, which is used to represent hot-carrier damage, during transient period by Y. Leblebici [14] for the nMOS transistor in an inverter. It can be clearly seen that a major peak of bond-breaking current is generated during input transient time,

especially when the transistor is in saturation. The gate oxide also experiences considerably more wear-out during transient time because of the generation of holes due to impact ionization. Shortening transient time, or *signal modulation*, thus becomes another approach to improve device reliability.

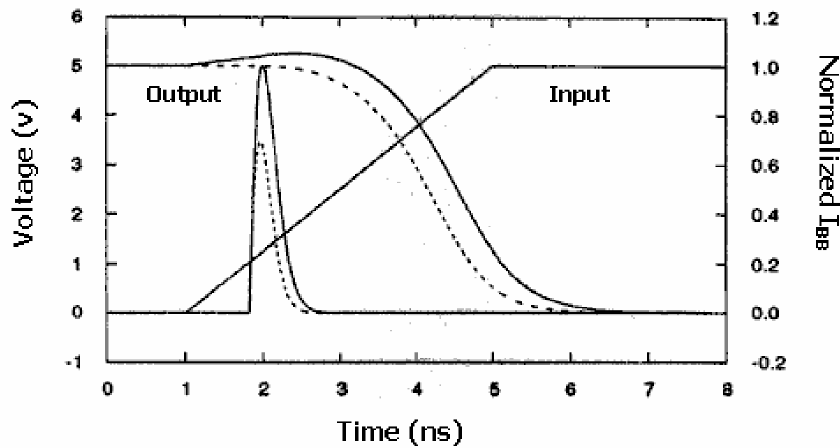


Figure 52. Hot-carrier degradation during signal transition in CMOS inverter.

In order to shorten the switching time of the hotspot gate, its preceding gate in RCP has to be speeded up. This can be accomplished by increasing the transistor channel width of the preceding gate. However, by doing so the driving gate of the preceding gate will be slowed down due to the increased fan-out capacitance, which may increase the overall delay of RCP and violate condition 2. Therefore, similar to dimension modulation, the speed gain at the preceding gate of the hotspot must be able to compensate the speed loss at its driving gate as well. Again, for complex logic circuit all involved paths (RCP and non-RCPs) must be carefully evaluated to maintain a better or equal speed.

Besides the modulations of dimension and signal, some other approaches may also be considered for DFR. Among them, the local use of dual power supply, or *power supply modulation*, can give a neat reliability enhancement to device. For failure mechanisms such as gate oxide wear-out, some recent research results suggest that for ultra-thin oxide the gate voltage has an even stronger impact than the electric field [32]. Thus, by applying a lower power supply at the driving gate of hotspot gate, which gives a lower gate voltage to the hotspot gate, and the normal power supply for the rest of circuit, the gate oxide wear-out can be released. However, the lower power supply will increase the delay of the driving gate of hotspot. To maintain the speed, channel widths of transistors in the gate are widened to speed up the gate and compensate the speed loss due to the lower power supply.

In CMOS logic the inverter is the simplest and basic structure. It contains only one n-p complementary pair that builds the foundation for the entire CMOS static logic family. In following chapters the local DFR approaches are first developed and evaluated for CMOS inverter network. The algorithms are then extended to all CMOS digital circuits. Also, due to the much more significant device degradations observed in nMOS transistor than in pMOS, only nMOS-related degradation and the corresponding delay t_{HL} are discussed. The same approaches and similar analyses apply to pMOS devices as well.

Unlike in digital circuits where only two logic levels of signal are seen, continuous signals are presented in analog/mixed-signal circuits. Various circuit-level specifications are used to evaluate circuit performance. Furthermore, any local change of circuit design may affect overall circuit performance dramatically by different mechanisms such as mismatching. Therefore, the application of local DFR in analog/mixed-signal circuits

requires much more thoughts and efforts. To meet the local DFR condition 1, the hotspot component and, if necessary, its adjacent components must be redesigned in such way that they become more reliable under certain failure mechanisms. Meanwhile, the circuit performance has to be maintained. Thus, a complex mapping between the circuit performance and component parameters must be created by certain techniques such as circuit synthesis.

CHAPTER 7

DFR WITH INTERCONNECT FAILURES

Feature size scaling has been pushing interconnect into a region of extremely high interconnect density and high current density. During the past years the current density, which is the driving force for EM damage, has increased by 1.5 ~ 2 times per generation. This has made interconnect failure a serious reliability issue.

7.1. Basic approach

For conducting a successful local DFR to circuit interconnect, the design update during DFR is not supposed to affect the overall circuit performance. Based on discussions in previous chapters, the limited change of resistance/capacitance for the hotspot interconnect lines can be neglected compared to the rest of the circuit, especially when the case of a few hotspots out of millions of interconnect lines in VLSI circuits is concerned. This allows the reliability hotspot line being redesigned with different dimension to achieve a possible longer lifetime under EM without affecting overall circuit level performance.

In terms of EM degradation, with given technology, i.e., given interconnect materials, thickness, and other process parameters, the width of the line is the key parameter making interconnect more reliable. Generally, due to the nature of EM, a wider interconnect line is supposed to release the problems caused by mass flow, and thus alleviate the EM damage. Figure 53 shows the simulation result for a set of 250μm long Al stripes with a 2μm defect presented under 300°C for 100 hours, which clearly indicates that as the line becomes wider the expected EM degradation has been released.

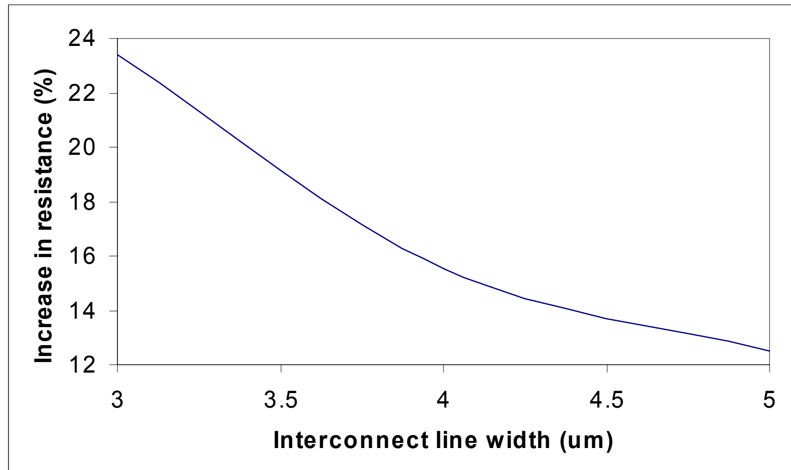


Figure 53. Degradations of interconnect lines with different widths in 100 hours.

Therefore, a straightforward local DFR algorithm for interconnect can be accomplished by redesigning the hotspot line with an increased width. Theoretically, this must be restricted by the layout area and cost. The more gain in reliability from the increase of the interconnect line width means the more area and cost is taken. However, because of the "localization" of this DFR approach, this trade-off is not expected to be significant at all. A perfect balance can be found by considering the overall circuit degradation including both interconnect and device comprehensively. The effect of DFR has to be evaluated by reliability simulation. A complete local DFR algorithm for IC interconnect reliability is thus developed accordingly.

7.2. Algorithm implementation

The flow chart of a complete local DFR algorithm for IC interconnect under EM is shown in Figure 54.

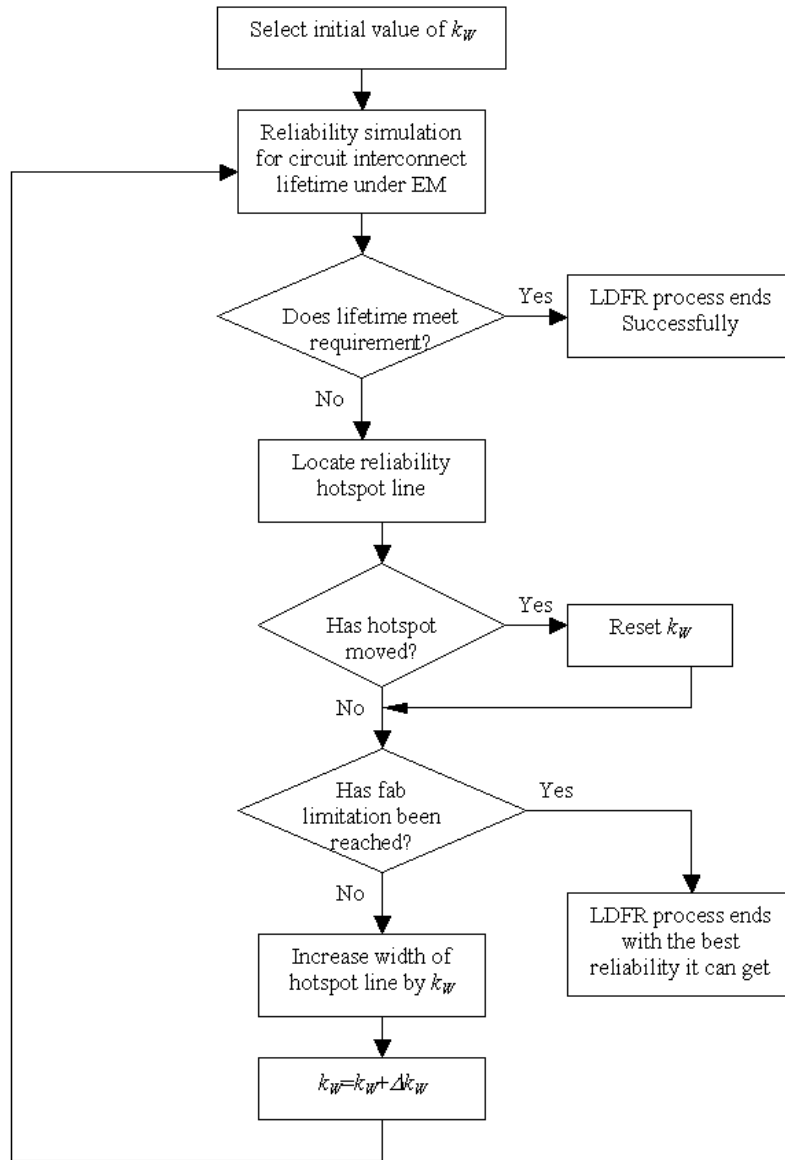


Figure 54. Local DFR algorithm for interconnect under EM.

As shown in the figure, before conducting DFR to a circuit condition, an expectation or requirement of circuit lifetime must be specified. This lets the algorithm know when to stop and the updated circuit lifetime is given by reliability simulation with ARET. In case the requirement is specified by other reliability indices such as failure rate, an internal conversion will take place inside the simulator.

k_W is a scaling factor of interconnect line width, i.e., $W' = k_W W$, where W' is the updated line width. This factor will be increased by Δk_W if the lifetime requirement is not met after one cycle of DFR. Both k_W and Δk_W are selected according to certain fabrication process. The fabrication limitation includes design rules related to interconnect such as specs for spacing. Every time a hotspot is identified, it has to be confirmed that this is the one that has been identified last time, or due to the design update to previous hotspot a new hotspot line has been generated. If it is the latter, then the scaling factor will be reset to start over on the new hotspot line.

7.3. Experimental results

To demonstrate the local DFR algorithm for interconnect, two experiments are conducted. For comparison, the first circuit in the experiments is the same two-stage op-amp circuit in Figure 22 used in Chapter 3. As same as in previous case, three metal interconnect lines in the op-amp, r1, r2, and r3, are selected as the group subject to EM damage, in which r1, r2 are at output and r3 is at input of the circuit, and the rest of the circuit is assumed to be EM damage-free. All three interconnects are pure aluminum traces and 118μm long, 1μm wide, and 1μm thick. The temperature is 300 °C. Regarding the defect condition, 0.3μm, 0.6μm, and 0.8μm are selected as the most significant key

sizes in terms of interconnect lifetime based on ARET simulation. For simplicity only one DFR run is conducted without a lifetime requirement specified.

Equation (48) is used in order to locate the hotspot line. Based on Table 3 the probabilities that the circuit fails at certain interconnect line due to interconnect break-off under EM are calculated for all lines with all possible defects presented. The line having the maximum such probability for all possible defect conditions is identified as hotspot, which in this case is line r3, as shown in the third column of Table 5.

After the hotspot line is located, it is then updated with an increased width. For this example the width scaling factor is set to 1.4, which gives an updated width of 1.4 μm right after DFR. With this scaling, another ARET simulation for circuit lifetime due to interconnect failure is conducted and shown in Table 6 that has the same structure as Table 3.

It is observed that by only updating the design of one interconnect trace the overall circuit lifetime due to interconnect failure under EM has been increased to 146.4 hours, about 1.8% over original design. In this example for simplicity all three interconnect lines involved have very similar geometries and parameters resulting in very close lifetimes. Thus, after one cycle of DFR the hotspot has moved to another line, r2, which has a lifetime only slightly longer than r3. Therefore the design update and reliability improvement at r3 are not fully reflected at circuit level. This is one reason that the lifetime improvement is not very significant in this case. To see a more significant result, more DFR cycles are needed. In practical ICs, much more interconnect traces with much more different parameters are involved. Therefore, there is a better chance that a more significant improvement in reliability after local DFR can be obtained.

Table 5. Interconnect hotspot identification.

Defect type		Probability causing circuit fail	Sum of probabilities
r1	0.3μm	0	0.0906
	0.6μm	0.067	
	0.8μm	0.0236	
	Defect-free	0	
r2	0.3μm	0.098	0.2128
	0.6μm	0.08	
	0.8μm	0.0248	
	Defect-free	0.01	
r3	0.3μm	0.582	0.6968
	0.6μm	0.08	
	0.8μm	0.0248	
	Defect-free	0.01	
Hotspot line	r3		

Table 6. Op-amp lifetime prediction after local DFR.

Defect type		Probability	Lifetime (hours)
r1	0.3μm	0.9364	171.59
	0.6μm	0.0887	108.68
	0.8μm	0.0248	56.176
	Defect-free	0.0564	172.84
r2	0.3μm	0.9364	167.01
	0.6μm	0.0887	104.92
	0.8μm	0.0248	54.093
	Defect-free	0.0564	168.68
r3	0.3μm	0.9364	175.52
	0.6μm	0.0887	112.28
	0.8μm	0.0248	66.303
	Defect-free	0.0564	177.67
Expected circuit lifetime (hours)	146.4		

In the second example, the CMOS mixer shown in Figure 17 is used. Similarly, three interconnect lines, m1 at the signal input, m2 connecting the sources of input signal pair, and m3 at the output are selected as the interconnect group degrading under EM. All interconnect line have the same material properties as in first example, except that m1 and m3 have 2μm widths. The temperature is set to 300 °C. The signal input stimulus is a 50 MHz sine wave with 100 mA amplitude, which gives an expected interconnect lifetime of 707.6 hours under assumed defect conditions.

By Equation (48) the line m2 is identified as the hotspot. During the following DFR, the width scaling factor is set to 1.5, which updates the line width of m2 to 1.5 μm. The redesigned circuit is then simulated by ARET for re-evaluation, and an improved interconnect lifetime of 766.8 hours is obtained.

Compared with the original design, a promising lifetime improvement, about 8%, is observed. In this case, the hotspot line has different width from the other lines, and is the interconnect that carries much more current than the other lines. These differences make this hotspot have a significantly shorter expected lifetime. Thus, the reliability enhancement on the single interconnect line by DFR redesign efficiently improved the overall circuit interconnect lifetime. Since only one DFR cycle is taken, this result indicates that the local DFR approach is quite effective to enhance interconnect reliability under EM.

CHAPTER 8

DFR FOR CMOS DIGITAL CIRCUITS

As the vanguard in technology scaling, CMOS digital circuits have been suffering from performance degradations as the trade-off pursuing high clock speeds. Shrinking channels and oxides under consecutive stresses on active devices make the circuit performance degrade over time. With the non-stoppable technology scaling into submicron region and even smaller in the future, design approaches for reliability are certainly necessary to release device degradations and maintain the proper reliability safety margin for CMOS digital circuits.

Various DFR approaches are discussed in this chapter for CMOS digital circuits with device degradation under hot-carrier and gate oxide wear-out, among which, dimension modulation and signal modulation are presented as two major algorithms. Dimension modulation makes device more reliable by device resizing such as channel length and width, while signal modulation improves reliability by adjusting the input signal. Both algorithms are proved to be very effective DFR approaches by experimental data, with acceptable trade-offs such as increase of area and power.

8.1. DFR by dimension modulation

8.1.1. Algorithm based on inverter network

An inverter network is demonstrated in Figure 55, where gate n_k is identified to be the reliability hotspot. All effective parasitic capacitances between drain, gate and bulk are also shown, and C_{line} is the interconnect capacitance.

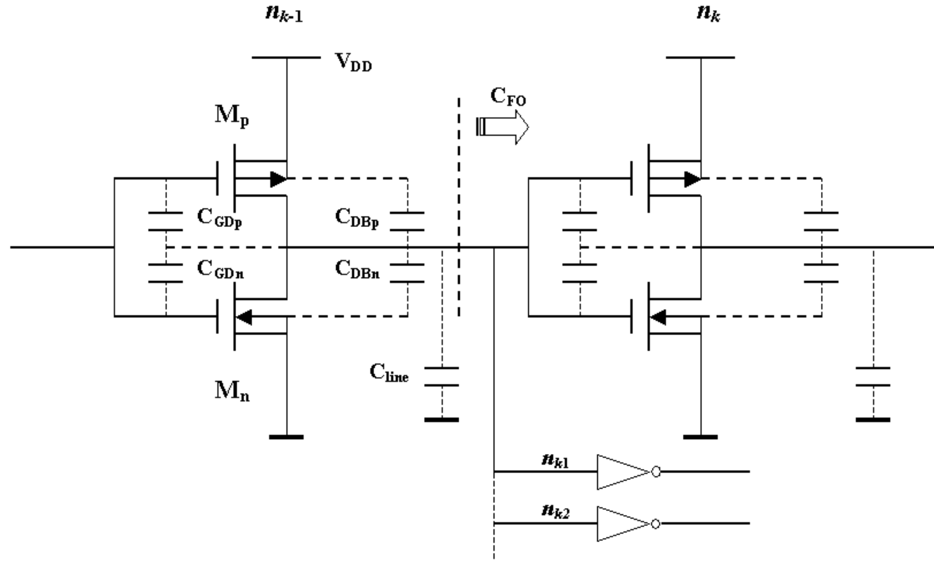


Figure 55. CMOS inverter network with effective capacitances.

In terms of nMOS-related degradation, the inverter fall time delay t_{HL} is analyzed in form of

$$t_{HL} = s_n \tau_n \tag{50}$$

where

$$s_n = \frac{2(V_{tn} - V_0)}{V_{DD} - V_{tn}} + \ln(\frac{2(V_{DD} - V_{tn})}{V_0} - 1) \tag{51}$$

is a voltage-dependent factor, and is independent of transistor sizing. The time constant τ_n is defined as

$$\tau_n = R_n C_{out} \tag{52}$$

in which

$$R_n = \frac{1}{k_n \frac{W}{L}(V_{DD} - V_{tn})} \tag{53}$$

and

$$C_{out} = (C_{GDn} + C_{GDp}) + (C_{DBn} + C_{DBp}) + (C_{line} + C_{FO}) \quad (54)$$

with all capacitances shown in Figure 55. By a simple approximation [46]

$$C_{GD} \approx \frac{1}{2} C_{OX} WL \quad (55)$$

where C_{OX} is the oxide capacitance, the time constant can be described as

$$\begin{aligned}
\tau_n &= \frac{1}{k_n \frac{W}{L}(V_{DD} - V_{tn})} \begin{bmatrix} (C_{GDn} + C_{GDp}) + (C_{DBn} + C_{DBp}) \\ + (C_{line} + C_{FO}) \end{bmatrix} \\
&\approx \frac{1}{k_n (V_{DD} - V_{tn})} \frac{L}{W} \left[C_{OX} WL + (C_{DBn} + C_{DBp} + C_{line} + C_{FO}) \right] \\
&= A\left[C_{OX} L^2 + \frac{L}{W}(C_{DBn} + C_{DBp} + C_{line} + C_{FO}) \right]
\end{aligned} \quad (56)$$

where

$$A = \frac{1}{k_n (V_{DD} - V_{tn})} \quad (57)$$

Since s_n is completely determined by supply voltage and threshold voltage, which are not affected by the circuit redesign in DFR, τ_n is used to represent the propagation delay t_{HL} throughout this work. Thus, for the local redesign at the hotspot gate n_k, if we increase the channel width W to $W \bullet K_W$, and the channel length L to $L \bullet K_L$, with $C_k = C_{DBn} + C_{DBp} + C_{line} + C_{FO}$, the gain in propagation delay at the hotspot gate after the local DFR is thus obtained as

$$\begin{aligned}
\Delta \tau_k &= A\left(C_{OX} L^2 - C_{OX} {K_L}^2 L^2 + \frac{L}{W} C_k - \frac{K_L L}{K_W W} C_k' \right) \\
&= A\left[C_{OX} \left(1 - {K_L}^2\right) L^2 + \left(C_k - \frac{K_L}{K_W} C_k' \right) \frac{L}{W} \right]
\end{aligned} \quad (58)$$

where $C_k \neq C_k'$ as it is a fairly complicated function of W and L.

For the preceding gate $n_{k\text{-}1}$, the increase of propagation delay at the gate due to the increased fan-out capacitance is obtained similarly as following. First, the capacitance due to fan-out load is given as

$$C_{FO} = FO(C_{OX}W_nL_n + C_{OX}W_pL_p) = 2FOC_{OX}WL \tag{59}$$

with FO representing the number of fan-outs. Because the increase of propagation delay is caused by change of fan-out capacitance only, the loss (increase) in propagation delay at the preceding gate is obtained as

$$\Delta\tau_{k-1} = A(C_{FO}' - C_{FO})\frac{L}{W} \tag{60}$$

However, from circuit perspective, the loss in propagation delay at the gate $n_{k\text{-}1}$ will also affect other paths involving gate $n_{k\text{-}1}$ and make them slower, such as the path containing gate $n_{k\text{-}1}$/gate n_{k1}, and the path of gate $n_{k\text{-}1}$/gate n_{k2} in Figure 55. This may cause some of those paths violate the timing requirement in the original design. To avoid taking the risk, all other fan-out gates of gate $n_{k\text{-}1}$ will have to be redesigned in a similar way as for the hotspot gate n_k.

For simplicity of design, all fan-out gates of gate $n_{k\text{-}1}$ are scaled up by the same K_W for channel width, except that a channel length scaling up of K_L is also taken for the hotspot gate n_k. Thus the updated fan-out capacitance for gate $n_{k\text{-}1}$ now becomes

$$C_{FO}' = \frac{1}{FO}C_{FO}K_WK_L + \frac{FO-1}{FO}C_{FO}K_W \tag{61}$$

which gives the updated loss in propagation delay

$$\Delta\tau_{k-1} = 2AC_{OX}L^2(K_WK_L + FOK_W - K_W - FO) \tag{62}$$

Finally the overall gain in propagation delay over RCP involving hotspot gate n_k becomes

$$\Delta\tau_{RCP} = \Delta\tau_k - \Delta\tau_{k-1}$$
$$= A\left\{\begin{aligned}&\left[(1-K_L{}^2 + 2K_W(1-K_L) + 2FO(1-K_W)\right]C_{OX}L^2\\ &+\frac{L}{W}\left(C_k - \frac{K_L}{K_W}C'_k\right)\end{aligned}\right\} \tag{63}$$

For those non-RCPs, which contain gates like n_{k1} and n_{k2}, the gain of propagation delay at these gates is obtained by rewriting Equation (58) and forcing K_L to 1.

$$\Delta\tau_k = AC_k\left(\frac{L}{W} - \frac{1}{K_W}\frac{L}{W}\right) \tag{64}$$

The overall change in propagation delay of non-RCPs thus becomes

$$\Delta\tau_{non-RCP} = \Delta\tau_k - \Delta\tau_{k-1}$$
$$= A\left\{\begin{aligned}&\left[(2K_W(1-K_L) + 2FO(1-K_W)\right]C_{OX}L^2\\ &+\frac{L}{W}\left(C_k - \frac{1}{K_W}C'_k\right)\end{aligned}\right\} \tag{65}$$

It is clear that for a successful local DFR, the change of propagation delay $\Delta\tau$ over both RCP and non-RCP must be greater than or equal to zero, which guarantees that the circuit after DFR has an overall performance no worse than the original design. By analyzing Equation (65) and Equation (63), and realizing that only C_{line} and C_{FO} are independent of K_W, it is concluded that for certain technology greater C_{line}, compared with parasitic capacitance, as well as large C_{FO} give more gain in speed during local DFR. According to our algorithm identifying hotspot, the hotspot gate where the redesign takes place is expected to have higher fanout than the rest of the circuit offering a larger C_{FO}. Also, in today's fast technology scaling trend, the interconnect capacitance has dominated the gate load capacitance. A C_{line}/C_{gate} of 10, even 20, is very easy to obtain. This makes the local DFR process much more hopeful to succeed in most of circuit situations.

Comparing Equation (65) to Equation (63), and knowing the fact that both K_L and K_W are greater than one, it is easily proven that $\Delta\tau_{RCP}<\Delta\tau_{non\text{-}RCP}$, which means that after the local DFR update if the overall gain in propagation delay over the RCP, $\Delta\tau_{RCP}$, is greater than zero, then the gain over the non-RCPs will also be greater than zero. This guarantees that after the local DFR, all affected paths will have the same or even smaller propagation delays than the original design, while the channel length of the hotspot gate has been increased for an improved reliability. Figure 56 shows the total change of propagation delay over the RCP of an inverter network as a function of device dimension adjustment, for 0.25µm technology and 0.1µm technology, respectively.

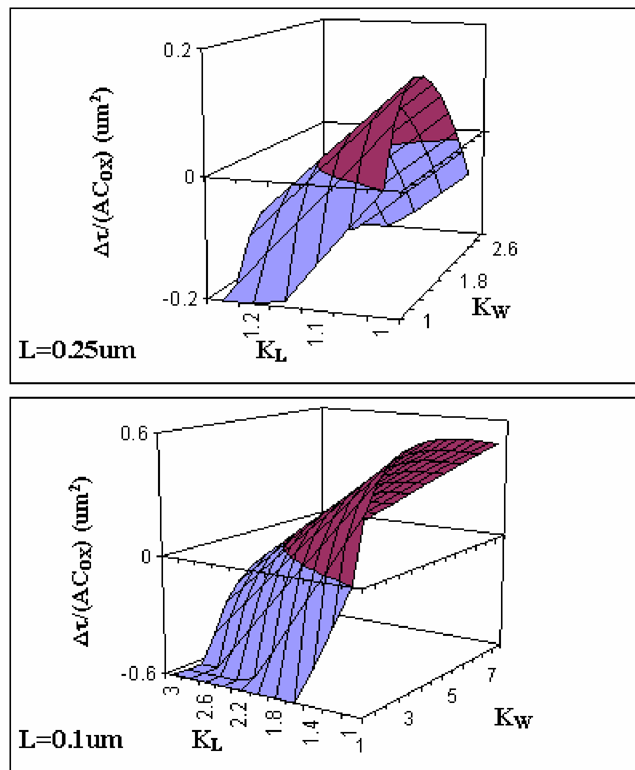


Figure 56. Overall change in propagation delay of RCP after local DFR.

With a 10 fF average interconnect capacitance and a relatively large fanout of 5, it is clearly shown that for both cases there are considerable ranges of device dimension adjustment for improving reliability (the parts of the plots in the figure above zero plane). This indicates that the local DFR can be successfully conducted for CMOS an inverter network, and with such available ranges of channel length adjustment, a significant reliability improvement of device can be expected.

8.1.2. Algorithm with technology scaling

Reliability problems are generated by aggressive technology scaling, and the solutions to these problems are driven forward by the leading-edge technologies from generation to generation. Therefore, it is necessary to prove that the local DFR approach is a working solution for future generations of technology.

For a clear view let us focus on the key part in the local DFR process – RCP. The available range of channel adjustment K_L offering a successful DFR is plotted with different channel lengths using Equation (63) as demonstrated in Figure 57, where it is assumed that the circuit is fabricated using AMI process parameter set with K_W=2.8 and an aspect ratio of 6. Figure 57 shows that to obtain a positive gain in overall propagation delay of RCP, the possibly maximum increase of channel length during DFR can be from 50% to 180%, as the channel length decreases from 0.25 μm to 10 nm. This tells us that a smaller technology can offer much more room for the local DFR and give an even more significant reliability improvement than a larger technology.

The same conclusion can also be drawn from Figure 56, where it can be observed that with 0.1μm channel length, a much larger available range of device redesign for circuit reliability ($K_L \geq 3$) is offered than the case of 0.25μm channel length ($K_L \approx 1.2$)

where $\Delta\tau$ is guaranteed greater than or equal to zero. All these analyses demonstrate that the local DFR technology will become more effective with future technology scaling-down.

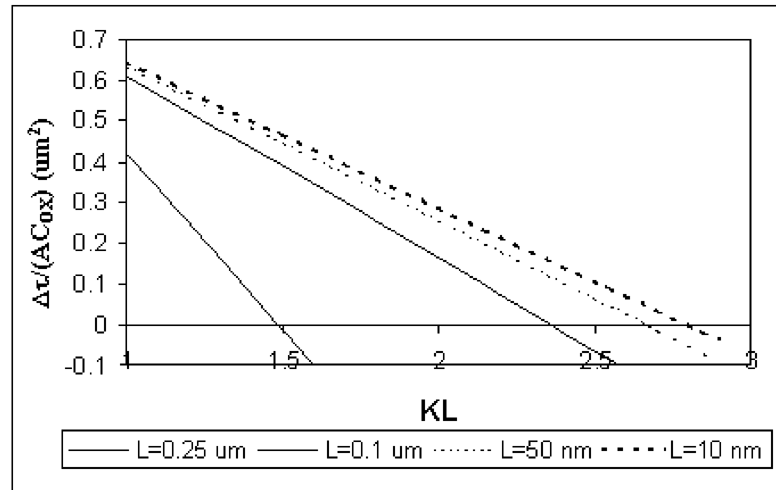


Figure 57. Available range of K_L in local DFR for different feature sizes.

8.1.3. Complete algorithm for CMOS digital logic

A complex CMOS logic gate is implemented by designing nMOS and pMOS switching arrays based on primary logic functions such as NOT, AND, and OR, as shown in Figure 58. In such circuits every input drives one complementary pair and the power is consumed only during the transient times. The inverter is thus the primary case in this family and the DFR algorithm for all CMOS logic circuits can be developed similarly.

Since in DFR the major concern is the degradation of discharging time through nMOS array, as far as the worst scenario is considered, the DFR local redesign will be performed for the longest chain in nMOS array. With *m* representing the number of nMOS transistors in series in the longest chain and *n* representing the number of

transistors connected to the output line, after re-writing Equation (56) and assuming $C_{DBn} \approx C_{DBp} = C_{DB}$, the time constant τ_n can be described as in Equation (66).

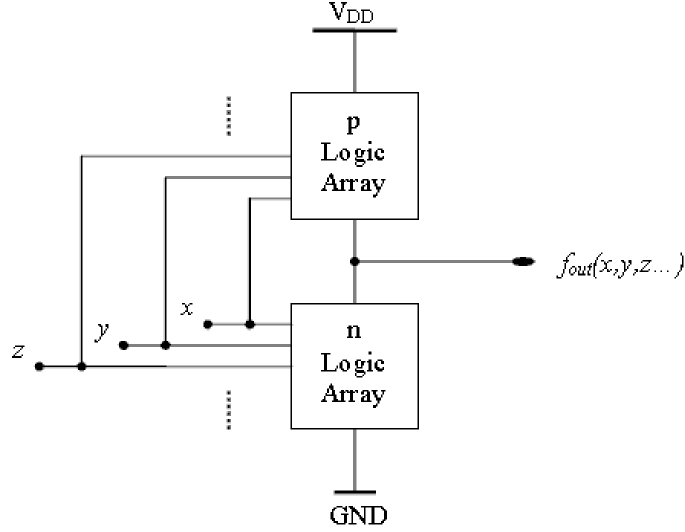


Figure 58. Structure of CMOS static logic.

$$\tau_n = \frac{m}{k_n \frac{W}{L}(V_{DD} - V_{tn})} \left[\frac{n}{2} C_{OX} WL + nC_{DB} + (C_{line} + C_{FO}) \right]$$
$$= Am \left[\frac{n}{2} C_{OX} L^2 + \frac{L}{W} \left(nC_{DB} + C_{line} + C_{FO} \right) \right] \quad (66)$$

again with

$$A = \frac{1}{k_n (V_{DD} - V_{tn})} \quad (67)$$

For simplicity, the inter-transistor capacitance between the adjacent series-connected transistors is ignored. In CMOS digital circuits, the device degradation happens during the transient periods, especially when the transistor is in saturation region. Thus this simplification is reasonable since only the first transistor in the series chain enters

saturation region for a time that is long enough to make hot-carrier degradations under transient signals [14].

For a general CMOS circuit demonstrated in Figure 59, at the hotspot gate G_{10}, after performing the DFR locally the channel width W and length L are increased to $W \bullet K_W$ and $L \bullet K_L$, respectively. The fall time constant then becomes

$$\tau_n' = Am\left[\frac{n}{2}C_{OX}K_L^2L^2 + \frac{K_L L}{K_W W}\left(nC_{DB} + C_{line} + C_{FO}\right)\right] \tag{68}$$

Therefore, the gain in propagation delay at the hotspot gate after DFR is obtained as

$$\begin{aligned}\Delta\tau_k &= \tau_n - \tau_n' \\ &= Am\begin{bmatrix}\frac{n}{2}C_{OX}L^2\left(1-K_L^2\right)+ \\ \frac{L}{W}\left(1-\frac{K_L}{K_W}\right)\left(nC_{DB}+C_{line}+C_{FO}\right)\end{bmatrix}\end{aligned} \tag{69}$$

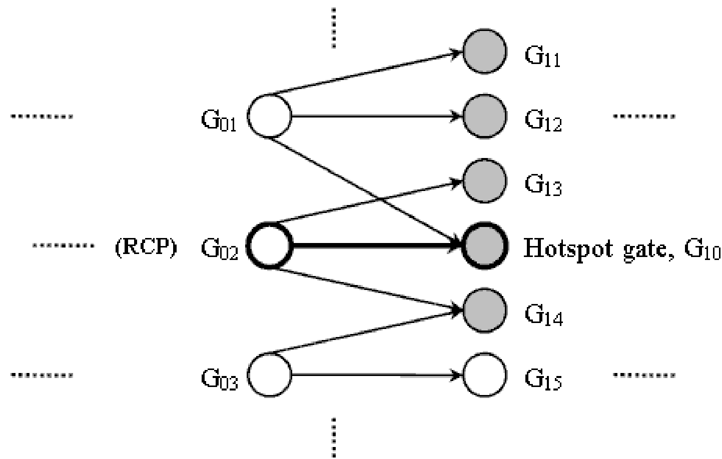


Figure 59. General digital logic network – dimension modulation.

Since in CMOS logic usually every input signal only drives one complementary pair, the change of propagation delay for the preceding gate G_{02} in RCP due to the increase of fan-

out capacitance keeps same as derived for inverter network. Thus, based on Equation (62) the increase of propagation delay at preceding gate is

$$\Delta\tau_{k-1} = 2AC_{OX}L^2(K_W K_L + FOK_W - K_W - FO) \tag{70}$$

Finally the overall change in propagation delay over RCP containing gate G_{02}-G_{10} becomes

$$\begin{aligned}\Delta\tau_{RCP} &= \Delta\tau_k - \Delta\tau_{k-1} \\ &= A\left\{\begin{aligned}&\left[\frac{mn}{2}\left(1-K_L{}^2\right)+2K_W\left(1-K_L\right)+2FO\left(1-K_W\right)\right]C_{OX}L^2 \\ &+\frac{L}{W}m\left(1-\frac{K_L}{K_W}\right)\left(nC_{DB}+C_{line}+C_{FO}\right)\end{aligned}\right\}\end{aligned} \tag{71}$$

Note in the equation FO is the fan-outs of the driving gate, and C_{DB}, C_{line}, and C_{FO} are parasitic capacitances of the driven gate (hotspot gate). All of these values will not be changed during DFR. Similarly, for non-RCPs such as G_{02}-G_{13} and G_{02}-G_{14}, by only increasing W the overall change in propagation delay is obtained as in the following equation.

$$\begin{aligned}\Delta\tau_{non-RCP} &= \Delta\tau_k - \Delta\tau_{k-1} \\ &= A\left\{\begin{aligned}&\left[2K_W\left(1-K_L\right)+2FO\left(1-K_W\right)\right]C_{OX}L^2 \\ &+\frac{L}{W}m\left(1-\frac{1}{K_W}\right)\left(nC_{DB}+C_{line}+C_{FO}\right)\end{aligned}\right\}\end{aligned} \tag{72}$$

Compared with Equation (71), it is easy to prove that $\Delta\tau_{RCP}<\Delta\tau_{non\text{-}RCP}$, if the hotspot gate has the similar complexity as the corresponding gate in non-RCP path. This indicates that as long as the redesign over the RCP path can give a speed same as that originally designed, all other involved paths (non-RCPs) in the circuit will maintain the original speeds as well. For the case that the hotspot gate has multiple inputs, the exactly same DFR processes can be performed around the other driving gates such as G_{01}.

Moreover, as shown in Figure 59, if all involved multi-input gates in non-RCPs such as G_{14} are uniformly resized for all transistors, the other input driving gates of those gates such as G_{03} will also be slowed down. This will inevitably spread the redesign further to gates like G_{15}. Since in CMOS logic one input normally only drives one complementary pair, and no channel length modification causing gate to slow down is made for non-hotspot gates, the transistor resizing described in Equation (72) is modified to only apply to those complementary pairs that are driven by the preceding gates that have been slowed by the hotspot gate, such as gate G_{01}. During the local DFR driving gates are slowed down due to increased gate capacitances of their driven gates, therefore by resizing the involved transistor pairs only, the redesign work can be effectively limited. For example, in Figure 59 gate G_{03} and G_{15} do not have to be involved. Thus, by only increasing the channel widths of one of its *m* complementary pairs the modified form of Equation (72) for non-RCPs is given below.

$$\Delta\tau_{non-RCP} = A\left\{ \begin{array}{l} \left[\begin{array}{l} 2K_W(1-K_L)+(m-1+2FO)(1-K_W)+ \\ \left(1-\frac{n}{2}\right)\left(\frac{1}{K_W}-1\right) \end{array} \right] C_{OX}L^2 + \\ \frac{L}{W}\left(1-\frac{1}{K_W}\right)(nC_{DB}+C_{line}+C_{FO}) \end{array} \right\} \tag{73}$$

A successful local DFR essentially requires the change of propagation delay $\Delta\tau$ for both RCP and non-RCP greater than or equal to zero. To verify the feasibility of the local DFR approach in complex logic circuits, a series of investigations have been conducted on Equation (71) and (73), and an even more optimistic result compared to the inverter network has been obtained. It shows that as the gates get more complex (larger *m* and *n*), the possible gain in propagation delay $\Delta\tau$ for both RCP and non-RCP also increase. This

is demonstrated in Figure 60. The process parameters are taken from AMI technology with aspect ratio of 6. In this case the channel length of the hotspot gate is increased by 50% (K_L=1.5) and the channel width is increased by 100% (K_W=2).

It is clearly seen that more gain in path propagation delay can be obtained for more complex gates after conducting local DFR approach, in other words, during the local DFR process that designates the same path speed as originally designed, more available range of device redesign can be obtained for reliability improvement with complex logic. It is also noticed that the simplest case *m*=1, *n*=2 is simply representing the inverter network. While in previous sections it has been shown that for the inverter network considerable room is available for conducting a successful local DFR, this well verifies the feasibility of the local DFR approach in all CMOS logic circuits for reliability improvement by showing an even larger available range of circuit redesign. All other conclusions drawn based on inverter network also stand correct for the general CMOS circuits.

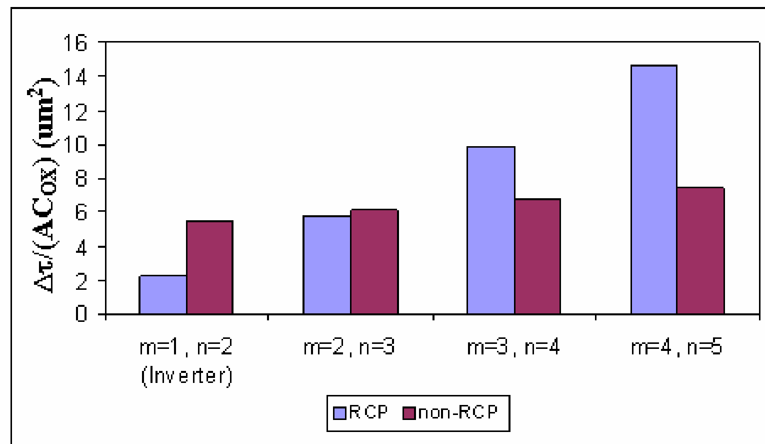


Figure 60. $\Delta\tau$ after DFR by dimension modulation for different gate complexities.

8.1.4. Circuit area involved

While conducting a successful DFR local redesign, the localization of redesign is a key condition, because it offers a very limited design work conducted in a very limited circuit area. This is critical to easily maintain the same circuit-level performance before and after the redesign, and save the power consumption and layout area due to transistor resizing as well.

The DFR redesign is mainly conducted around the reliability hotspot gate G_{10} as shown in Figure 59 locally. For improving reliability and compensating speed loss at gate G_{02} both channel length and width of G_{10} are modified. In addition, since all paths containing G_{02} are affected by this speed loss, all its fan-out gates G_{13} and G_{14} have to be redesigned as well for compensating the speed loss over the corresponding path. Since for those gates only the complementary pair driven by the gate that has been slowed down is resized, the redesign will not affect their other inputs such as G_{03} for G_{14}, and thus no further redesign is required along this line. On the other hand, in case that the hotspot G_{10} has multiple inputs, the same redesign process will have to be repeated for all its driving gates such as G_{01} leading to a series of gate redesign at G_{11} and G_{12}. Thus the maximum number of gates involved in DFR, N, is calculated in following equation.

$$N = n_{in} n_{FO} \tag{74}$$

where n_{in} represents the number of inputs for hotspot gate, and n_{FO} is the average fan-outs of the driving gate.

In practical circuit design, due to the restriction of parasitic capacitance on speed, the number of gate inputs and fan-outs is normally limited to 3~4 giving a total number of

gates involved around 10. Compared with totally hundreds of thousands gates in circuit, this certainly guarantees the localization of redesign.

8.1.5. Algorithm implementation

Based on the algorithm discussed above, the complete local DFR approach for CMOS digital circuits can be conducted accordingly, whenever a circuit design is possibly not reliable enough with existing reliability design rules, such as cutting-edge technologies and military systems. The main flow chart of the complete local DFR is shown in Figure 61.

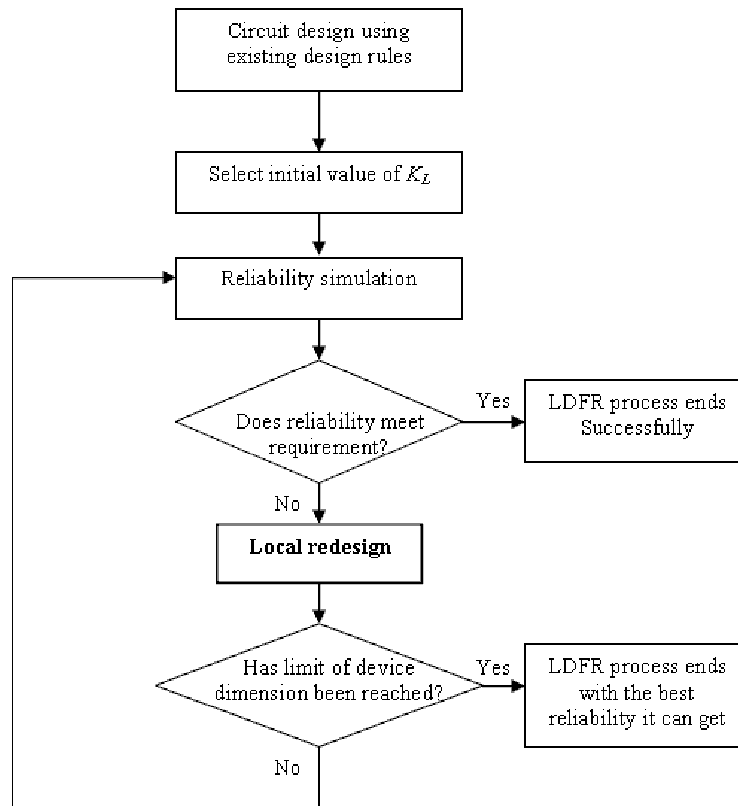


Figure 61. Local DFR by dimension modulation.

Based on current circuit design and process technology, a $K_L > 1$ is selected at the very beginning of the process. Since the common device degradation mechanisms are very sensitive to the change of channel length in short-channel transistors, a K_L between 1.1 and 1.5 is suggested as starting point for submicron technology. Using simulator ARET the reliability of current circuit design is simulated to determine if further redesign is needed. If the current circuit reliability does not meet requirement, the DFR redesign cycles are conducted. The detail procedures conducting the local redesign algorithm are described in Figure 62.

It is also noticed that a required reliability is not always guaranteed to be obtained due to the restriction of existing design and technology. This can also be due to the situation shown in Figure 56, in which as the channel length and width increase, somewhere the gain in speed drops below zero plane. Therefore, the DFR has to stop since the original performance can no longer be maintained. In case that the device resizing limit is reached, the DFR process will end with the most reliable design it can get. Further investigations are then required to solve the problem.

In DFR process shown in Figure 62, the reliability hotspot is identified by ARET tool, and it has to be conducted every time the circuit design is updated, in order to locate the most current hotspot that is responsible for circuit final failure. If it is indicated that the hotspot has moved to another gate from the one that is identified during the last cycle, the current redesign cycle will have to be taken at the new hotspot gate and the initial values of K_L and other parameters used in the algorithm will be reset. Equation (71) and (73) are used to calculate K_{W_RCP} and $K_{W_non\text{-}RCP}$.

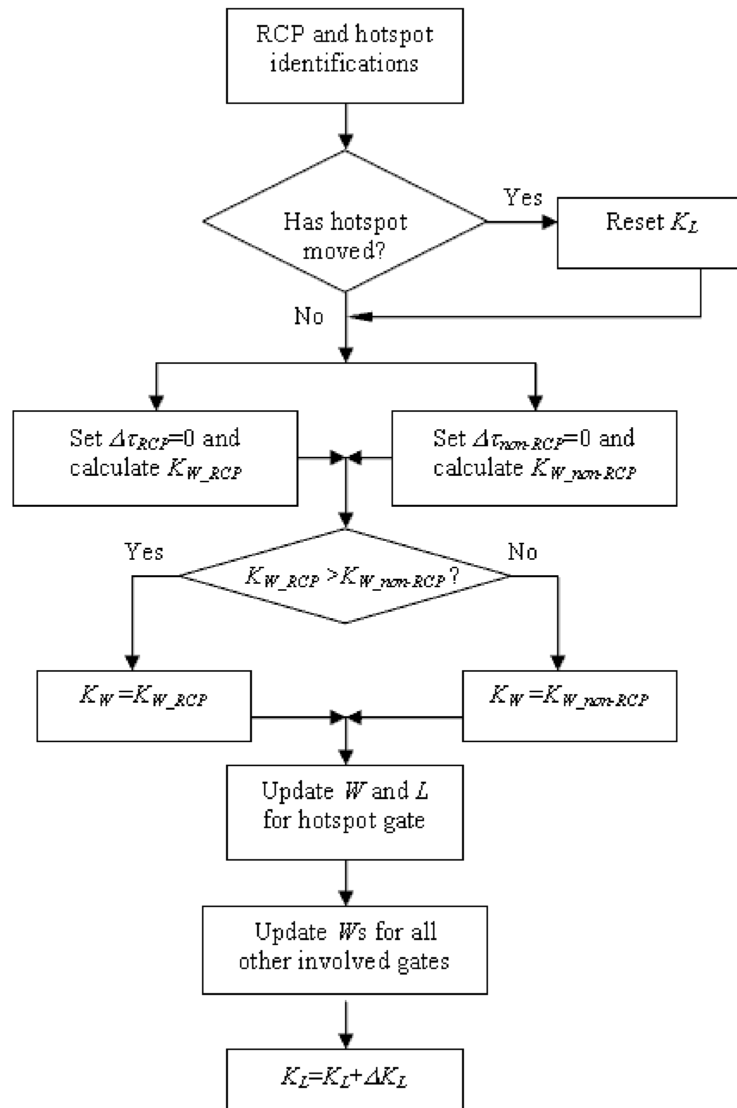


Figure 62. Local redesign process in DFR.

To maintain the original performance at all paths involved in DFR, the maximum of them are selected as K_W. After all involved devices are redesigned K_L is increased by ΔK_L for the next redesign cycle, if it is needed. While determining the increment ΔK_L, there is a trade-off between accuracy and computation time.

8.1.6. Discussions and trade-offs

Due to the increase of device dimension during local DFR process, the circuit layout area will be increased. Normally this leads to an increased cost. However, in local DFR such area changes only happen inside a very small part of circuit around the reliability hotspot(s). Compared with a large VLSI circuit containing millions of transistors, this should not affect the overall production cost significantly.

The power consumption will also go up due to the increased parasitic capacitances. However, the same argument for cost also applies to power consumption. The total increase of power consumption for the whole circuit will not be significant after DFR, especially when the circuits are continually becoming larger in technology scaling.

The DFR local redesign can also be challenging due to the non-uniformly sizing in the circuit. There could be some practical design issues for achieving the specific dimensions at hotspot(s), which are different from the rest of circuit components. This may be solved by alternative approaches, for example, increasing the channel length by putting several transistors with original sizes in series.

Due to the amount of reliability simulations involved, the time required to conduct such a DFR process is quite considerable. For a circuit having about 2400 gates, the total time needed is more than 40 minutes on an Ultra 10 workstation for a 30% lifetime increment. The time consumed in DFR can be effectively reduced by reliability-based

circuit partition technique, which results in a smaller searching space. The control of time step used in simulations is also critical.

8.2. DFR by signal modulation

8.2.1. Algorithm based on inverter network

A CMOS inverter network is demonstrated in the following figure. Gate N_4 is the reliability hotspot with RCP N_1-N_0-N_4. Gate N_0 is the preceding driving gate of N_4, and N_1 is the driving gate of N_0. There are other two gates, N_2 and N_3, shown as the other fan-outs of gate N_1. Effective parasitic transistor capacitances are shown for gate N_0 as gate-drain capacitance, C_{GD}, and drain-bulk capacitance, C_{DB}. C_{line} represents the interconnect capacitance. All transistors are assumed uniformly sized before DFR.

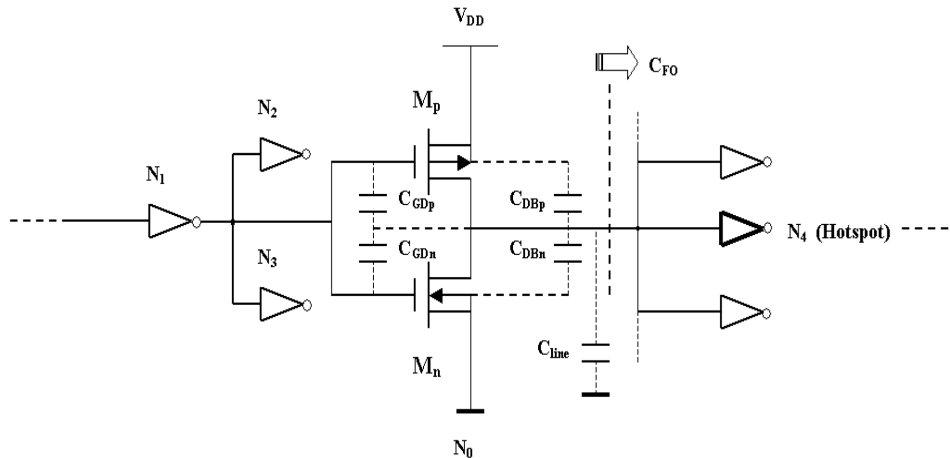


Figure 63. CMOS inverter network with effective capacitances.

Similar to the processes in dimension modulation approach, to analyze nMOS-related degradation, the inverter fall time delay t_{HL} is represented by τ_n in form of

$$\tau_n = \frac{1}{k_n \frac{W}{L}(V_{DD} - V_{tn})} \begin{bmatrix} (C_{GDn} + C_{GDp}) + (C_{DBn} + C_{DBp}) \\ + (C_{line} + C_{FO}) \end{bmatrix}$$
$$\approx \frac{1}{k_n (V_{DD} - V_{tn})} \frac{L}{W} \left[C_{OX} WL + (C_{DBn} + C_{DBp} + C_{line} + C_{FO}) \right] \tag{75}$$
$$= A\left[C_{OX} L^2 + \frac{L}{W}(C_{DBn} + C_{DBp} + C_{line} + C_{FO}) \right]$$

with

$$A = \frac{1}{k_n (V_{DD} - V_{tn})} \tag{76}$$

As described in the basic local DFR approach in Chapter 6, to shorten the transient time for the hotspot gate N_4, the preceding driving gate N_0 is accelerated by widening the channel widths. Based on Equation (75), if the channel width W at the gate N_0 is increased to $W \bullet K_W$, with the notation $C_0 = C_{DBn} + C_{DBp} + C_{line} + C_{FO}$, the gain of speed at the driving gate N_0 after DFR is thus obtained below.

$$\Delta \tau_0 = A\left(\frac{L}{W} C_0 - \frac{L}{K_W W} C_0' \right) = A \frac{L}{W}\left(C_0 - \frac{1}{K_W} C_0' \right) \tag{77}$$

For its preceding gate N_1 in RCP, the increase of propagation delay due to the increased fan-out capacitance is obtained in the following way. First, the capacitance due to fan-out load is given as

$$C_{FO} = FO(C_{OX} W_n L_n + C_{OX} W_p L_p)$$
$$= 2 FO C_{OX} WL \tag{78}$$

with the same FO in dimension modulation approach representing the number of fan-outs of gate N_1. Because the increase of propagation delay is solely determined by the increase of fan-out capacitance, the loss of speed at the preceding gate N_1 is obtained in following equation.

$$\Delta\tau_1 = A(C_{FO}' - C_{FO})\frac{L}{W} \tag{79}$$

From the circuit perspective, the slow-down of gate N_1 will also affect other paths such as the path N_1-N_2 and N_1-N_3, as shown in Figure 63. This may cause some of those paths fail to meet the timing requirement in the original design. To avoid taking this risk, all other fan-out gates of gate N_1 will have to be resized in a similar way as N_0. For simplicity of design, all fan-out gates of gate N_1 including N_0 are scaled up by the same K_W for channel width. Thus after DFR local redesign the total fan-out capacitance for gate N_1 becomes

$$C_{FO}' = C_{FO} K_W \tag{80}$$

which updates the loss in propagation delay at gate N_1 to

$$\Delta\tau_1 = 2AC_{OX}FOL^2(K_W - 1) \tag{81}$$

Finally, the overall gain of speed over RCP as well as non-RCPs involved in local DFR process becomes

$$\begin{aligned}\Delta\tau_{gain} &= \Delta\tau_0 - \Delta\tau_1 \\ &= A\left\{\frac{L}{W}\left(C_0 - \frac{1}{K_W}C_0'\right) + 2FOC_{OX}L^2(1 - K_W)\right\}\end{aligned} \tag{82}$$

In order to meet the two critical conditions required for conducting a successful local DFR in Chapter 6, $\Delta\tau_{gain}$ in Equation (82) must be greater than zero, so that for RCP the signal-switching period of the hotspot gate can be shortened giving an improved reliability (condition 1), and for all paths involved the propagation delay after DFR can be equal to or smaller than the originally designed (condition 2). This is further explored in the next section.

8.2.2. Algorithm feasibility

For performing a successful and practically feasible DFR, further investigations are conducted on Equation (82). By using Equation (82), Figure 64 shows the change of propagation delay as the channel width increases during DFR redesign for various technology feature sizes, where all process-related parameters are taken from AMI technology and an aspect ratio *W/L*=6 is assumed. By analyzing Equation (82), the same conclusion as for dimension modulation is obtained that the greater the interconnect and fanout capacitances are, the more possibility that a local DFR process can be conducted successfully. A fanout of 2 and a 10 fF average interconnect capacitance are assigned.

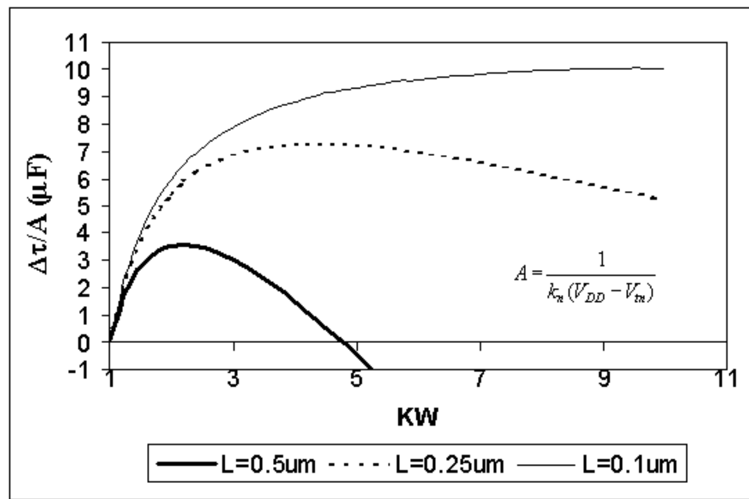


Figure 64. Change of delay in DFR by signal modulation for different feature sizes.

It can be clearly seen from the figure that for a range of feature sizes that are used today and possibly used in future, there is considerable room for channel width adjustment to have a positive gain in speed and conduct a successful local DFR. For example, for a half-micron technology a significant speed-up can be obtained with about

two times of device widening. The more $\Delta\tau_{gain}$ is obtained, the more improvement of reliability can be expected. To help selecting the proper K_W, the derivative of $\Delta\tau_{gain}$ with respect to K_W is taken below.

$$\frac{d(\Delta\tau_{gain})}{d(K_W)} = A\left(\frac{L}{W}C_0\frac{1}{K_W{}^2} - 2FOC_{OX}L^2\right) \tag{83}$$

By forcing it to zero, the K_W that gives the maximum gain of speed, therefore the maximum gain in reliability is obtained as in the following equation.

$$K_W = \sqrt{\frac{C_0}{2FOC_{OX}WL}} \tag{84}$$

For a technology with very small feature size, this value can be quite large. In the actual selection of K_W in DFR, the power consumption and layout area have to be considered as well.

Another observation from Figure 64 is that as the technology feature size gets smaller, the possible gain in speed ($\Delta\tau_{gain}$) becomes much higher representing a much more promising reliability improvement. In other words, as technology keeps scaling down, a much greater benefit of reliability improvement can be obtained by local DFR. This result is critical to the local DFR technique, as well as any other DFR approach because that the reliability technology is really driven by generation of new technology. So far the local DFR has been proven to be very effective for current cutting-edge technologies, especially when the feature size is becoming even smaller.

8.2.3. Algorithm with technology scaling

Figure 65 is the plot of the gain in propagation delay obtained after local DFR vs. feature size, using Equation (82) with K_W equals to 2 and all other parameters are same as in Figure 64.

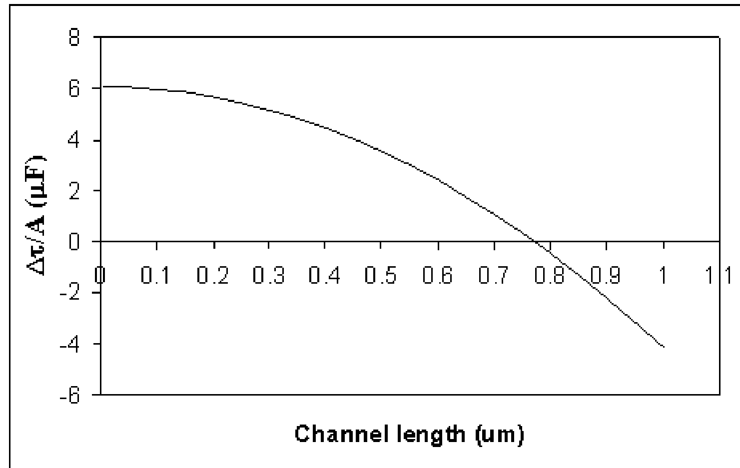


Figure 65. $\Delta\tau$ vs. channel length in DFR by signal modulation.

In Figure 65 when the feature size drops to about 0.8 μm, under the specific process, DFR starts to gain some speed during redesign. As the feature size continues to shrink, the gain of speed, or the reliability improvement expected, increases rapidly. This result indicates that the local DFR algorithm can be useful for the latest generation of technology and even more effective for future generations. The same observation can also be obtained from Figure 64, in which it is clearly shown that, as technology keeps scaling down, a much greater benefit of reliability improvement can be obtained using the local DFR approach by signal modulation.

8.2.4. Complete algorithm for CMOS Digital Family

For the general CMOS logic circuit shown in Figure 58, the local DFR algorithm for CMOS logic family can be developed similarly based on the algorithm for inverter network. With m representing the number of nMOS transistors in series in the longest chain and n representing the total number of transistors connected to the output line, Equation (75) is re-written with assumption $C_{DBn} \approx C_{DBp} = C_{DB}$ as in the following equation.

$$
\begin{aligned}
\tau_n &= \frac{m}{k_n \dfrac{W}{L}(V_{DD}-V_{tn})}\left[\frac{n}{2}C_{OX}WL + nC_{DB} + (C_{line}+C_{FO})\right] \\
&= Am\left[\frac{n}{2}C_{OX}L^2 + \frac{L}{W}\left(nC_{DB}+C_{line}+C_{FO}\right)\right]
\end{aligned} \tag{85}
$$

again with

$$
A = \frac{1}{k_n(V_{DD}-V_{tn})} \tag{86}
$$

Here a schematic of general CMOS logic circuit is shown in Figure 66, where gate G_{20} is the hotspot in RCP G_{02}-G_{10}-G_{20}. Based on the local DFR algorithm discussed in previous sections, the driving gate, G_{10}, of the hotspot gate will be widened, while the other fan-out gates of the driving gate of G_{10}, such as gate G_{12} and G_{13}, will also be resized for compensating the speed loss at gate G_{02} due to increased fan-out capacitance.

Thus, after resizing the fall time constant of gate G_{10} becomes

$$
\tau_n{}' = Am\left[\frac{n}{2}C_{OX}L^2 + \frac{L}{K_W W}\left(nC_{DB}+C_{line}+C_{FO}\right)\right] \tag{87}
$$

which gives the updated speed gain of gate G_{10} after DFR

$$
\begin{aligned}
\Delta\tau_0 &= \tau_n - \tau_n{}' \\
&= Am\frac{L}{W}\left(1-\frac{1}{K_W}\right)\left(nC_{DB}+C_{line}+C_{FO}\right)
\end{aligned} \tag{88}
$$

Since in CMOS logic every input signal only drives one complementary pair just like the situation in inverter network, the speed loss at the driving gate G_{02} in RCP due to the increase of fan-out capacitance remains the same as derived in Equation (81), which gives the updated final overall change in propagation delay over RCP G_{02}-G_{10}-G_{20} as

$$
\begin{aligned}
\Delta\tau_{RCP} &= \Delta\tau_0 - \Delta\tau_1 \\
&= A\left\{2FO(1-K_W)C_{OX}L^2 + \frac{L}{W}m\left(1-\frac{1}{K_W}\right)\left(nC_{DB}+C_{line}+C_{FO}\right)\right\}
\end{aligned} \tag{89}
$$

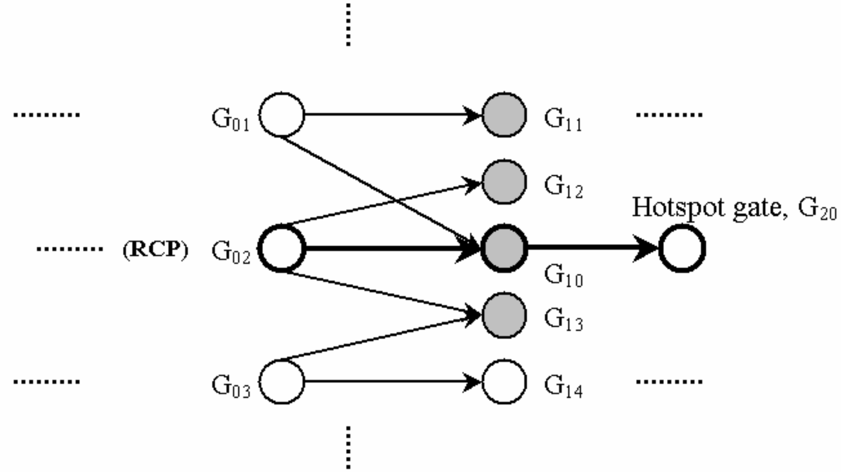


Figure 66. General CMOS logic network.

Similarly, the K_W at that $\Delta\tau_{RCP}$ achieves its maximum value is obtained by forcing the derivative of Equation (89) with respect to K_W equal to zero, described in the equation below.

$$K_W = \sqrt{\frac{m(nC_{DB} + C_{line} + C_{FO})}{2FOC_{OX}WL}} \tag{90}$$

Unlike in inverter network where every gate only has one input, in complex logic circuits a gate may have multiple inputs from different driving gates, such as driving gates G_{02} and G_{03} to G_{13} in Figure 66. Thus, the uniform resizing of all transistors in gate G_{13} will also slow down gate G_{03}, which is not a driving gate of G_{10}. This, however, will make gate G_{14} to be resized just like G_{13}, which continues to affect other driving and driven gates until it reaches a single-input gate. One essential idea in the local DFR is to keep the redesign area as small as possible, by which it can be easier to maintain the original circuit performance and the design work involved is minimized. Therefore, to restrict the redesign area, gates in non-RCPs such as G_{12} and G_{13} are resized for involved complementary transistor pair only instead of uniform resizing. By doing this, the

redesign work is effectively limited around the only uniformly resized gate G_{10}. Gates like G_{03} will not be slowed and gates like G_{14} thus do not have to be resized. Finally, the total change in propagation delay over non-RCPs is given by

$$\Delta\tau_{non-RCP} = A\left\{\begin{array}{l}\left[(m-1+2FO)(1-K_W)+\left(1-\frac{n}{2}\right)\left(\frac{1}{K_W}-1\right)\right]C_{OX}L^2 + \\ \frac{L}{W}\left(1-\frac{1}{K_W}\right)(nC_{DB}+C_{line}+C_{FO})\end{array}\right\} \qquad (91)$$

To verify the feasibility of the algorithm, the change of propagation delay after conducting DFR is plotted using Equation (89) and (91) for circuits with different complexities, as shown in Figure 67. All process parameters are taken from AMI technology with aspect ratio W/L=6 and K_W=2.

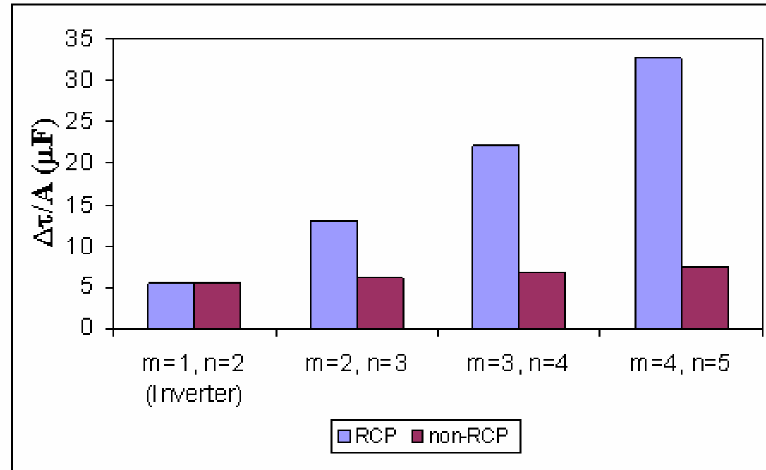


Figure 67. Gain of speed after local DFR for circuits with different complexities.

The figure clearly shows that with an increasing circuit complexity, the gains in propagation delay over both RCP and non-RCPs increase. With the gain in propagation delay over RCP representing how much the reliability can be improved, this indicates that

the local DFR works more effectively for more complex circuits. Another important observation is that the first case m=1/n=2 is simply the inverter network. With the feasibility of the local DFR algorithm for inverter network well verified in previous sections, this certainly proves that this DFR technique is able to work with all CMOS digital circuits and offer a promising reliability improvement. It also indicates that all other conclusions based on the inverter case stand correct for the whole CMOS logic family.

8.2.5. Circuit area involved

During the local DFR minimizing the involved circuit components is critical to maintain the circuit-level performance before and after the redesign, and save the power consumption and layout area due to transistor resizing. As shown in Figure 66, DFR redesign is mainly conducted around the reliability hotspot gate G_{20}. For improving reliability of G_{20} the transistor is widened at its driving gate G_{10}, which slows down its own driving gate G_{02}. Since all paths containing G_{02} are affected by this speed loss, all its other fan-out gates G_{12} and G_{13} are to be redesigned too for compensating the speed loss over the corresponding path.

For these resized gates in non-RCPs only the complementary pair affected by redesign is resized, such as the transistor pair in gate G_{13} that is driven by G_{02}. Therefore, the other driving gates of these resized gates like gate G_{03} of G_{13} will not be slowed down and thus no further redesign spreading to other components beyond gate G_{13}.

On the other hand, in case that the uniformly resized driving gate of the hotspot, such as G_{10}, has multiple inputs from gate G_{01}, the same redesign process will have to be repeated for all driven gates of G_{01} leading to a resizing at G_{11}.

The maximum number of gates involved in DFR, N, is thus calculated in following equation.

$$N = n_{in}' n_{FO} \tag{92}$$

where n_{in}' represents the number of inputs for the driving gate of the hotspot gate in RCP, and n_{FO} is the average fan-outs.

Again, the number of gate fan-outs is normally about 3~4 in practice. With 5 inputs for the driving gate of the hotspot, the total number of gates involved in DFR redesign is around 15 to 20, which is a really small number compared with totally number of gates in VLSI circuit.

8.2.6. Algorithm implementation

The execution flow of the DFR process by signal modulation is shown in Figure 68. This local DFR is basically used for the circuit designed by the new technology for that the existing design and process rules cannot give the satisfying reliability. After the initial design, as shown in Figure 68, the K_W that gives the maximum speed gain is calculated using Equation (90). As the algorithm is executed, K_W is increased by a selected step to approach to the desired reliability. However, once K_W exceeds K_{W_Max}, further increment will be meaningless since the speed gain will start to drop as shown in Figure 64.

Another process terminator is the speed change of non-RCPs, $\Delta\tau_{non\text{-}RCP}$, after DFR calculated by Equation (91). To meet the critical condition 2 of the local DFR approach, all involved paths must have at least the same speeds as in the original design. Therefore, every time K_W is updated, besides checking with K_{W_Max}, $\Delta\tau_{non\text{-}RCP}$ has to be checked as well to make sure the circuit is not performing any worse than original design.

ARET reliability simulation and hotspot identification are essential in the local DFR. After a redesign cycle is done, the circuit is re-evaluated by ARET to see if the required reliability has been obtained. Inside the "local redesign" step in Figure 68, the hotspot identification function is called at the beginning of every redesign cycle so that the right part of the circuit is redesigned. All channel width updates are performed in this step too.

8.2.7. Discussions and trade-offs

The devices are widened during local DFR process, which will cause the circuit layout area to be increased. For the local DFR algorithm based on signal modulation, the involved circuits area and components have been effectively minimized. Compared with a large VLSI circuit containing millions of transistors, this increase in area will not affect the overall circuit quality indices such as cost significantly. In fact, it is believed negligible in most cases.

The power consumption is another issue for widened devices. However, the same argument for layout area also applies to power consumption. The total increase of power consumption for the whole circuit will not be significant, especially when the local DFR gets more efficient in more complex and large-scale circuits.

For digital logic circuits, how long the propagation delay falls beyond the required boundary is used to judge a circuit failure. An extra gain in this algorithm is that, with the transient time shortened for the hotspot gate, the overall propagation delay of the RCP is also reduced, which makes the path even more reliable by increasing the difference between the maximum delay required and the initial delay.

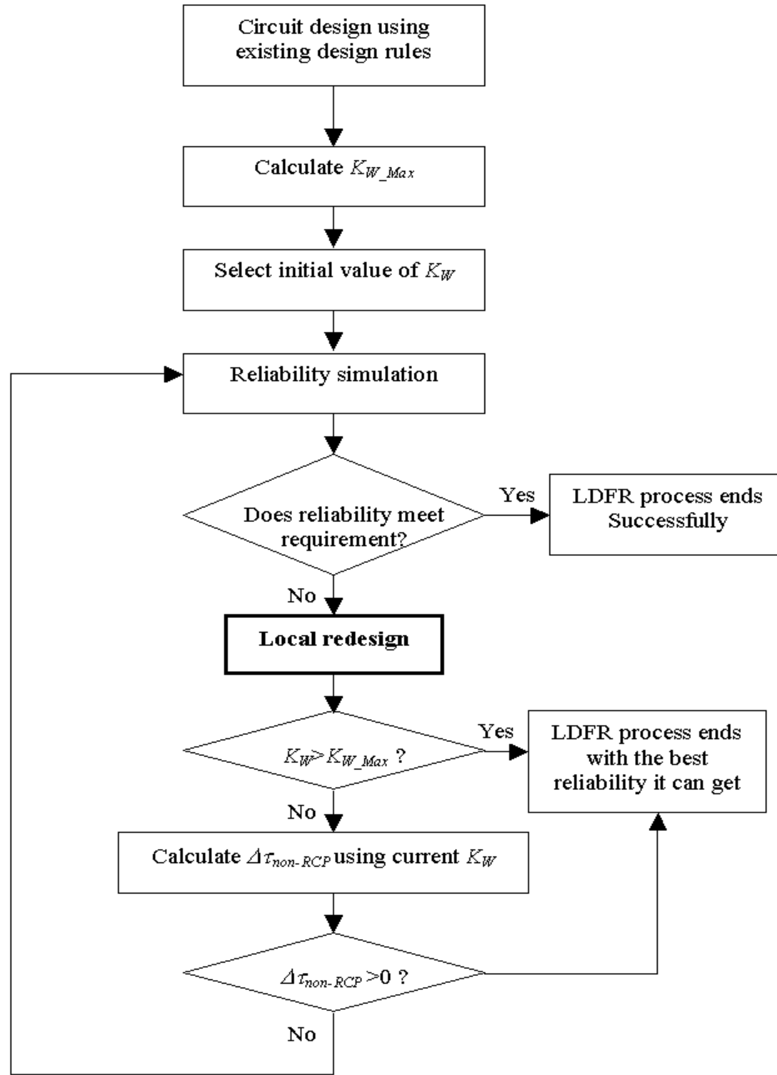


Figure 68. DFR algorithm by signal modulation.

Also, since the DFR by signal modulation essentially release device degradations by shortening the signal transient period at the reliability hotspot, the current density inside the device is increased. This in turn may accelerate the degradation in some degree, which is one of the reasons that the signal modulation is not as powerful as the dimension modulation. This conclusion is verified by the experimental results later in this thesis.

8.3. Other DFR approaches

In addition to dimension modulation and signal modulation, some other approaches may also be used for local DFR as long as they can meet the two key conditions given in Chapter 6. As an example, the power supply modulation approach is briefly introduced as follow.

The use of dual-power supply has been accepted by the semiconductor industry to lower the power consumption. Except that it costs the extra design complexity, this technology can be very promising in design-for-reliability as well. This local DFR algorithm is especially effective for device degradation caused by gate oxide wear-out. It is demonstrated using the figure below.

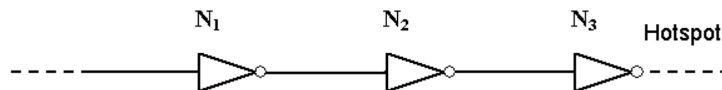


Figure 69. Schematic for power supply modulation.

As shown in Figure 69, if inverter N_3 is identified as the reliability hotspot, a lower power supply $V_{DD}'=K_V \bullet V_{DD}$ is used at the driving gate N_2, where the voltage scaling factor K_V is less than one. However, the lower power supply will give a slower switching

at N_2. This may be compensated by the similar approach used in signal modulation – device widening.

At gate N_2 after DFR redesign, from equation (51) the voltage-dependent factor is written as

$$s_n' = \frac{2(V_{tn} - V_0)}{V_{DD}' - V_{tn}} + \ln\left(\frac{2(V_{DD}' - V_{tn})}{V_0} - 1\right) \tag{93}$$

and based on Equation (56) the redesigned fall time constant is

$$\begin{aligned} \tau_n' &= A\left[C_{OX}L^2 + \frac{L}{WK_W}(C_{DBn} + C_{DBp} + C_{line} + C_{FO})\right] \\ &= \frac{1}{k_n(V_{DD}' - V_{tn})}\left(C_{OX}L^2 + \frac{L}{WK_W}C_0\right) \end{aligned} \tag{94}$$

Thus, the fall time change due to DFR local redesign at gate N_2 can be described as

$$\Delta t_{HL_2} = s_n\tau_n - s_n'\tau_n' = f_2(K_W, V_{DD}') \tag{95}$$

The speed loss at the driving gate of gate N_2 is calculated from Equation (60) as following.

$$\Delta t_{HL_1} = s_n \Delta\tau_1 = f_1(K_W) \tag{96}$$

The total change of propagation delay over the path is thus obtained in following equation.

$$\Delta t_{HL} = \Delta t_{HL_2} - \Delta t_{HL_1} = f(K_W, V_{DD}') = f(K_W, K_V) \tag{97}$$

By forcing Equation (97) equal to zero, the original propagation delay can be maintained.

During the application of the algorithm, based on simulation an initial K_V is selected. Then by solving Equation (97) equal to zero with K_V known, a corresponding K_W is obtained, which will be used for device resizing. This DFR cycle is repeated with K_V

decreasing by a step, until the desired reliability is obtained, or the limit of K_W or K_V is reached.

Besides the increase of area and power dissipation, the stand-by leakage at gate N_3 in Figure 69 due to the lowered gate voltage is induced. Depending on the set-up of power supplies and the length of RCP, it is possible that this leakage becomes more serious.

8.4. Experiments

8.4.1. SPICE check

In order to confirm that the circuit timing performance can indeed be maintained after local DFR redesign, a circuit shown in Figure 70 is designed with TSMC 0.18μm technology, and HSPICE simulations are run to check the propagation delay model using in DFR.

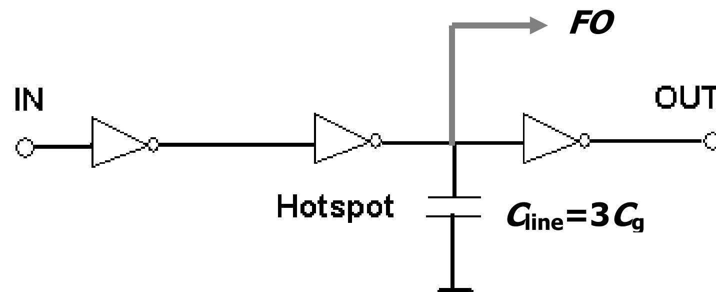


Figure 70. Inverter network for SPICE simulation.

The hotspot gate is re-sized with K_L=1.2 and K_W=1.7, for reliability improvement and performance maintenance during local DFR. An average interconnect capacitance of 10 fF is assigned. The propagation delay over the path is simulated by HSPICE for various fanout situations of hotspot gate. The results are plotted in Figure 71.

HSPICE results:

TSMC 0.18um KL=1.2 KW=1.7 C_line=3C_g

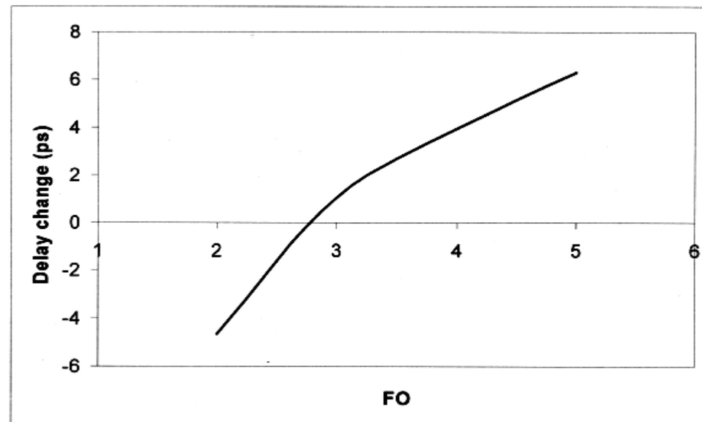


Figure 71. Verification of delay model used in local DFR.

As seen from the figure, for maintaining circuit performance, the fanout of hotspot gate is really a critical factor in local DFR approach to improve the chance of success. With a fanout greater than 3, it is very easy to maintain the circuit speed while holding a 20% of channel length increase for reliability improvement. In most of circuits, it is quite normal that part of its gates have much higher fanouts than the rest of circuit. This is the uneven distribution of potential reliability, which is discussed in detail based on ISCAS benchmarks later in this section. Therefore, it is clearly indicated that, by proper transistor sizing it is possible to maintain the circuit timing performance accommodating the change caused by redesign for reliability.

8.4.2. CMOS inverter chain

To verify the local DFR approaches, an inverter chain containing six (6) CMOS inverters in series is designed and shown as follow. Inverter chain has been widely used in digital circuits such as buffer structures.

Figure 72. CMOS Inverter chain.

Originally all transistors in the path have the same aspect ratios of 4.5µm/0.6µm. A load capacitor of 2 pF is assigned to the last gate, which gives the gate the largest capacitive load. The interconnect capacitances are ignored due to the simple connection. To accelerate the degradation, a drain voltage of 7 volts was used, and the temperature was maintained at –40°C through the test. The circuit was laid out using AMI C5N technology. The layout of one single inverter is shown in Figure 73.

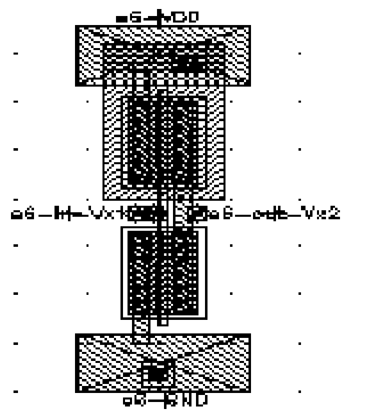


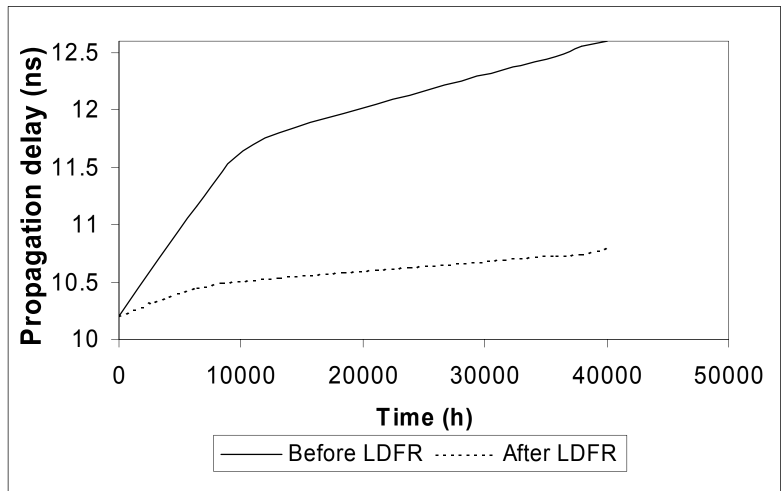
Figure 73. CMOS inverter layout.

Due to the additional load capacitor, the last inverter was identified as the reliability hotspot for this circuit by ARET. Thus, a DFR process was conducted locally around the last gate in the chain. In this case we only performed one DFR cycle to show the effect in

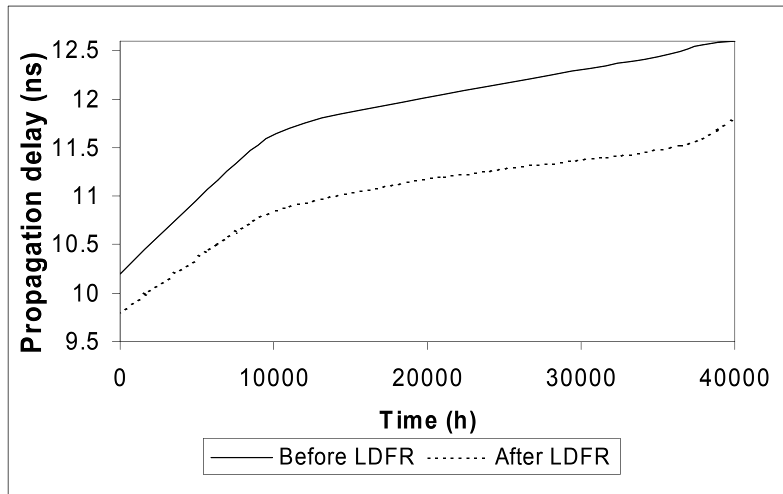
reliability improvement. By dimension modulation, after DFR the channel length of the last gate was increased from 0.6 μm to 0.75 μm, and the channel width was also increased from 4.5 μm to 6 μm. By signal modulation, the channel width of the second gate from the right was increased from 4.5 μm to 9.45 μm. The degradations (increases) of the propagation delay over the path were then simulated by ARET under hot-carrier effect. The results are shown in Figure 74.

From Figure 74 it is observed that in 40000 hours, the overall propagation delay increased from 10.2 ns to about 12.6 ns for the original design. After DFR by dimension modulation, the propagation delay increased only to 10.6 ns, which equals an 83% of improvement in reliability. After DFR by signal modulation, the propagation delay of the redesigned chain showed an increase from 9.8 ns to 11.6 ns, giving a 25% of improvement. It is clearly seen that, although more area cost in signal modulation redesign, it is not as effective as the dimension modulation approach in terms of reliability improvement.

The power dissipation was about 0.15 nW for the original circuit, by simulation using SPICE. Since only 1 of 6 gates was redesigned with increased channel length and width, no significant increase in power dissipation was observed in both approaches after local DFR.



(a) Dimension modulation



(b) Signal Modulation

Figure 74. Degradation of propagation delay over the inverter chain.

8.4.3. ISCAS benchmark circuits

DFR approaches by both dimension modulation and signal modulation were also applied to ISCAS benchmark circuits. The circuits involved were all combinational static logic circuits. The information about the benchmark circuits involved in this experiment are listed in Appendix B. Experiments were conducted for achieving a 30% and a 50% reduction of degradation in propagation delay in 40,000 hours, respectively. AMI C5N process parameters with a 0.5μm feature size were assumed during simulations. An average interconnect capacitance of 20 fF is assumed. For an accelerated degradation, a 6.0 volts power supply under –40 °C was used. The hot-electron degradation was assumed the only failure mechanism taking effect. All reliability simulations were accomplished using validated simulator ARET. The power consumptions were simulated by SPICE. The primary results are shown in Table 7 for dimension modulation, and Table 8 for signal modulation, respectively. For both tables the reliability achievements are listed first for involved benchmark circuits, followed by the trade-offs presented in the last two columns.

For both dimension and signal modulation approaches, the data clearly show that the reliabilities of the benchmark circuits, which are measured by the degradation in speed (propagation delay), have been successfully improved by designated 30% and 50%, compared to the original designs. It is also easy to understand that, to achieve a higher reliability goal such as the 50% over 30% improvement, more DFR cycles (hotspots) were needed, requiring more area, power and execution time.

Table 7. Results of DFR by dimension modulation on benchmark circuits.

Circuit/Total Gates	DFR cycles conducted		Increase of Device Area (%)		Increase of Pwr Dissip. (%)	
	30% improv.	50% improv.	30%	50%	30%	50%
C432/160	14	36	≈ 3.01	≈ 9.06	2.631	7.213
C880/383	25	—	≈ 1.67	—	2.079	—
C1908/880	7	14	≈ 0.46	≈ 0.91	0.33	0.581
C2670/1193	8	18	≈ 0.39	≈ 0.86	0.278	0.551
C3540/1669	5	6	≈ 0.17	≈ 0.21	0.124	0.131
C6288/2406	—	—	—	—	—	—

Table 8. Results of DFR by signal modulation on benchmark circuits.

Circuit/Total Gates	DFR cycles conducted		Increase of Device Area (%)		Increase of Pwr Dissip. (%)	
	30% improv.	50% improv.	30%	50%	30%	50%
C880/383	29	39	≈ 3.35	≈ 4.85	2.242	2.786
C1908/880	32	41	≈ 2.09	≈ 2.68	1.509	1.701
C3540/1669	42	—	≈ 1.45	—	1.044	—
C6288/2406	—	—	—	—	—	—

Another interesting observation is the dependence on the structure of the circuit. Generally, more DFR cycles are not necessarily required for a relatively more complex circuit to achieve certain amount of reliability improvement. Instead, the specific circuit structure has a significant impact on the local DFR process. This is due to that in both dimension and signal modulations the fanout at the gate under redesign really plays a critical role in terms of success and efficiency of the DFR process. According to the equations in the delay models in previous sections, with more gates having fanouts much larger than the rest of gates, the local DFR can be accomplished more easily and successfully. Based on the information given in Figure 75 about the distribution of gate fanouts, this has been clearly proven. For example, circuit C6288 did not return a successful result for the cases of both dimension and signal modulations. This is due to the fact that, in C6288 the fanouts are very evenly distributed giving no significant difference to gate structure-related reliability, which makes the local DFR very difficult.

This is seen even more clearly when compared with the results from gate C3540 and C1908.

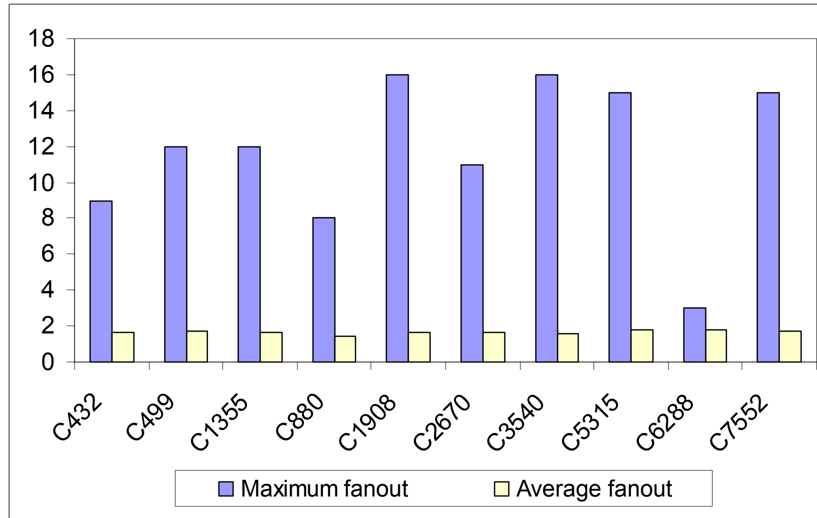


Figure 75. Fanout distributions of ISCAS benchmark circuits.

The increase of device area was approximately estimated based on the modifications of channel length and channel width, while the power consumption was simulated by SPICE. Both show clear increases in Table 7 and Table 8 as the trade-offs of the DFR local redesign. As the result of increased device dimension for part of the circuit, more area and power increments are cost to accomplish a successful local DFR for a higher reliability goal, which requires more DFR cycles conducted.

Overall, the area increment is below 4% for the 30% degradation reduction, below 10% for the 50% degradation reduction, and the power increase is below 3% to achieve the 30% reliability improvement, below 8% to achieve the 50% reliability improvement, even for the complex circuit with more than 2000 gates. The trade-off is completely acceptable compared to the gain in reliability.

Between the two approaches, the signal modulation approach costs more in both area and power than the dimension modulation to achieve a same reliability goal. In other words, the DFR by dimension modulation is more powerful than by signal modulation. This is demonstrated in Figure 76 with sorted data on area required by different circuits for different reliability goals, where the area increments cost by signal modulation are higher than by dimension modulation in all cases.

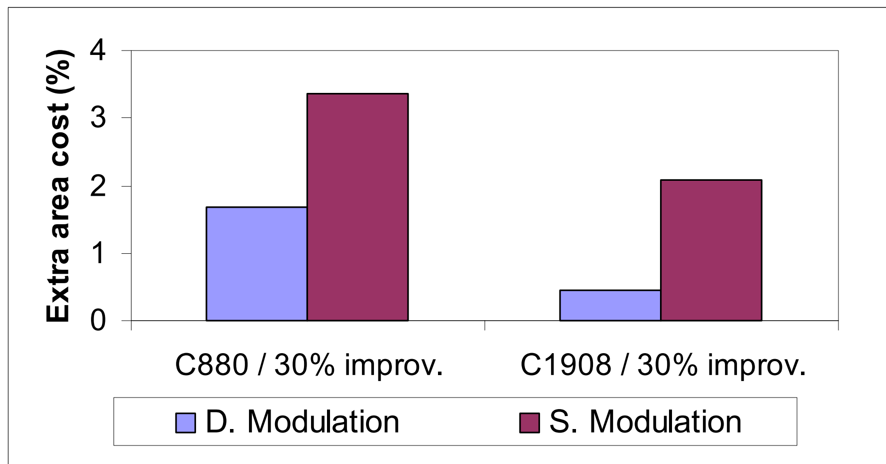


Figure 76. Comparison between dimension modulation and signal modulation.

One can also notice that for some circuits such as C6288, there is no successful result returned. This is due to the limit of the local DFR algorithms in order to maintain circuit performance, as discussed before, which indicates that depending on different circuit structures, it is not always guaranteed to achieve the given reliability goal with a performance at least same as the original. Otherwise, the circuit will be slowed down. The redesign process has to abort or stop with the best reliability that can be obtained when the physical/structural limitation of the device is reached, or the original circuit performance simply cannot be maintained. This could happen when a circuit has evenly

distributed reliability factors such as fanout and switching activity among its components. Such circuits usually have optimized reliabilities and thus do not need to be revised by local DFR, which works the best for the circuit situation where part of circuit is having much better or worse reliability than the other part.

CHAPTER 9

DFR FOR ANALOG CIRCUITS

Experiencing the same shrinkage of reliability safety margin under technology scaling, analog circuits faces a much more complicated situation than digital circuits to improve reliability by design approaches. There no longer are the signals with only two logic levels, instead, the continuous signals everywhere. There are so many circuit level specs describing the performance, and each of them may change in a different way in case of device degradations. Therefore, the design-for-reliability in analog circuits has to be performed very carefully, and the corresponding algorithms may be developed on a case-by-case base, although the principle of the local DFR still works as guidance.

9.1. Basic DFR approach

First, an analog circuit is specified by a group of circuit level performance specs, described by space P, and a group of device parameters that can be modified, described by space X. The original design point is thus located in the space P, which is n-dimensional with n denoting the total number of circuit key specs. In the corresponding m-dimensional parameter space X where m represents the number of modifiable device parameters, it is assumed that there exists an area so that the circuit will stay at the original design point in space P as long as the parameter point in space X stays inside this area. The goal of DFR is to move the parameter point of reliability hotspot to a reliability-favored direction while maintaining the circuit original design point.

To perform the local DFR to such an analog circuit, a mapping from the circuit performance space to the design parameter space of the hotspot component is first created. A 2-D mapping is demonstrated in Figure 77.

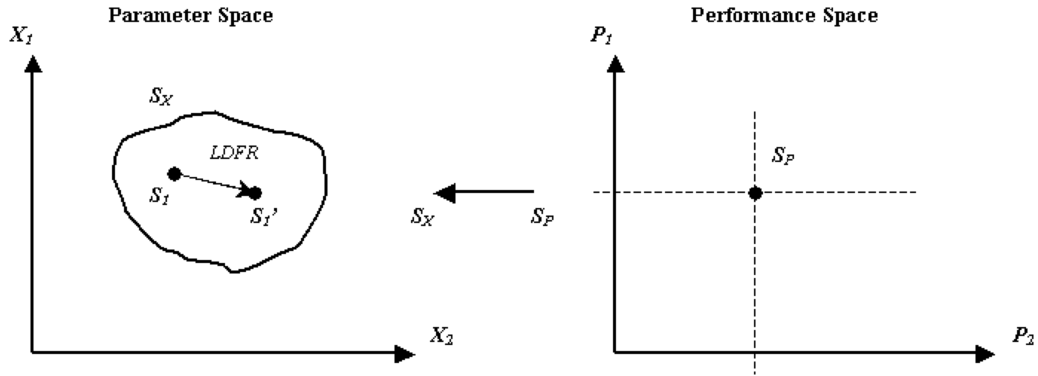


Figure 77. Mapping between circuit performance and device parameter.

In Figure 77, it is assumed that the performance specification group contains two items, P_1 and P_2, and the design parameters of the hotspot component are X_1 and X_2. The original performance design point is specified by S_P in performance space, which is by mapping corresponding to the area S_X in parameter space.

During DFR, the original parameter design point $S_1 = (X_1, X_2)$ of the hotspot component moves to S_1' in a reliability-favored direction. This makes the hotspot more reliable and, according to Equation (49), will release the overall degradation of circuit key performance specification, while the circuit performance is still maintained at S_P.

The major restriction in this approach is the set S_X. In order to maintain the original circuit performance, the redesign of the hotspot has to be conducted inside S_X. The larger the area of S_X, the more reliability improvement can be expected during DFR. In an extreme case where the area of S_X becomes a single point, more surrounding components

of the hotspot will have to be involved, which means the dimension of the parameter space increases, to obtain an available area of S_X.

The mapping between the parameter space and the performance space can be complicated, especially when the circuit performance specification group becomes very large and more circuit components have to be involved to conduct DFR. Simulation or synthesis technique may be employed to determine the parameter set S_X at a price of large amount of computation time.

9.2. Implementation

Based on the approach given above, the key step in local DFR for analog circuits is creating the mapping between the parameter space and the performance space. This can be very complicated and the development of certain algorithm much depends on the circuit architecture with the assistance of other circuit techniques such as circuit synthesis and simulation. A high-level description of the possible algorithms is presented in Figure 78.

In the chart, the key circuit performance specs are selected to determine the circuit failure and lifetime. The parameter set actually decides how much room for redesign is available during the DFR. After redesign the updated circuit reliability (lifetime) has to be re-evaluated to make decision to go or not to go.

The first redesign phase is to enhance the reliability of the hotspot components, based on which the circuit overall reliability is expected to be improved. This is done according to the certain failure mechanisms involved. For example, under hot-carrier degradation the device channel length can be increased to improve the reliability. It is also restricted by the available parameter set.

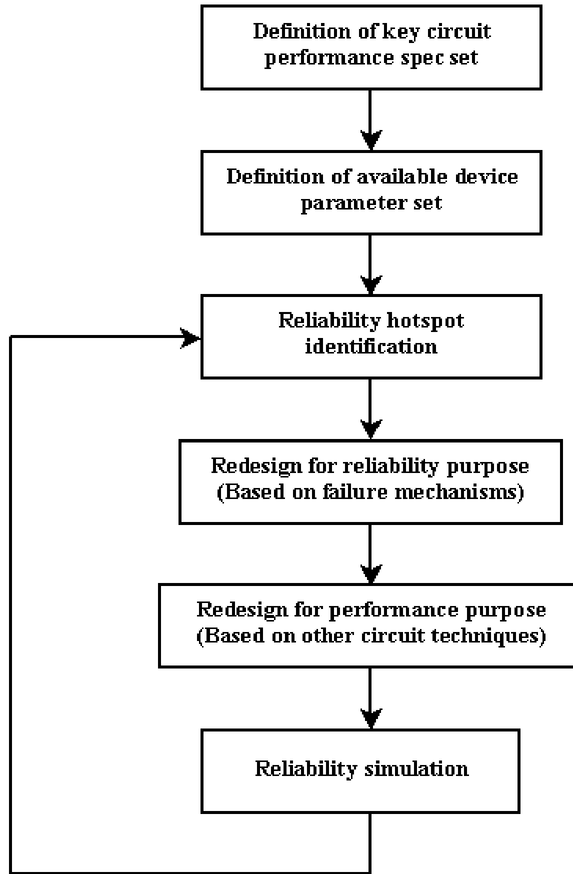


Figure 78. DFR for analog circuits.

The second phase of redesign is to compensate the change of circuit performance due to the first phase. The goal is to make the circuit design point after DFR stays in its original position in the performance space. Apparently this is to satisfy the condition 2 of the basic local DFR approach. As an example, a circuit synthesis technique is briefly discussed here [47]. A simulation-based transistor level analog circuit sizing method was proposed in that work. It is based on the evaluation of a response surface model and its update for accuracy. Some initial results using this method are given in the next section.

Several issues need to be addressed when an algorithm is to be developed. First, depending on certain circuit technique and available modifiable parameter set of the

device, the mapped parameter set S_X in Figure 77 may not be large enough, in an extreme case, it can even be a single point giving no solution at all. Thus, to gain more room for conducting DFR, the available device parameter set will need to be expended, if it is possible, which increases the dimension of parameter space and requires heavy computation load. Second, due to the variety of the structure and the specification of analog circuits, after a DFR local redesign the overall circuit reliability is not guaranteed to improve. This is because that besides the device structural parameters, the electrical stress also plays an important role in reliability. The redesign in the local DFR will most probably change the stress distribution in the circuit, and an uncertainty that this change of stress condition may or may not degrade the circuit even more is thus generated. The reliability simulation cycle in the proposed DFR approach can serve as a checker for this, but cannot guide the process to a guaranteed correct direction.

9.3. Experiments

Since the mapping between the parameter space and the performance space is critical to DFR for analog circuit, some experimental evidences on a two-stage op-amp are supplied in Table 9. These results are obtained from the extended work of [47].

Table 9. Op-amp deigns with different sizing.

Circuit design	W6 (μm)	W5 (μm)	W4 (μm)	W3 (μm)	W2 (μm)	W1 (μm)	Open-loop gain (v/v)
1	301.014	364.953	455.296	60.3	88.6	79.9	1933.19
2	451.204	477.525	424.124	351.253	404.08	152.876	1935.39
3	167.492	164.864	394.947	75.9	166.435	277.415	1934.69

Table 9 shows three op-amp designs obtained using the synthesis approach in [47]. It is seen that they all have very close open-loop gains, but different sizing of W1 to W6, which are defined in the Spectre netlist script given in Figure 79 from the same source of [47]. This result clearly demonstrates that, by using some kind of circuit techniques such as circuit synthesis, it is possible to create the mapping, or the approximate mapping like shown in this example, between the two spaces.

```
subckt twostage_opamp_ami06 _net1 _net0 Vdd Vo Vss

  I0 (V2 Vss) isource dc=Ibias m=1 type=dc
  M16 (Vo V2 Vdd Vdd) ami06P w=W4 l=600n as=W4 * 2.5 * (600n) ad=W4 * \
      2.5 * (600n) ps=(2 * W4) + (5 * (600n)) pd=(2 * W4) + (5 * (600n)) \
      m=1 region=sat
  M15 (V1 V2 Vdd Vdd) ami06P w=W3 l=600n as=W3 * 2.5 * (600n) ad=W3 * \
      2.5 * (600n) ps=(2 * W3) + (5 * (600n)) pd=(2 * W3) + (5 * (600n)) \
      m=1 region=sat
  M18 (V2 V2 Vdd Vdd) ami06P w=W6 l=600n as=W6 * 2.5 * (600n) ad=W6 * \
      2.5 * (600n) ps=(2 * W6) + (5 * (600n)) pd=(2 * W6) + (5 * (600n)) \
      m=1 region=sat
  M17 (V4 _net0 V1 Vdd) ami06P w=W1 l=600n as=W1 * 2.5 * (600n) ad=W1 * \
      2.5 * (600n) ps=(2 * W1) + (5 * (600n)) pd=(2 * W1) + (5 * (600n)) \
      m=1 region=sat
  M19 (V3 _net1 V1 Vdd) ami06P w=W1 l=600n as=W1 * 2.5 * (600n) ad=W1 * \
      2.5 * (600n) ps=(2 * W1) + (5 * (600n)) pd=(2 * W1) + (5 * (600n)) \
      m=1 region=sat
  M21 (Vo V3 Vss Vss) ami06N w=W5 l=600n as=W5 * 2.5 * (600n) ad=W5 * \
      2.5 * (600n) ps=(2 * W5) + (5 * (600n)) pd=(2 * W5) + (5 * (600n)) \
      m=1 region=sat
  M14 (V4 V4 Vss Vss) ami06N w=W2 l=600n as=W2 * 2.5 * (600n) ad=W2 * \
      2.5 * (600n) ps=(2 * W2) + (5 * (600n)) pd=(2 * W2) + (5 * (600n)) \
      m=1 region=sat
  M20 (V3 V4 Vss Vss) ami06N w=W2 l=600n as=W2 * 2.5 * (600n) ad=W2 * \
      2.5 * (600n) ps=(2 * W2) + (5 * (600n)) pd=(2 * W2) + (5 * (600n)) \
      m=1 region=sat
  R0 (net81 Vo) resistor r=R1
  C0 (V3 net81) capacitor c=C1

ends twostage_opamp_ami06
```

Figure 79. Op-amp Spectre netlist.

CHAPTER 10

CONCLUSIONS

IC reliability simulation technique is supposed to be applied not only as the designer's guide in design phase, also as the supplemental means to existing reliability tests in industry such as qualification test and burn-in test in product engineering to the real world ICs. This mission had not been fully accomplished in the previous work. Therefore, in this research work, the IC reliability simulation has been redefined and also developed in the following way.

First, the device level physical models should be able to evaluate the reliability of post-fab ICs, which inevitably have physical defects inside the circuits. The effort to achieve this goal is incorporating post-fab defects in EM physics-of-failure model based on the review of general EM models. This includes modifying structural factors at the defect area, revising thermal conditions, and etc.

Second, the simulation results have to clearly reflect the impact of device level degradations on circuit level spec degradations under certain failure mechanisms. This is achieved through the circuit level simulation algorithms such as hierarchical approach. Corresponding to the upgraded EM model at device level, a probability model has been developed to predict the expected interconnect lifetime under EM degradation, based on statistical defect distribution and probability theory.

To have more confidence in reliability simulation, the reliability simulator has to be calibrated to produce the correct and accurate enough results. In order to do so, the very

costly and time-consuming stress tests are conducted, and the calibrated simulator ARET has shown an excellent agreement with measurements.

The last revision to the reliability simulation made in this work is that, in addition to basic simulation functions, the reliability simulation technique should be able to contribute to IC design-for-reliability for reliability improvement. The distinct reliability hotspot identification feature has been developed to fulfill this requirement. The identification of the reliability hotspots offers the opportunity performing the local design-for-reliability technique.

Compared to the reliability simulation, the design-for-reliability technique is very immature. However, under the aggressive technology scaling and the resulting shrinkage of reliability safety margin, effective and practically feasible DFR techniques are certainly needed. In this work, a new DFR concept, local design-for-reliability, has been proposed, and various DFR algorithms are developed to achieve this goal.

The local DFR approach proposed is based on reliability simulation and hotspot identification, where the "localization" of design is the key technical point. With the redesign process localized around the reliability hotspot, the overall reliability can be effectively improved and the overall performance can be maintained as well. Thus, with the very limited local design work, both reliability enhancement and performance maintenance are obtained. This makes the local DFR a feasible and thus attractive DFR approach.

Different DFR algorithms are developed for different circuit/failure situations. For interconnect failures under EM, the interconnect dimension modulation approach is proposed. Based on simulations, this approach can effectively prolong the IC

interconnect lifetime with very simple interconnect redesign. For CMOS digital circuits, the major algorithms implemented in this work include dimension modulation and signal modulation. The experimental results have shown very promising reliability improvements (up to 50%) with acceptable design work involved for various circuits. For analog circuits, with much more complex situations than in digital circuits, a fundamental approach/algorithm has been set up, and demonstrated by incorporating circuit synthesis technique.

Regarding the future work, the effect of the post-fab physical defects on device degradations needs to be addressed. The variation of device dimensions such as uneven oxide thickness may have impact on the corresponding wear-out degradations. This will have to be accomplished based on statistical processes. More investigations on oxide breakdown and the relationship between existing failure mechanisms will be important as well. Currently the reliability simulator ARET is only a technical work focusing on reliability simulation. In order to use ARET efficiently by different users, some existing problems associated with the tool will have to be solved, such as the lengthy simulation time and software bugs. In terms of design-for-reliability, the feasible local design-for-reliability algorithm for analog circuits needs to be developed. Since with local DFR, the specified reliability goal is not always guaranteed to achieve, it will be wise and helpful to explore other DFR approaches. These approaches may be local or global ones, but they have to be feasible to apply practically.

APPENDIX A

ARET OPERATION GUIDE

A.1. System requirements

Cadence/Spectre 4.4.5 (or not too older than that) and Tcl/TK Wish 8.0 must be installed. The OS is SunOS 5.8 and for a fast analysis, UNIX Ultra 10 work station or higher should be used.

A.2. General operations

Before running ARET, the netlist of the circuit under simulation must be created and some basic information about the interconnect must be added into the existing database. Please refer to section 3 and 4 for detailed instructions.

To run ARET, simply run “aret” in the tool directory. The main interface will pop up with a title tag. Click to close the title window and get into the normal operation.

The output will be plots of performance degradations under failure mechanisms.

A.2.1. Open netlists

The first step is to open the netlist file that you want to analyze. Use the “File” function at the top of the interface to find the netlist file under “/netlists”. There are several sample netlist files currently in this directory, such as *e_m.ckt*, *opamp.ckt*.

A.2.2. Select failure mechanisms

In the function “Mechanisms”, select the failure mechanism(s) you want to deal with. For now, only electromigration and hot-carrier effects are available. You can choose one of them or both. However, for the component-level evaluation (detailed in the next section.) you can only choose the mechanism that is affecting the component you are

selecting. For example, if you are looking at metal interconnect traces, then you can only select "Electromigration" in the "Mechanisms" function.

A.2.3. Select evaluation level

Under function "Degradation", you can decide the evaluation is at component-level or circuit-level. The former means the analyses on component parameters, such as the resistance of interconnect trace. The latter means the analyses on circuit performance, which is selected in further options under this function. In order to run this simulation, the appropriate failure mechanisms must be selected.

A.2.3.1. Component-level evaluation

When you selected this function, an information sheet will pop up. You need to fill out all these information required to start a correct analysis. For this case, the component name, the temperature, and the time length are needed. Once the simulation is done, a plot of component parameter degradation vs. time will be returned.

A.2.3.2. Circuit-level evaluation

In this function, the circuit system-level performance degradation due to selected failure mechanisms will be simulated. In this case, you will have to specify the output characteristics, in which you want to see any degradation. Again, a plot of performance vs. degradation will be returned as soon as the simulation is done.

A.2.4. Sensitivity analysis

This function will analyze and identify the circuit reliability "hotspots", which are most likely to cause circuit fail. Again, In order to run this simulation, the appropriate failure mechanisms must be selected first. This function may take longer time to return.

Once it returns it will give a list of the "hottest" spots in this circuit, under specified failure mechanisms.

A.2.5. Re-design

This function was developed to design the circuit with improved reliability. The corresponding design-for-reliability algorithms have now been developed but not been integrated in the tool yet.

A.2.6. Help

The "Help" button at the upper-right corner does not help at this time. A help system may be added under this function, depending on the certain situation.

A.3. How to create netlist file

The ARET netlist file takes the basic form of the Cadence/Spectre netlist and adds some special format information to minimize the cost on running time. Thus, any circuit under simulation must be first described in Spectre format. The following information is then to be added to form an ARET netlist file. The netlist file must be in the directory "netlists".

Two example netlist files are attached to demonstrate this procedure. They are *opamp.ckt*, an analog op-amp circuit and *inverter.ckt*, a CMOS inverter circuit.

A.3.1. Device model parameters

Every transistor model definition must be right BEFORE the defined device description line, such as m1, m2, etc. that are the corresponding inputs while running the simulation. The device model name has to be "X" plus the device name. For example, a transistor m1 will have its model name *Xm*1. The threshold voltage *vto*, the transconductance *kp*, and any other hot-carrier-sensitive model parameters in the future

version must be listed in this line in this sequence. There must be at least ONE space between the equal mark and the following parameter value, which has to have enough digits for the length of scientific number in C/C++, usually TWELVE digits.

model Xm1 mos1 type=n vto= 7.50000e-01   kp= 1.00000e-04

A.3.2. Device parameters

Every device description line must be following its model definition line. There must be at least ONE space between the left parenthesis and the first node, and between the last node and the right parenthesis as well. Again, there has to be at least ONE space between the equal mark and the parameter value such as channel length. However, those values are NOT required to have any specific length.

m1 (7 1 3 5) Xm1 l= 0.5e-6 w= 25e-6

A.3.3. Interconnect parameters

All interconnect lines under investigation must be written in form of resistor. As same as the rules for device parameters, there has to be at lease ONE space between parentheses and nodes, equal marks and the resistance values. The interconnect resistance value here does not have to be exact. A rough estimation will be enough. The tool will calculate the accurate resistance later in simulation. But it must have at least EIGHT-digit and a TWELVE-digit space after it until an end notation specifying that this line is a interconnect line.

r1 (9 99) resistor r= 3.186e-1   // interconnection

A.3.4. Signal parameters

A sign line “//signal” must be right above the circuit input signal line. Besides that there has to be at least one space between parentheses and nodes, at the end of the line, the period of the signal must be given after the sign “//period=” and ONE space.

//signal

Vi (1 2) vsource type=sine freq=500000.0 ampl=1 //period= 2e-6

Similarly, the output signal nodes must be specified starting with “//signal_out”, in form

//signal_out

//(4 gnd)

NOTE: if you are simulating the “delay” in a digital circuit, you must make sure that, with the applied input signal the signal at the output will change as the input changes its logic level. In addition, the maximum delay over this path must also be assigned in the following form.

//assigned_delay

// 5E-7

A.3.5. Analysis specifications

For any circuit netlist, there must be two basic analyses included: DC analysis *dc* and transient analysis *tran* as following with the sign line “//DC” and “//Sweep” above. The names of the two analyses are “bias” and “sweep”. All analyses are shielded by “//” followed by a space, and the tool will select the ones needed in simulation. For the transient analysis, the same rules apply here: at least ONE space between equal marks and values, the values must be at least TWELVE digits, and there is an end sign at the end of the line.

//DC

// bias dc oppoint=screen

//Sweep

// sweep tran start= 0 stop= 1.00000e-02   strobeperiod= 1.00000e-02 //end

In case other circuit-level specs are wanted to simulate, the corresponding Spectre analysis must be added into the netlist file. For the current version, only the simulations on open-loop voltage gain and propagation delay are available. The following is the formats to describe the voltage gain analysis and the propagation delay.

Again, a sign line "//Gain" is required, there should be ONE space between parentheses and nodes, and the analysis is shielded by "//" and a space.

//Gain

// gain (9 gnd) xf stimuli=sources freq=500000.0

//Delay

// delay tran start= 0 stop= 1.00000e-03   // end

A.4. How to add interconnect information into data file

To run any simulation under electromigration, some necessary physical information about the interconnect line must be added into the data file named *em_analysis.in* in the directory "netlists".

In the data file, every interconnect occupies one line, in which every item is separated by a space. This line starts with netlist name and interconnect name, and must be ended by a "*".

The information following the interconnect name are in such a sequence:

a) Seed number of grain texture – seed number for grain growth,

b) Hazard level – proportion of hazard-free structure factors,

c) Mean grain size (m),

d) Line width (m),

e) Line length (m),

f) Line thickness (m),

g) Atomic volume (m^3),

h) Atomic concentration at grain boundary (m^{-3}),

i) Diffusion coefficient (m^2/s),

j) Effective charge of the ion

k) Resistivity (ohm-m),

l) Temperature coefficient of resistivity (ohm-m/K),

m) Structural factor,

n) Mean grain boundary activation energy (eV),

o) Threshold current density,

p) Thermal switch – to control if the thermal profile is created, 1:yes, 0:no.

Among these parameters, a, b, and m do NOT change with interconnect type. Parameter o is NOT in use for the current version. All the rest of the parameters will have to be modified according to the interconnect type, such as the material properties.

After parameter p, the next item can be set to either "+", meaning the interconnect line is defective, or "-", meaning it is defect-free. After this switch, the next parameter is the total number of the defects present on this line, followed by a series of defect sizes. For example, if there is one defect on the line and it is 40% of the line width, then it will be in the form "+ 1 0.4 *". If there are three defects with sizes of 30%, 40%, and 50% of

the width, respectively, then it will be “+ 3 0.3 0.4 0.5 *”. NOTE: This feature is for the specific study on defective interconnect, in other words, it is for component-level analyses only. In circuit-level simulation, the physical defects are generated statistically by the probability model in ARET.

- August 20, 2002

APPENDIX B

ISCAS85 BENCHMARK CIRCUITS

The functions and the schematics of the six ISCAS85 benchmark circuits used in this work are listed below [49].

C432

- Function: 27-channel interrupt controller
- Statistics: 36 inputs; 7 outputs; 160 gates
- Schematic:

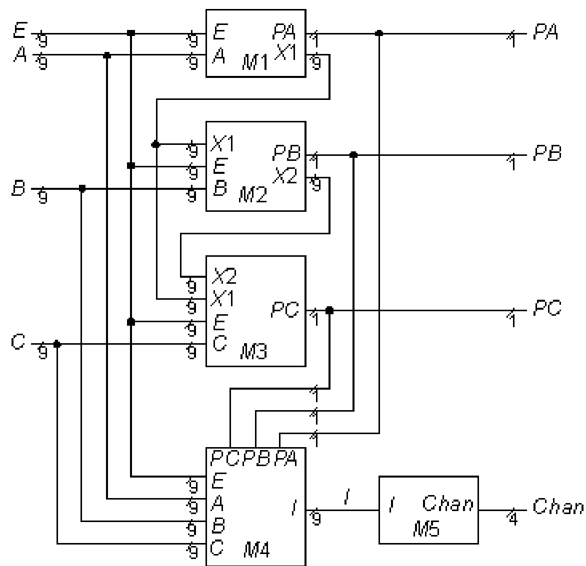


Figure 80. ISCAS benchmark circuit C432.

C880

- Function: 8-bit ALU
- Statistics: 60 inputs; 26 outputs; 383 gates

- Schematic:

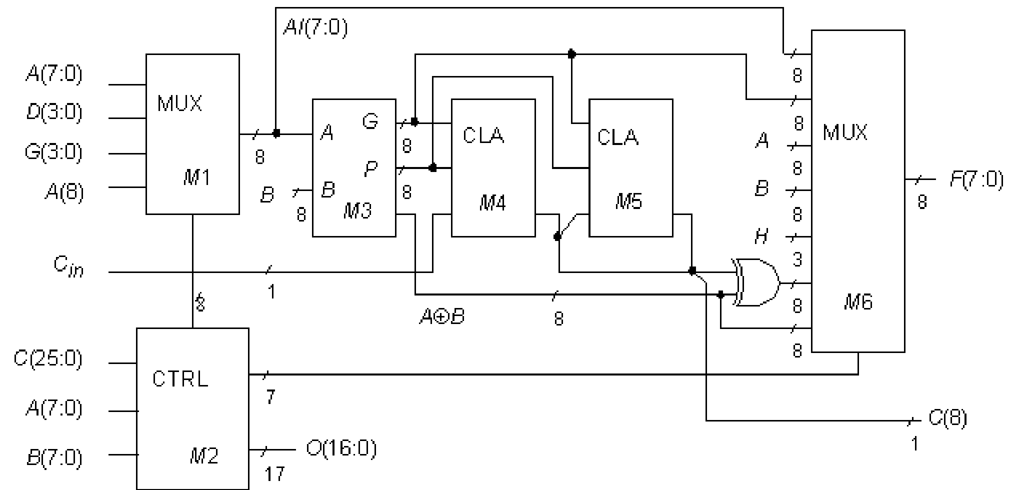


Figure 81. ISCAS benchmark circuit C880.

C1908

- Function: 16-bit error detector/corrector
- Statistics: 33 inputs; 25 outputs; 880 gates
- Schematic:

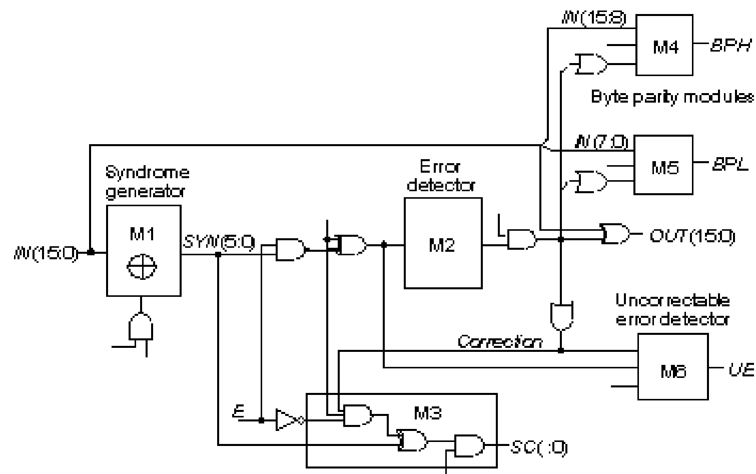


Figure 82. ISCAS benchmark circuit C1908.

C2670

- Function: 12-bit ALU and controller
- Statistics: 233 inputs; 140 outputs; 1193 gates
- Schematic:

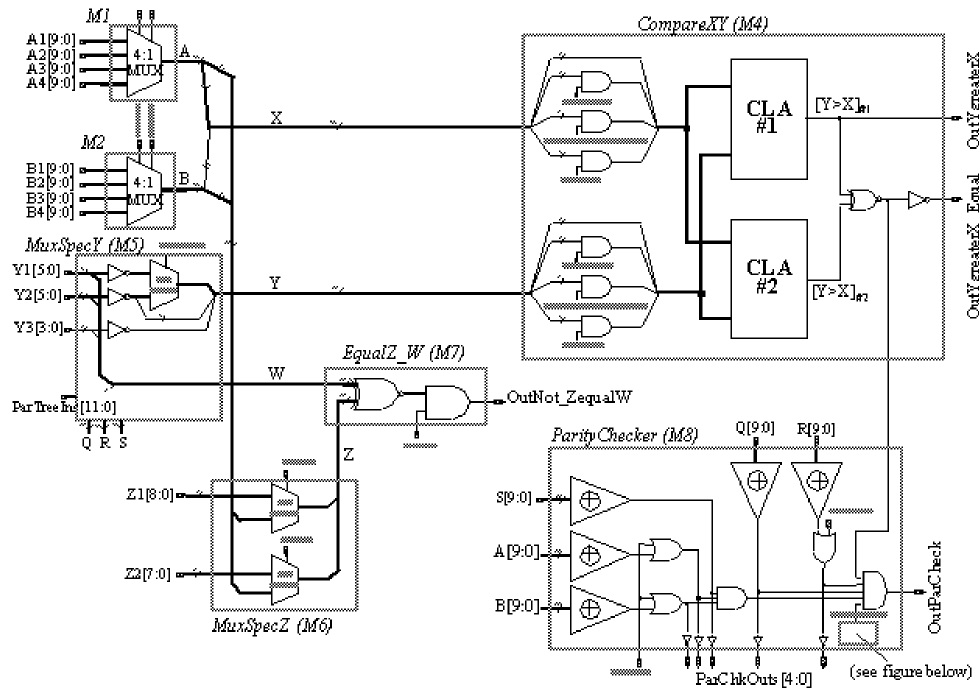


Figure 83. ISCAS benchmark circuit C2670.

C3540

- Function: 8-bit ALU with binary and BCD arithmetic, with logic/shift operations
- Statistics: 50 inputs; 22 outputs; 1669 gates
- Schematic:

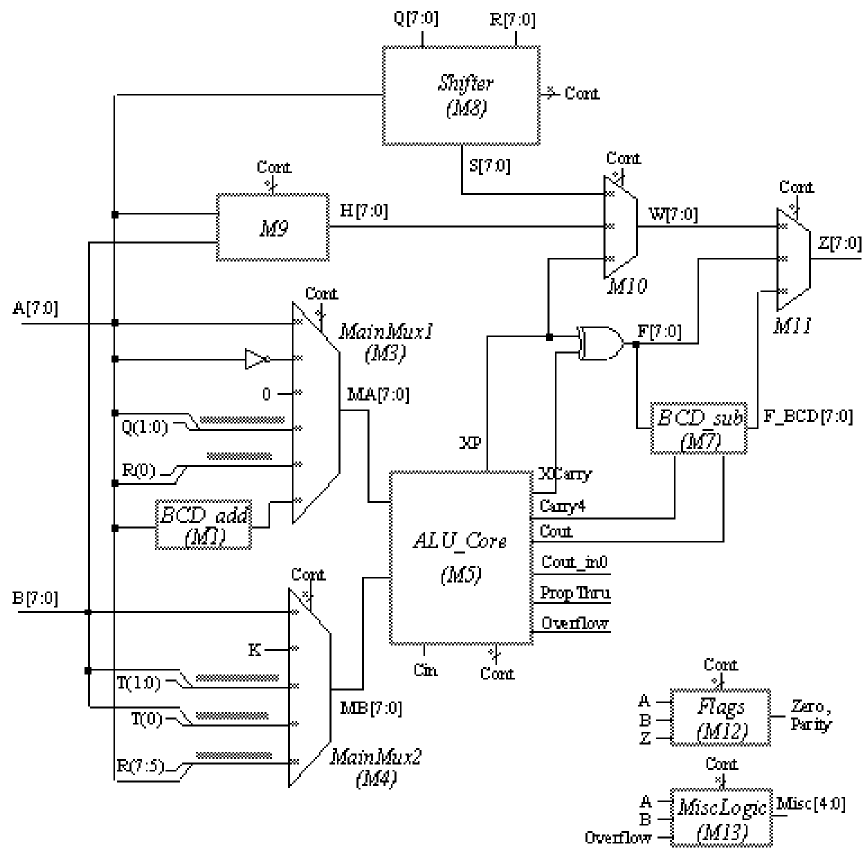


Figure 84. ISCAS benchmark circuit C3540.

C6288

- Function: 16×16 multiplier
- Statistics: 32 inputs; 32 outputs; 2406 gates
- Schematic:

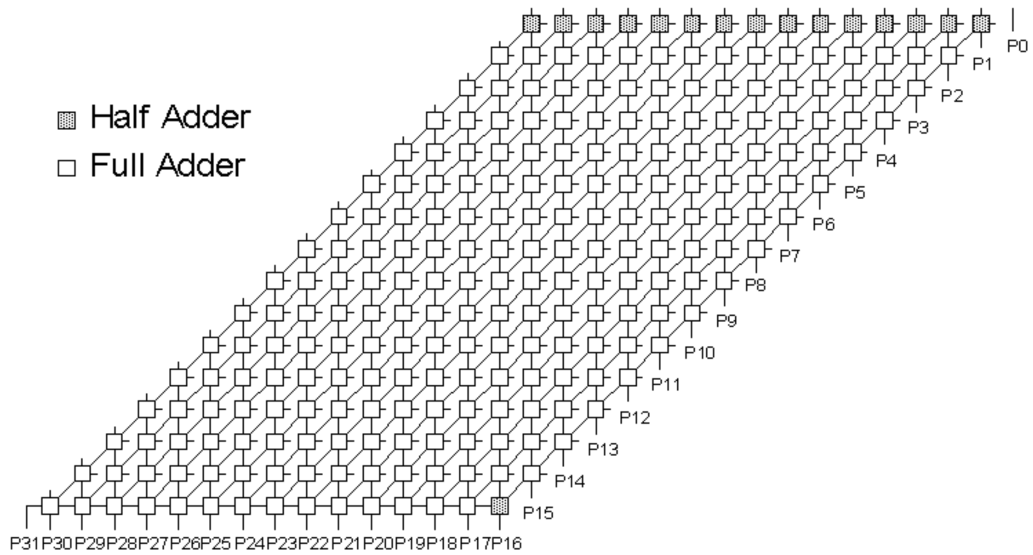


Figure 85. ISCAS benchmark circuit C6288.

REFERENCES

[1] A. Pages and M. Gondran, translated by E. Griffin, *System Reliability – Evaluation and Prediction in Engineering*, New York: Springer-Verlag, 1986.

[2] C. F. Hawkins, J. Segura, J. M. Soden and T. A. Dellin, "Test and Reliability: Partners in IC Manufacturing," *IEEE DESIGN & TEST OF COMPUTERS*, Oct.,1999, pp.66-73.

[3] M. Bohr, "High Performance Logic Technology and Reliability Challenges," *IEEE Int. Reliab. Pyhs. Symp.*, Key Note Presentation, 2003.

[4] A. Suyko and S. Sy, "Development of a Burn-In Time Reduction Algorithm Using the Principles of Acceleration Factors," *International Reliability Physics Symposium*, pp. 264-270, 1991.

[5] B. J. Sheu, W-J. Hsu, B. W. Lee, "An Integrated-Circuit Reliability Simulator – RELY," *IEEE J. Solid-State Circuits*, vol. SC-24, no. 2, pp. 473-477, 1989.

[6] C. Hu, "IC Reliability Simulation," *IEEE Journal of Solid-State Circuits,* Vol. 27, No. 3, pp. 241-246, March 1992.

[7] X. Xuan, A. Chatterjee, and A. D. Singh, "ARET for System-Level IC Reliability Simulation," *International Reliability Physics Symposium*, Vol. 41, pp. 572-574, March 2003.

[8] S. Aur, D. E. Hocevar, and P. Yang, "HOTRON – A Circuit Hot Electron Effect Simulator," *IEEE International Conference of Computer-Aided Design*, pp. 256-259, 1987.

[9] F. N. Najm, R. Burch, P. Yang, and I. N. Hajj, "Probabilistic Simulation for Reliability Analysis of CMOS VLSI Circuits," *IEEE Transactions on Computer-Aided Design*, vol. CAD-9, no. 4, pp. 439-450, 1990.

[10] X. Xuan, A. D. Singh, and A. Chatterjee, "Reliability Evaluation for Integrated Circuit with Defective Interconnect under Electromigration," *International Symposium on Quality Electronic Design,* pp. 29-34, San Jose, March 2003.

[11] J. P. Shen, W. Maly and F. J. Ferguson, "Inductive Fault Analysis of MOS Integrated Circuits," *IEEE Design and Test*, December 1985, pp. 13-26.

[12] Y. Leblebici and S. M. Kang, *Hot-Carrier Reliability of MOS VLSI Circuit,* Boston: Kluwer Academic Publishers, 1993.

[13] J. E. Chung, M. C. Jeng, J. E. Moon, P. K. Ko, and C. Hu, "Performance and Reliability Design Issues for Deep-Submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, pp. 545-554, 1991.

[14] Y. Leblebici, "Design Considerations for CMOS Digital Circuits with Improved Hot-Carrier Reliability," *IEEE Journal of Solid-State Circuits*, Vol. 31, No. 7, July 1996.

[15] M. A. Styblinski and M. Huang, "Drift Reliability Optimization in IC Design: Generalized Formulation and Practical Examples," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 12, No. 8, pp. 1242-1252, August 1993.

[16] L. J. Opalski and M. A. Styblinski, "Generation of Yield Optimization Problem: Maximum Income Approach," *IEEE Transactions on Computer-Aided Design*, vol. CAD-5, no. 2, 1986.

[17] A. Christou, *Electromigration and Electronic Device Degradation*, New York: John Wiley & Sons, Inc., 1994.

[18] P. J. Marcoux, P. P. Merchant, V. Naroditsky, and W. D. Rehder, "A new 2D simulation model of electromigration," *Hewlett-Packard J.* June, pp. 79 (1989).

[19] K. Nikawa, "Monte Carlo calculations based on the generalized electromigration failure model," *IEEE Int. Reliab. Pyhs. Symp.*, vol. CH1619-6, 175 (1981).

[20] J. Harrison, Jr., "A simulation model for electromigration in fine-line metallization of integrated circuits due to repetitive pulsed currents," *IEEE Trans. Electron Devices,* vol. 35, no. 12, 2170 (1988).

[21] I. A. Blech, "Electromigration in thin aluminum films on titanium nitride," *J. Appl. Phys.,* vol. 47, no. 4, 1203 (1976).

[22] P. F. Tang, A. G. Milnes, C. L. Bauer, and S. Mahajan, "Electromigration in thin films of Au on GaAs," *Proc. Mater. Res. Soc.*, vol. 167, 341 (1989).

[23] Y. Leblebici and S. M. Kang, *Hot-Carrier Reliability of MOS VLSI Circuit*, Boston: Kluwer Academic Publishers, 1993.

[24] P. E. Cottrell, R. Troutman and T. Ning, "Hot-electron emission in n-channel IGFET's," *IEEE Electron Device Lett.,* vol. ED-26, pp. 520-532, April 1979.

[25] K. - L. Chen, S. A. Saller, I. A. Groves and D. B. Scott, "Reliability effects on MOS transistors due to hot-carrier injection," *IEEE Trans. Electron Devices,* vol. ED-32, pp. 386-393, February 1985.

[26] C. T. Sah, “VLSI device reliability modeling,” in *Proceedings of 1987 Int. Symp. On VLSI Technology, System and Applications*, pp. 153-162, May 13-15, 1987.

[27] P. Heremans, G. van den Bosch, R. Bellens, G. Groeseneken and H. Maes, “Temperature dependence of the channel hot-carrier degradation of n-channel MOSFETs,” *IEEE Trans. Electron Devices*, vol. 37, pp. 980-993, April 1990.

[28] P. Heremans, R. Bellens, G. Groeseneken and H. E. Maes, “Consistent model for the hot-carrier degradation in n-channel and p-channel MOSFETs,” *IEEE Trans. Elecron Devices*, vol. ED-35, pp. 2194-2209, December 1988.

[29] M. Ancona, N. Saks, and D. McCarthy, “Lateral distribution of hot-carrier-induced interface traps in MOSFET’s,” *IEEE Trans. Electron Devices*, vol. ED-35, pp. 2221-2228, December 1988.

[30] M. Shatzes, “On the Nature of Conduction and Switching in SiO_2,” *J. Appl. Phys.*, vol. 45, pp. 2065-2077, 1974.

[31] C. Osburn and D. Ormond, “Sodium-Induced Barrier Height Lowering and Dielectric Breakdown on SiO_2 Films on Silicon,” *J. Electrochem. Soc.*, vol. 121, pp. 1195-1198, 1974.

[32] J. S. Suehle, “Ultrathin Gate Oxide Reliability: Physical Models, Statistics, and Characterization,” *IEEE Transactions on Electron Devices*, vol. 49, no. 6, 2002.

[33] J. W. McPherson and D. A. Baglee, “Acceleration factors for thin gate oxide stressing,” *in Proc. IRPS,* vol. 23, pp. l-5, 1985.

[34] I. C. Chen, S. Holland, and C. Hu, “A Quantitative Physical Model for Time-Dependent Breakdown in SiO2,” *Int. Reliability Physics Symposium*, vol. 23, pp. 24-27, 1985.

[35] I. C. Chen, S. E. Holland, K. K. Young, C. Chang, and C. Hu, “Substrate Hole Current and Oxide Breakdown,” *Appl. Physics. Lett.*, vol. 49, pp. 669-671, 1986.

[36] M. Depas, T. Nigam, and M. H. Heynes, “Soft Breakdown of Ultra-Thin Gate Oxide Layers,” *IEEE Trans. Electron Devices*, vol. 43, pp. 1499-1504, 1996.

[37] K. R. Farmer, R. Saletti, and R. A. Burhman, “Current Fluctuations and Silicon Wear-out in Metal-Oxide Semiconductor Tunnel Doides,” *Appl. Phys. Lett.*, vol. 52, pp. 1749-1751, 1988.

[38] S. H. Lee, B. J. Cho, J. C. Jim, and S. H. Choi, “Quasibreakdown of Ultrathin Gate Oxide under High Field Stress,” *IEDM Tech. Dig.*, pp. 605-606, 1994.

[39] J. R. Lloyd, P. M. Smith, and G. S. Prokop, "The Role of Metal and Passivation Defects in Electromigration-Induced Damage in Thin Film Conductors," *Thin Solid Films*, pp. 385-395, 1982.

[40] C. H. Stapper, "Modeling of Integrated Circuit Defect Sensitivities," *IBM J. Research and Development*, Vol. 27, No. 6, pp. 549-557, 1983.

[41] Z. Stamenkovic and N. Stojadinovic, "New Defect Size Distribution Function for Estimation of Chip Critical Area in Integrated Circuit Yield Models," *Electronics Letters*, Vol. 28, No. 6, March 1992, pp. 528-530.

[42] X. Xuan, A. Chatterjee, and A. D. Singh, "Application of Local Design-for-Reliability Techniques for Reducing Wear-out Degradation of CMOS Combinational Logic Circuits," *European Test Symposium*, May 2004.

[43] National Instruments, *Function and VI Reference Manual*, January 1998.

[44] X. Xuan, A. Chatterjee, A. D. Singh, N. P. Kim, and M. T. Chisa, "IC Reliability Simulator ARET and Its Application in Design-For-Reliability," *Asian Test Symposium*, November 2003.

[45] H. Hwang, J. Lee, P. Fazan, and C. Dennison, "Hot-Carrier Reliability Characteristics of Narrow-Width MOSFETs," *Solid-State Electronics*, Vol. 36, No. 4, pp. 665-666, 1993.

[46] J. P. Uyemura, *CMOS Logic Circuit Design*, Kluwer Academic Publishers, Boston, 1999.

[47] D. Han and A. Chatterjee, "Simulation-in-loop Analog Circuits Sizing Method Using Adaptive Model-Based Simulated Annealing," Technical report, Georgia Institute of Technology, 2004.

[48] M. Burns and G. W. Roberts, *An Introduction to Mixed-Signal IC Test and Measurement*, Oxford University Press, 2001.

[49] M. Hansen, H. Yalcin, and J. Hayes, Web resources at www.eecs.umich.edu, University of Michigan.

[50] D. G. Pierce and P. G. Brusius, "Electromigration: A Review," *Microelectron Reliability*, Vol.37, No.7, pp. 1053-1072, 1997.

[51] N. McCusker, H.Gamble and B.Armstrong, "Surface Electromigration in Copper Interconnects," *International Reliability Physics Symp.*, 1999, pp.270-275.

[52] J. Lioyd, "Electromigration in Thin Film Conductors," *Semiconductor Science Technology*, vol.12, 1997, pp.1177-1185.

[53] Eiji Takeda, Cary Y.Yang and Akemi Miura-Hamada, *Hot-Carrier Effects in MOS Devices,* 1995.

[54] Mario G. Ancona, Nelson S. Saks and Daniel McCarthy, "Lateral Distribution of Hot-Carrier-Induced Traps in MOSFET's," *IEEE Transactions on Electron Devices,* Dec. 1988, Vol.35, pp.2221-2228.

[55] A. Schwerin, W. Hansch and W. Weber, "The Relationship Between Oxide Charge and Device Degradation: A Comparative Study of n- and p- Channel MOSFET's," *IEEE Transactions on Electron Devices,* Dec.1987, Vol.34, pp.2493-2500.

[56] Y.Leblebici, S.M.Kang, C.T.Sah and T.Nishida, "Modeling and Simulation of Hot Electron Effects for VLSI Reliability," *International Test Conf.*, 1987, pp.252-225.

[57] Mary M. Kuo, Koichi Seki, Peter M. Lee, Jeong Yeol Choi, Ping K.Ko. and Chenming Hu, "Simulation of MOSFET Lifetime under AC Hot-Electron Stress," *IEEE Transactions on Electron Devices,* July 1988, Vol.35, pp.1004-1010.

[58] Mohamed El-Banna and Mahmoud A. El-Nokali, "A Simple Analytical Model for Hot-Carrier MOSFET's," *IEEE Transactions on Electron Devices,* May 1989, Vol.36, pp.979-986.

[59] Patrick Roblin, Amer Samman and Steven Bibyk, "Simulation of Hot-Electron Trapping and Aging of nMOSFET's," *IEEE Transactions on Electron Devices,* Dec.1988, Vol.35, pp.2229-2237.

[60] Y. Leblebici and S.M.Kang, "Modeling of nMOS Transistors for Simulation of Hot-Carrier-Induced Device and Circuit Degradation," *IEEE Transactions on Computer-Aided Design,* Feb.1992, Vol.11, pp.235-246.

[61] Tadanori Yamaguchi and Seiichi Morimoto, "Analytical Model and Characterization of Small Geometry MOSFET's," *IEEE Transactions on Electron Devices*, 1983, Vol.30, pp.559-561.

[62] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high performance and low power-the next ten years," *Proc. IEEE,* vol. 83, pp. 595-606,1995.

[63] M. Redder, S. Aur, and I. C. Chen, "A scaled 1.8 V, 0.18 pm gate length CMOS technology: device design and reliability considerations," *in IEDM Tech. Dig.,* pp. 415418, 1995.

[64] R.-H. Yan, D. Monroe, J. Weis, A. Mujtaba, and E. Westerwick, "Reducing operating voltage from 3,2, to 1 Volt and below - challenges and guidelines for possible solutions," in *IEDM Tech. Dig.,* pp. 55-58,1995.

[65] B. Davari, "CMOS technology scaling, 0.1 pm and beyond," *in IEDM tech. Dig.,* pp. 555-558, 1996.

[66] T. Mizuno, A. Toriumi, M. Iwase, M. Takahashi, N. Niiyama, M. Fukumoto, and M. Yoshimi, "Hot-carrier effects in 0.1 pm gate length CMOS devices," *in IEDM Tech. Dig.,* pp. 695-698, 1992.

[67] D. Esseni, L. Selmi, R. Bez, E. Sangiorgi, and B. Ricc, "Bias and temperature dependence of gate and substrate currents in nMOSFETs at low drain voltage," *in IEDM Tech. Dig.,* pp. 307-310, 1994.

[68] P. K. Chatterjee, W. R. Hunter, A. Amerasekera, S. Aur, C. Duvvury, P. E. Nicollian, L. M. Ting, and P. Yang, "Trends for deep submicron VLSI and their implications for reliability," *in IEEE International Reliability Physics Proceedings, pp.* l-11, 1995.

[69] D. L. Crook, "Evolution of VLSI reliability engineering," *Proc. IRPS*, pp. 2-11, (1990).

[70] H. H. Huston, C. P. Clarke, "Reliability defect detection and screening during processing-theory and implementation," *Proc. IRPS,* pp. 268-275, (1992).

[71] J. A. van der Pal, G. N. van der Voort, J. de Groot, "Yield screening, a novel reliability improvement technique," *JESSI T4 Project Report RiYR 48/91-XII-074,* (1991).

[72] J. Y. Choi, P. K. Ko, C. Hu, "Hot Carrier Induced MOSFET Degradation Under AC Stress," *IEEE Electron Device Letters,* pp. 333-335 (Aug. 1987).

[73] W. Weber, R Thewes, "Hot Carrier Related Device Reliability for Digital and Analogue CMOS Circuits," *Semicond. Sci. Technol.,* pp. 1432-1443 (1995).

[74] K. Quader, P. Ko, C. Hu, "Projecting CMOS Circuit Hot-Carrier Reliability from DC Device Lifetime," *International Electron Device Meeting,* pp. S 11-514 (1993).

[75] K. Quader, P. Fang, J. Yue, P. Ko, C. Hu, "Hot-Carrier Design Rules for Translating Device Degradation to CMOS Digital Circuit Degradation," *IEEE Trans. Electron Devices,* pp. 681-691(May 1994).

[76] B. K. Liew, N. W. Cheung, C. Hu, "Projecting Interconnect EM Lifetime for Arbitrary Current Waveforms," *IEEE Trans. Electron Devices,* p. 1343 (1990).

[77] L. M. Ting, J. S. May, W. Hunter, J. McPherson, "AC Electromigration of Multilayered Interconnects," *311th Int. Reliability Phys. Symp.,* p. 311 (1993).

[78] M. Attardo, R. Rutledge, and R. C. Jack, "Statistical metallurgical model for electromigration failure in aluminum thin-film conductors," *J. Appl. Phys.*, vol. 42, no. 11, 4343 (1971).

[79] J. Tao, J. F. Chen, N. W. Cheung, C. Hu, "Modeling and Characterizing Electromigration Failures Under Bi-directional Current Stress," *IEEE Trans. Electron Devices,* pp. 800-808 (May 1996).

[80] E. Rosenbaum, Z. Liu, C. Hu, "The Effect of Stress Waveform on MOSFET Performance," *IEDM,* pp. 719-722 (1991).

[81] E. Rosenbaum, C. Hu, "Silicon Dioxide Breakdown Lifetime Enhancement Under Bipolar Bias Conditions," *IEEE Trans. Electron Devices,* pp. 2287-2295 (Dec. 1993).

[82] E. Takeda, K. Ikuzaki, H. Katie, Y. Ohji, K. Hinode, A. Hamada, T. Sakuta, T. Funabiki, and T. Sasaki, "VLSI reliability challenges: from device physics to wafer scale systems," *Proc. IEEE,* vol. 83, pp. 653-673,1993.

[83] M. K. Orlowski and C. Werner, "Model for the electric fields in LDD MOSFETs-part II: field distribution on the drain side," *IEEE Trans. Electron Devices,* vol. 36, pp. 382-391, 1989.

[84] G. J. Dunn and J. T. Krick, "Channel hot-carrier stressing of reoxidized nitrided oxide p-MOSFETs," *IEEE Trans. Electron Devices,* vol. ED-38, pp. 901-906, 1991.

[85] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2," *J. Applied Physics,* vol. 40, pp. 278-283,1969.

[86] P. Chaparala, J. S. Suehle, C. Messick, and M. Roush, "Electric-field dependent dielectric breakdown of intrinsic SiO_2 films under dynamic stress," *in Proc. IRPS,* pp. 61-66, 1996.

[87] C. Hu *et al*., "Hot-Electron-Induced MOSFET Degradation – Model, Monitor, and Improvement," *IEEE Transactions on Electron Devices*, Vol. ED-32, pp. 375-384, February 1985.

[88] Y. Leblebici and S. M. Kang, "Modeling of nMOS Transistors for Simulation of Hot-carrier-Induced Device and Circuit Degradation," *IEEE Transactions on Computer-Aided Design*, Vol. 11, No. 2, pp. 235-246, February 1992.

[89] G. Groeseneken, R. Degraeve, T. Nigam, G. Van den Bosch, and H. E. Maes, "Hot-Carrier Degradation and Time-Dependent Dielectric Breakdown in Oxides," *Microelectronic Engineering*, Vol. 49, pp. 27-40, 1999.

[90] A. Schwerin, W. Haensch, and W. Weber, "The Relationship between Oxide Charge and Device Degradation," *IEEE Trans. Electron Devices*, Vol. 34, pp. 2493-2499, December 1987.

[91] X. Xuan and A. Chatterjee, "Sensitivity and Reliability Evaluation for Mixed-Signal ICs under Electromigration and Hot-Carrier Effects," *IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, pp. 323-329, October 2001.

[92] B. Marchand, D. Blachier, C. Lerous, G. Ghibaudo, F. Balestra, and G. Reimbold, "Generation of Hot Carriers by Secondary Impact Ionization in Deep Submicron Devices: Model and Light Emission Characterization," *IEEE International Reliability Physics Symposium*, pp. 93-97, 2000.

[93] S. Naseh, M. J. Deen, and O. Marinov, "Effects of Hot-Carrier Stress on the RF Performance of 0.18 μm Technology NMOSFETs and Circuits," *IEEE International Reliability Physics Symposium*, pp. 98-104, 2002.

[94] T. Sakurai and A. R. Newton, "Delay Analysis of Series-Connected MOSFET Circuits," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 2, 1991.

[95] M. Hafed, M. Oulmane, and N. C. Rumin, "Delay and Current Estimation in a CMOS Inverter with an RC Load," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 1, 2001.

[96] X. Xuan, A. Chatterjee, and A. D. Singh, "Local Redesign for Reliability of CMOS Digital Circuits under Device Degradation," *International Reliability Physics Symposium*, April 2004.

VITA

Xiangdong Xuan, son of Yimin Zhao and Zian Xuan, was born on February 4th, 1971, in Shenyang, Liaoning, P. R. China. He received his pre-university education from Liaoning Shiyan elementary school and high school. He started his college career at Northern JiaoTong University, Beijing, from 1988 and graduated with Bachelor of Science degree in Electrical Engineering in July 1992. Subsequently, he joined Shenyang Institute of Railway Technology, Shenyang, China, and worked as an electrical engineer in Electrical and Electronic Design group until 1997. In September 1997, he was admitted to Auburn University, Auburn, Alabama, U. S. A., as a graduate student in Materials Engineering. After he completed his study at Auburn University with a Master of Science degree, he entered Georgia Institute of Technology, Atlanta, Georgia, U. S. A., pursuing his Ph. D. degree in School of Electrical and Computer Engineering in September 1999.