Design Tool and Methodologies for Interconnect Reliability Analysis in Integrated Circuits

by
Syed Mohiul Alam
S.M., Massachusetts Institute of Technology, 2001
B.S., The University of Texas at Austin, 1999

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2004

© 2004 Massachusetts Institute of Technology
All Rights Reserved

Author . . . .
Department of Electrical Engineering and Computer Science
August 31, 2004

Certified by . . . .
Donald E. Troxel
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Certified by . . . .
Carl V. Thompson
Stavros Salapatas Professor of Materials Science and Engineering
Thesis Supervisor

Read by . . . .
Duane S. Boning
Associate Department Head and Professor of Electrical Engineering
and Computer Science
Thesis Reader

Accepted by . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Design Tool and Methodologies for Interconnect Reliability Analysis in Integrated Circuits

by

Syed Mohiul Alam

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Total on-chip interconnect length has been increasing exponentially with technology scaling. Consequently, interconnect-driven design is an emerging trend in state-of-the-art integrated circuits. Cu-based interconnect technology is expected to meet some of the challenges of technology scaling. However, Cu interconnects still pose a reliability concern due to electromigration-induced failure over time.

The major contribution of this thesis is a new reliability CAD tool, SysRel, for thermal-aware reliability analysis with either Al or Cu metallization technology in conventional and three-dimensional integrated circuits. An interconnect tree is the fundamental reliability unit for circuit-level reliability assessments for metallization schemes with fully-blocking boundaries at the vias. When vias do not block electromigration as indicated in some Cu experimental studies, multiple trees linked by a non-blocking via are merged to create a single fundamental reliability unit. SysRel utilizes a tree-based hierarchical analysis that sufficiently captures the differences between electromigration behavior in Al and Cu metallizations. The hierarchical flow first identifies electromigration-critical nets or "mortal" trees, applies a default model to estimate the lifetimes of individual trees, and then produces a set of full-chip reliability metrics based on stochastic analysis using the desired lifetime of the circuit.

We have exercised SysRel to compare layout-specific reliability with Cu and Al metallizations in various circuits and circuit elements. Significantly improved test-level reliability in Cu is required to achieve equivalent circuit-level reliability. The required improvement will increase as low-k dielectric materials are introduced and liner thicknesses are reduced in future.

Thesis Supervisor: Donald E. Troxel
Title: Professor of Electrical Engineering and Computer Science

Thesis Supervisor: Carl V. Thompson
Title: Stavros Salapatas Professor of Materials Science and Engineering

Dedication

To my mother

Shahana Alam

Acknowledgments

I am grateful to my co-advisors Professors Donald E. Troxel and Carl V. Thompson for their guidance and support throughout the years at MIT. There were numerous discussions and meetings leading to the successful completion of this thesis. I would like to thank Professor Duane S. Boning for serving in my thesis committee. He provided valuable feedback during one-on-one and committee meetings.

This research was supported by Microelectronics Advanced Research Corporation (MARCO) Interconnect Focus Research Center Program. I got numerous opportunities to actively participate in annual and quarterly reviews and to interact with other researchers and industry members in the program. I acknowledge the valuable interactions with Professor James Meindl, Professor Kenneth Goodson, Professor Anantha Chandrakasan, Professor Rafael Reif, Sungjun Im, Jae-Mo Koo, Eric Pop, and Tyrone Hill.

I would like to thank my colleagues at MIT, Frank Wei, Zung-Sun Choi, Ramkumar Krishnan, Rajappa Tadepalli, and Andrew Takahashi, for their support through discussions and collaboration. This research is partially built upon prior work in our group. The Magic layout parsing and data storage in SysRel are based on Yonald Cherry's earlier work with ERNI. Electromigration analysis in Cu metallization is based on experimental and modelling work by Gan Chee Lip, Frank Wei, and Zung-Sun Choi. Stefan Hau-Riege developed the concept of hierarchical reliability analysis with interconnect trees in Al metallization technology.

I was very fortunate for several summer internship opportunities during my graduate studies. The thermal analysis work using ANSYS originated during my internship at IBM T. J. Watson Research Center in Summer 2003. I am grateful to Kathryn W. Guarini and Meikei Ieong for giving me the opportunity to design the pilot testchip in 3D IC technology at IBM Watson Research.

My graduate school experience has been truly rewarding, both from academic and professional perspective. MIT fosters professional development and growth of its graduate student community through numerous programs and activities. I was the

Academics, Research, and Careers Committee Chair of the Graduate Student Council during the 2002-2003 academic year. I managed and led a group of graduate students in organizing major programs such as MIT Career Fair, Airport Shuttle Program, International Student Mentorship Program, Travel Grant Program, Academic Careers Series, and Professional Development Lecture Series. I would like to thank every one at the Graduate Student Council, and in particular, Emmi Snyder, for the wonderful experience.

Most importantly, I am indebted to the support and encouragement I have received over the years from my parents and family. I am grateful to my wife, Raiyan T. Zaman, for her love and dedication. I would like to thank my brother, Syed Masudul Alam, for his assistance in drawing many figures and for his support while writing this thesis.

Contents

1 Introduction **23**

1.1 Interconnect Reliability 24

1.1.1 Electromigration Phenomenon 25

1.1.2 Electromigration Failure Modes 27

1.1.3 Immortality Condition 28

1.1.4 Electromigration Lifetime Model 30

1.2 Interconnect Architecture Schemes 31

1.2.1 Aluminum Metallization Technology 31

1.2.2 Copper Metallization Technology 31

1.3 Fundamental Reliability Unit 32

1.4 Three-Dimensional (3D) Integrated Circuit Technology 34

1.4.1 Technology Concept 34

1.4.2 Layout Methodology for 3D Circuits 35

1.4.3 3D IC Technology with Thermal Management 36

1.4.4 Interconnect Reliability in 3D ICs 38

1.5 Circuit-Level Reliability Analysis 39

1.5.1 Electromigration Reliability in Networked Interconnects (ERNI and ERNI-3D) 40

1.6 Thesis Statement 41

1.7 Organization of Thesis 42

2 Hierarchical Reliability Analysis Flow **45**

2.1 Electromigration Experiments 45

2.2 Via Asymmetric Failure Characteristics in Copper Technology 47
2.3 Steps in Hierarchical Reliability Analysis for Copper Metallization . . 49
2.4 Contrast with Hierarchical Reliability Analysis in Aluminum Metallization 52
2.5 Electromigration Default Model 53
2.6 Summary 55

3 Electromigration Reliability Mathematics 57
3.1 Basic Definitions 57
3.2 Lifetime Distribution Model 61
3.2.1 Lognormal Distribution 62
3.3 Reliability Mathematics in Series and Parallel Systems 63
3.4 Reliability Metrics for Full-chip Analysis 65
3.4.1 Probability of no failure 66
3.4.2 Failure Rate along chip lifetime 66
3.4.3 Time to cumulative % failure 67
3.5 Reliability Metrics in Use 67
3.6 Summary 69

4 Electromigration Reliability Comparison of Aluminum and Copper Interconnects 71
4.1 Immortality Condition Filters 72
4.2 Diffusivity Mechanisms and Models 73
4.2.1 Diffusion in Aluminum Interconnects 73
4.2.2 Diffusion in Copper Interconnects 75
4.2.3 Diffusivity Comparison 76
4.3 Lifetime Comparison of Straight Line Interconnects 78
4.4 Reliability Prediction Using the ITRS 81
4.5 Summary 82

5 SysRel: Circuit-Level Reliability CAD Tool **85**

5.1 Software Architecture 85

5.2 Graphical User Interface 86

5.3 Circuit Layout Parsing 88

5.4 Interconnect Tree Extraction 90

5.5 Reliability Analysis Flow 91

5.6 Simulation with 32-bit Comparator Layout 93

5.6.1 Copper Metallization Technology 95

5.6.2 Aluminum Metallization Technology 96

5.7 Implications of SysRel 96

5.8 Three-Dimensional IC Analysis in SysRel 97

5.9 Summary 100

6 Cell-Based Reliability Analysis in SysRel **101**

6.1 Cell Characterization in SysRel 102

6.2 Hierarchical Layout Support in SysRel 103

6.3 Cell-based Reliability Simulation 104

6.4 Computational Efficiency 106

6.5 Summary 107

7 Non-blocking Via Analysis in Copper Metallization Technology **109**

7.1 Non-blocking Via Due to Liner Ruptures 110

7.2 Non-blocking Via Assignment in Circuit Layout 111

7.3 Steps for Non-blocking Via Analysis in SysRel 112

7.4 Interconnect Tree Merging 114

7.5 Line-length Effect on Reliability 116

7.6 32-bit Comparator Layout Simulation 118

7.7 Discussion of Results 119

7.8 Summary 120

8 Thermal-Aware Reliability Analysis in SysRel 123

8.1 Fundamental Concepts for Thermal Analysis 124

8.2 Thermal Simulation with ANSYS 127

8.3 Transient Thermal Behavior in Integrated Circuits 131

8.4 Cell-based Thermal Profiling Technique 134

8.5 Interconnect Joule Heating 138

8.6 Thermo-file Implementation in SysRel 140

8.7 Thermal-Aware Reliability Simulation with 32-bit Comparator 141

8.8 Summary 142

9 Reliability Analysis with Arithmetic and Logic Unit 145

9.1 Arithmetic and Logic Unit Design and Synthesis 145

9.2 Reliability Simulation with Different Metallization Technologies . . . 148

9.3 Thermal-Aware Reliability Analysis with the 64-bit ALU 152

9.4 Summary 155

10 Conclusions and Future Work 157

10.1 Summary and Implications of Results 160

10.1.1 SysRel: Interactive Reliability CAD Tool 160

10.1.2 Thermal-Aware Reliability Analysis in SysRel 163

10.1.3 Reliability Comparison of Copper and Aluminum Metallizations 164

10.1.4 Non-blocking Via Analysis with Copper Metallization 165

10.2 Potential Applications of SysRel 166

10.3 Future Directions 167

10.3.1 Optimization of SysRel 167

10.3.2 Enhanced Electromigration Lifetime Model 168

10.3.3 Thermal Analysis and Management in 3D IC Technology . . . 169

10.3.4 Impact of Process Variations 170

A SysRel Source Code Overview 171

A.1 Main Application Classes 171

A.2 Graphical Interface Classes 172
A.3 File Parser Classes 173
A.4 Corner-stitched Data-structure Classes 175
A.5 Layout and Tree Representation Classes 175
A.6 Reliability Computation Class 177

B Cadence to ANSYS Geometric Model Conversion: cds2inp 179

C Matlab Source Code for TProfile 189

D SysRel Release 193

D.1 SysRel v1.1 193
D.2 SysRel v2.0 193

List of Figures

1-1 Cross-section of an Integrated Circuit showing multi-layer metal interconnects on top of a device layer. 24

1-2 Gate and interconnect delay, and interconnect reliability requirement in different technology generations 25

1-3 Electromigration-induced failure modes 27

1-4 Immortality Condition in a straight line interconnect 28

1-5 Interconnect architecture schemes 32

1-6 Example of an interconnect tree, the fundamental reliability unit for electromigration analysis 33

1-7 Cross-section of a 3D IC with 3D trees and inter-wafer vias 35

1-8 Different types of vias/contacts for 3D ICs 36

1-9 Thermal management using optimal placement of microfluidic thermal connects in a face-to-face and back-to-back bonded 3D IC. 37

1-10 Flow diagram of the hierarchical circuit-level reliability analysis in ERNI 40

1-11 A screen-shot of MAJIC with a view of menu items for reliability analyses in ERNI 41

2-1 Schematic diagrams of designed interconnect trees for electromigration testing 46

2-2 Side-view schematic of void formation in via-above and via-below interconnects 48

2-3 Calculating L_{max} in an interconnect tree. 50

3-1 Failure Rate versus time for typical Integrated Circuits (bathtub curve). 60

3-2 Schematic illustration of various system reliability models 64

4-1 Cross-section of an Al interconnect with schematic illustration of diffusion pathways 74
4-2 Cross-section of a Cu interconnect with schematic illustration of diffusion pathway 75
4-3 A composite model of atomic diffusivity versus interconnect line width 77
4-4 Atomic diffusivity of Al and Cu metallization technologies at different temperatures 78
4-5 Interconnect lifetimes of various types of lines in Cu and Al metallization technologies 80
4-6 Cross-over temperature versus current density plot 80
4-7 Interconnect lifetimes of various types of lines in Cu and Al metallization technologies in typical operating condition 81

5-1 Screen shot of SysRel's desktop pane 87
5-2 Corner-stitched representation of a single plane with multiple tiles in a Magic layout 89
5-3 Extraction of an interconnect tree from corner-stitched representation of tiles 91
5-4 Reliability analysis flow diagram in SysRel 92
5-5 The 32-bit comparator circuit layout analysis in SysRel 94
5-6 Layout technique for Al polygranular to bamboo type conversion in an interconnect line 97
5-7 The 3D 32-bit comparator circuit layout for SysRel reliability simulation 99

6-1 The largest cell, FAX1, characterization with SysRel 103
6-2 Hierarchical layout view in SysRel 104
6-3 Cell-based adder layout's level 2 view in SysRel 105

7-1 Multi-level Cu interconnects with different liner thicknesses 110

7-2 Straight line via-to-via Cu dual-damascene test structure for electromigration experiments . . . 111
7-3 Menu options for reliability analysis in SysRel with non-blocking vias in a circuit layout . . . 113
7-4 Schematic illustration of interconnect tree merging with a non-blocking via . . . 114
7-5 Dynamic conversion of a non-blocking via into fully blocking in a merged tree . . . 115
7-6 Illustration of interconnect tree merging in a 3-input nand gate layout 116
7-7 Time-to-failure dependence on interconnect length . . . 117
7-8 Mortal trees in the 32-bit comparator circuit layout when all vias are assigned non-blocking . . . 120
7-9 Test structures to investigate the line length dependence of reliability in Cu interconnects . . . 121

8-1 Z-axis dimensions with a substrate height of $10um$ in cds2inp . . . 128
8-2 ANSYS thermal simulation with a bulk CMOS transistor . . . 130
8-3 Energy dissipation in the NMOS transistor of an inverter at different output frequencies . . . 131
8-4 MOSFET transient temperature plot at $2.5MHz$ square wave heat generation frequency . . . 132
8-5 MOSFET transient temperature plot at $1GHz$ square wave heat generation frequency . . . 133
8-6 Package-level thermal modeling to compute uniform temperature at the bottom surface of a substrate . . . 135
8-7 Thermal impulse response at the top surface of a substrate . . . 136
8-8 Power density profile of an inverter cell in a $30um \times 30um$ substrate . 137
8-9 Temperature profile calculated using TProfile of an inverter cell in a $30um \times 30um$ substrate . . . 137
8-10 Modeling the temperature profile in a straight line interconnect . . . 138

8-11 Modeling the temperature rise due to joule heating in multi-level interconnects 140
8-12 Thermo-file menu options in SysRel for thermal-aware reliability analysis 141
8-13 Layout-level non-uniform temperature profile in the 32-bit comparator circuit layout 142

9-1 The 64-bit Arithmetic and Logic Unit layout in Magic 147
9-2 Immortality in interconnect trees in the 64-bit ALU circuit with Cu metallization 148
9-3 Nonuniform temperature profile of the 64-bit ALU circuit in bulk CMOS technology at $Vdd = 5V$ and operating frequency= $250MHz$. 153

10-1 Straight line interconnect and interconnect tree structures in circuit layout 158
10-2 Screen shot of SysRel desktop pane 161
10-3 Multiple vias at a junction in wide metal interconnects 169

B-1 Cadence circuit layout to ANSYS geometry conversion 180

List of Tables

3.1 Tree types with lognormal lifetime distribution 68

3.2 Improved tree types with lognormal lifetime distribution 68

3.3 Full-chip reliability metrics of circuit layout with various types of mortal trees 68

4.1 Reported (jL) product thresholds for immortality conditions in Al and Cu technologies 72

4.2 Immortality length in various technology generations with Al and Cu metallizations 73

4.3 Material parameters used for calculating diffusivity of Al polygranular type lines 74

4.4 Material parameters used for calculating diffusivity of Al bamboo type lines 75

4.5 Material parameters used for calculating diffusivity in Cu metallization technology 76

4.6 Technology-specific parameter description and values for lifetime calculation 79

4.7 Worst case interconnect lifetimes of various types of Cu and Al lines . 82

5.1 Menu bars and corresponding operations in SysRel 88

5.2 A sample *.mag* layout file 89

5.3 SysRel simulation results of hierarchical reliability analysis with Cu metallization in the 32-bit comparator circuit layout 95

5.4 SysRel simulation results of hierarchical reliability analysis with *Al* metallization in the 32-bit comparator circuit layout 96
5.5 SysRel simulation results of hierarchical reliability analysis with *Cu* metallization in the 3D 32-bit comparator circuit layout 98

6.1 Cell reliability characterization data format in *sysrelcell.char* file . . . 102
6.2 Statistics of Fundamental Reliability Units in the 4-bit adder layout . 105
6.3 Statistics of Fundamental Reliability Units in the 32-bit comparator circuit layout 106
6.4 Time required for layout drawing in SysRel 107
6.5 Time required for extracting interconnect trees in SysRel 107
6.6 Computational efficiency with cell-based reliability analysis in SysRel 107

7.1 Input parameters for non-blocking via analysis defined in *sysrel.emparam* file 112
7.2 Reliability simulation with various degrees of non-blocking vias in the 32-bit comparator circuit layout 119

8.1 Electrical and thermal analogy for modeling steady-state heat flow . . 125
8.2 Material parameters for thermal simulation in ANSYS 129
8.3 Maximum temperature increase due to interconnect joule heating in different technology generations 140
8.4 Thermal-aware reliability simulation results with *Cu* metallization in the 32-bit comparator circuit layout 143

9.1 Operations in the 64-bit Arithmetic and Logic Unit 146
9.2 The number of cells of various types in the 64-bit ALU circuit 147
9.3 Full-chip reliability analysis with the 64-bit ALU in different metallization technologies 149
9.4 Times to void nucleation and void growth in various interconnect technologies 151

9.5 Cu atomic diffusivity requirements for equivalent circuit-level reliability in the 64-bit ALU with Al and Cu metallizations 151
9.6 Thermal-aware full-chip reliability report for the 64-bit ALU with Cu/SiO_2 metallization in bulk CMOS technology 154
9.7 Thermal-aware full-chip reliability report for the 64-bit ALU with Cu/SiO_2 metallization in SOI technology 155

Chapter 1

Introduction

In the pursuit of higher performance and integration, Integrated Circuit (IC) technology is heading towards the nanotechnology era. The gate length of a state-of-the-art active device, such as a metal oxide silicon field-effect transistor (MOSFET), is in the range of 45 to 50 nm allowing millions of such devices to be fabricated in a single chip [1]. In addition to achieving higher density, smaller devices increase speed due to higher drive current during the "on" state. While more and more devices are desirable for integrating more functionality in a single chip, interconnecting the devices using metal wires takes up an even greater percentage of space. The on-chip metal lines referred to as interconnects are fabricated by deposition of metal and dielectric materials, lithography of patterned features, and selective etching. ICs can have up to 14 layers of metal interconnects as predicted by the International Technology Roadmap for Semiconductors (ITRS) 2003 edition (figure 1-1) [2]. The dielectric material, referred to as inter-layer dielectric (ILD), provides insulation between interconnects and mechanical stability for the multi-level interconnect structure.

Concurrent to device dimension scaling, interconnect width is shrinking while total interconnect length is increasing exponentially. According to the International Technology Roadmap for Semiconductors, 688 meters of active metal wiring per centimeter square area are required to construct a high-performance chip in 2004. Interconnect delay related to its resistance (R), capacitance (C), and in some cases inductance (L) has become dominant over gate delay as shown in figure 1-2(a). Moreover, the

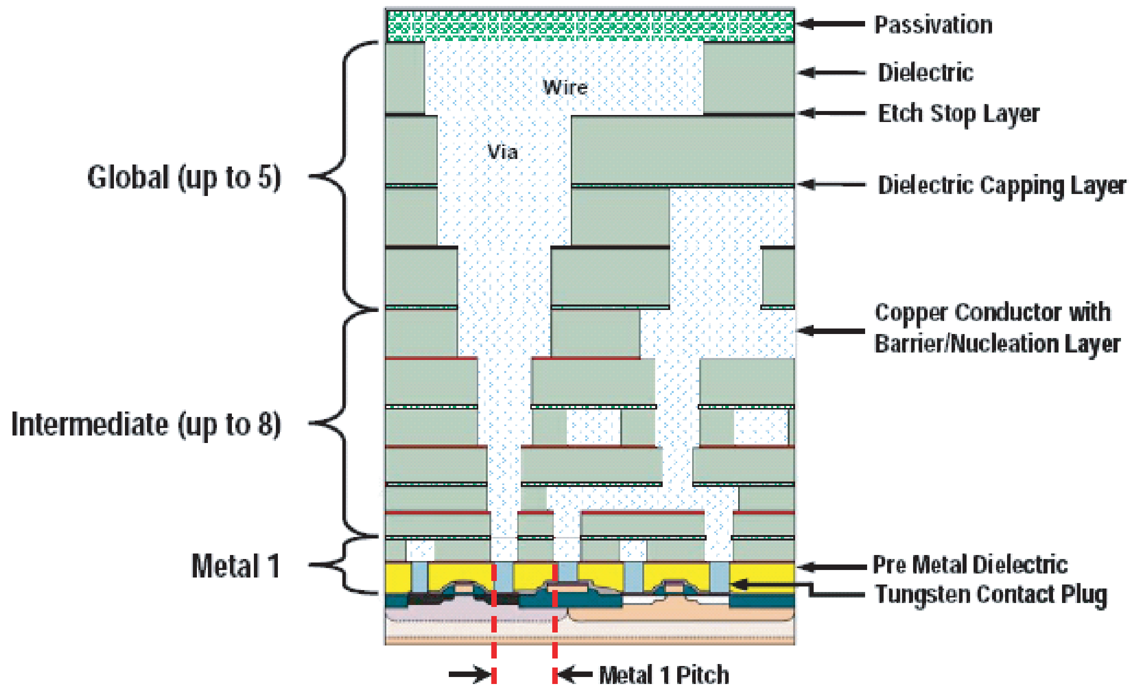

Figure 1-1: *Cross-section of an Integrated Circuit showing multi-layer metal interconnects on top of a device layer. Source: the ITRS [2].*

increased capacitance in interconnects leads to an increase in net switching power. Since the invention of ICs, Aluminum (Al) and its alloys have been used in the metallization layers. Silicon Dioxide (SiO_2) is used as the ILD material. Copper (Cu) has been replacing Al as the material of choice for interconnects due to its lower sheet resistance while new ILDs with lower dielectric constant (low-k dielectrics) are under active study. However, as seen in figure 1-2(a), interconnect delay is still dominant with the introduction of Cu and low-k dielectrics beyond 180nm technology generations. Therefore, interconnect design and analysis have become as important as device design [3].

1.1 Interconnect Reliability

From the interconnect design perspective, reliability is the extent to which the interconnects maintain signal integrity and produce desired functionality over the lifetime

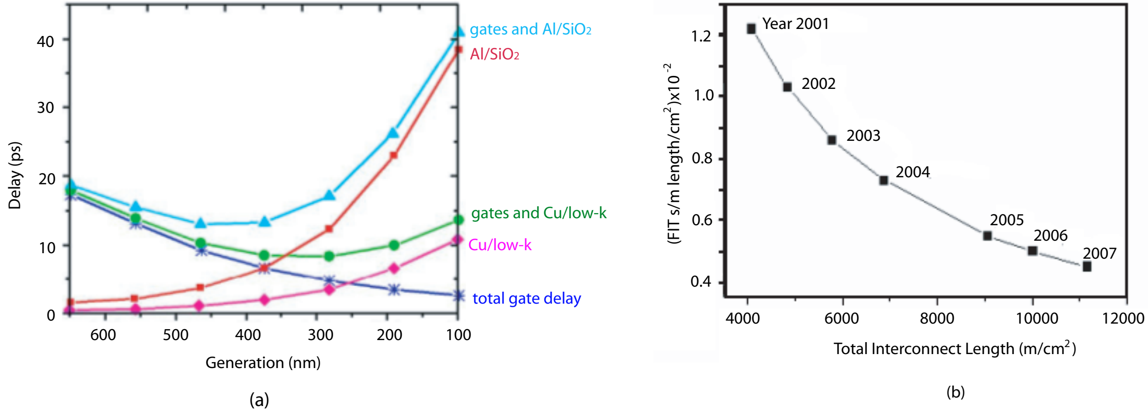

Figure 1-2: *(a) Gate and interconnect delay in different technology generations [4]. (b) Interconnect reliability requirement versus total interconnect length in different years. Here FIT refers to Failure Unit. Source: the ITRS 2001 edition [2].*

of a chip. Stress conditions, such as current density and temperature, during circuit operation affect the interconnect reliability. As the performance of present day ICs increases, a more stringent operating condition in the metal interconnects is expected with a service temperature of 105^oC and maximum current density of 0.5 MA/cm^2 [2]. In addition, the interconnect reliability requirement has become more stringent as shown in figure 1-2(b). Electromigration is the primary interconnect reliability concern.

1.1.1 Electromigration Phenomenon

Electromigration is the transport of atoms in a conducting material due to momentum transfer from flowing electrons. When a difference in electrical potential is applied to an interconnect, electrons flow from low potential (cathode) to high potential (anode) terminal. The metal atoms start to diffuse along the electron flow direction due to scattering. Assuming that the electron flow direction has a positive sign, the "electron wind" force on atoms can be expressed as

$$F_{elec} = -q^* E = -Z^* e\rho j \tag{1.1}$$

where $q^* = Z^*e$ is the effective atomic charge, Z^* is the effective atomic charge number, e is the fundamental electron charge, $E = \rho j$ is the electric field, ρ is the electrical resistivity of the metal, and j is the current density. As current density opposes the electron flow direction, a negative sign for j attributes to a positive sign in F_{elec}.

In an interconnect terminating at diffusion barriers such as Tungsten (W) filled vias or Tantalum (Ta) liners, the "electron wind" force creates tensile stress near the cathode where the atoms deplete and compressive stress near the anode where the atoms accumulate [5]. The resulting stress gradient leads to a mechanical driving force, referred to as the back-stress force, which opposes the electromigration wind force. The back-stress force is expressed as

$$F_{mech} = \Omega \frac{d\sigma}{dx} \tag{1.2}$$

where Ω is the atomic volume, σ is the stress, and x is the distance measured along the length of the line. Due to a negative value of $\frac{d\sigma}{dx}$, F_{mech} has a negative value which is consistent with our sign convention here.

According to the 1-D Korhonen model [6], the atomic flux, J_a, can be expressed as a function of F_{elec} and F_{mech} using

$$\begin{aligned} J_a &= \frac{DC_a}{kT}(F_{elec} + F_{mech}) \\ &= \frac{DC_a}{kT}(-Z^* e\rho j + \Omega \frac{d\sigma}{dx}) \\ &= -\frac{DC_a}{kT}(Z^* e\rho j - \Omega \frac{d\sigma}{dx}) \end{aligned} \tag{1.3}$$

where C_a is the atomic concentration, D is the atomic diffusivity, k is the Boltzmann's constant, and T is the temperature. Equation 1.3 suggests that $|F_{elec}| > |F_{mech}|$ is required to have a non-zero atomic flow for electromigration.

1.1.2 Electromigration Failure Modes

A void nucleates at the cathode end of an interconnect line when the tensile stress exceeds the critical stress necessary for void nucleation, σ_{crit_nuc}. Once a void nucleates, it can grow to larger volume even spanning the whole interconnect width as shown in figure 1-3(a). In both Al and Cu metallization schemes, void nucleation and growth would result in a resistance increase of the line. In the event of a fully spanning void, conductive refractory metal under layers or over layers in Al technology will shunt the current to prevent an open-circuit failure. An open-circuit failure is more common in Cu technology due to the absence of conductive refractory layers. While an open-circuit due to voiding would cause functional failure, the resistance increase would add extra delay in signal propagation resulting in timing violations in timing critical circuits.

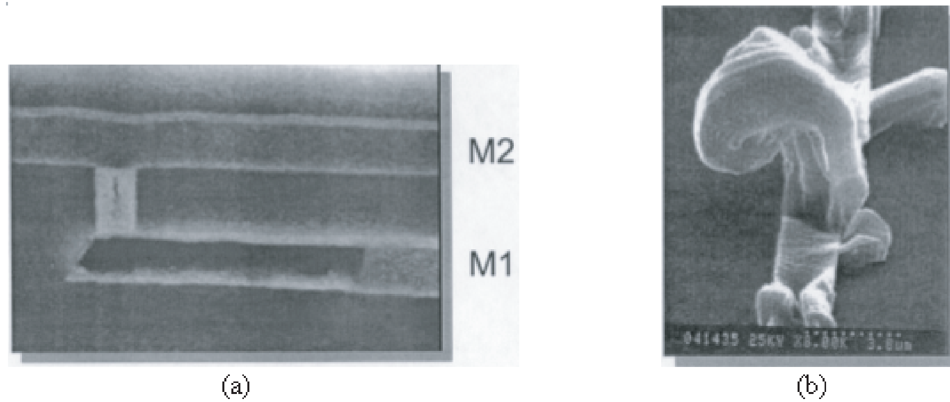

Figure 1-3: *Electromigration-induced failure modes; (a) Tensile stress at the cathode-end of an interconnect can cause voiding [7]; (b) Compressive stress at the anode-end can cause extrusion [8].*

Compressive stress at the anode end of both Cu and Al interconnect lines can cause metal extrusion where the atoms accumulate (figure 1-3(b)). However, extrusions are not observed in service operating condition for either Al or Cu as metal interconnects are surrounded by rigid ILD material that suppresses the formation of hillocks. Current densities greater than 5 MA/cm^2 are required to observe extrusion during electromigration testing condition [9]. Therefore, voiding is the primary mode

of electromigration failure in ICs.

1.1.3 Immortality Condition

In steady state, the "electron-wind" force balances the back-stress force resulting in zero atomic flux and a linear time-invariant stress along the interconnect line as illustrated in figure 1-4 [10]. Equation 1.3 in steady state becomes

$$Z^* e\rho j = \Omega \frac{d\sigma}{dx} \tag{1.4}$$

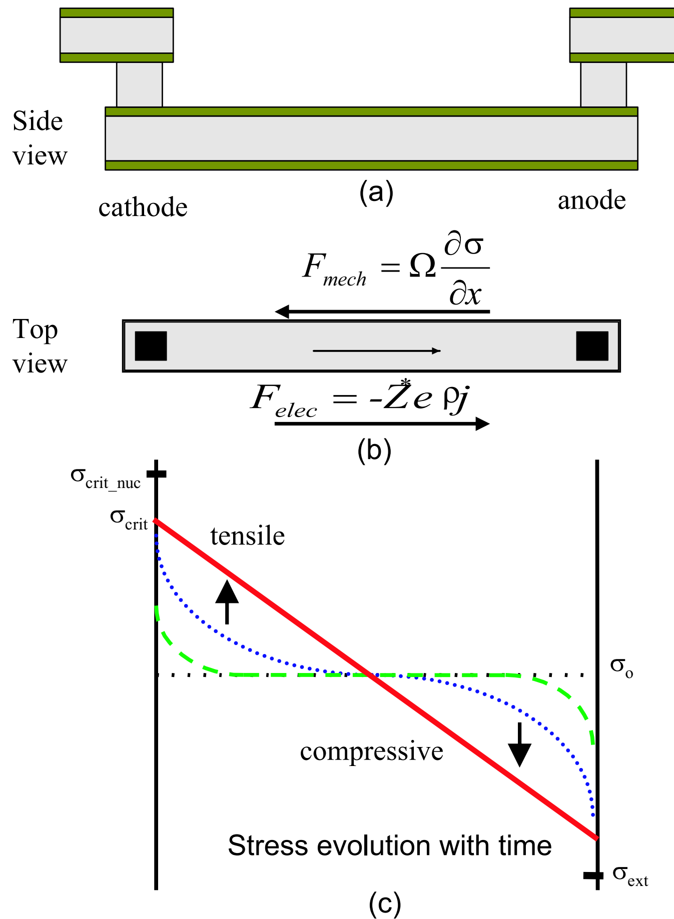

Figure 1-4: *Immortality Condition in a straight line interconnect. (a) Side view (b) Top view (c) Stress as a function of location along the interconnect at different times.*

Rewriting equation 1.4 for a straight line interconnect with length, L, we can get

an expression for current density and line length product (jL)

$$jL = \frac{\Omega \Delta\sigma_{max}}{\rho e Z^*} \tag{1.5}$$

where $\Delta\sigma_{max}$ is the maximum stress difference between the cathode and anode ends. If the stress at the cathode end is σ_{crit}, $\Delta\sigma_{crit}$ can be written as

$$\Delta\sigma_{crit} = 2(\sigma_{crit} - \sigma_o) \tag{1.6}$$

Assuming the initial stress, σ_o, is zero,

$$\Delta\sigma_{crit} = 2\sigma_{crit} \tag{1.7}$$

If the critical tensile stress for void nucleation, σ_{crit_nuc}, is greater than the maximum steady state tensile stress, σ_{crit}, developed at the cathode end, no void will form and the line will not fail. Using equation 1.5, we can derive a critical current density line length product (jL) that defines the immortality condition as

$$jL < (jL)_{crit_nuc} = \frac{\Omega \Delta\sigma_{crit_nuc}}{\rho e Z^*} \tag{1.8}$$

which means short lines or lines stressed at low current densities that satisfy $(jL) < (jL)_{crit_nuc}$ are immune from void nucleation, and thus "immortal" from void nucleation limited failure.

On the other hand, if (jL) exceeds $(jL)_{crit_nuc}$, a void will nucleate in the metal line. However, if the void does not completely block the current flow, the void nucleation will not be fatal. In the presence of conducting over layers and under layers, the void can continue to grow as the current is shunted around it through the refractory metal layers until the electron wind force and back-stress balance again. Under such circumstances, the line is immortal if the resistance increase associated with the void volume is lower than the maximum acceptable resistance increase, $\Delta R_{crit}^{\max}$, and the maximum compressive stress in the line does not cause yielding or fracture of the sur-

rounding dielectric material. The immortality condition due to resistance saturation is given by [11]

$$(jL)_{crit_sat} = \frac{\frac{\rho}{A}}{\frac{\rho_l}{A_l}} \frac{\Delta R_{crit}^{\max}}{R} \frac{2\Omega B}{Z^* e\rho} \tag{1.9}$$

where ρ and A are the resistivity and cross sectional area of the high-conductivity metal, respectively, ρ_l and A_l are the resistivity and cross sectional area of the shunt layer, respectively, and R is the initial resistance of the line. It is apparent that $(jL)_{crit_sat} > (jL)_{crit_nuc}$ as resistance saturation will occur after void nucleation.

1.1.4 Electromigration Lifetime Model

Electromigration testing is conducted at accelerated conditions with temperatures as high as $300^o C$ and current densities greater than 2.0 MA/cm^2. To estimate the lifetime of a population of interconnects at service conditions, the well known Black's equation is used to extrapolate the median-time-to-failure (MTTF) to service conditions [12]. Black's equation is expressed as

$$t_{50} = Aj^{-n} e^{\frac{E_a}{kT}} \tag{1.10}$$

where A is a constant independent of temperature and current density, j is the current density, n is the current density exponent, E_a is the activation energy for the rate-limiting diffusion path, k is the Boltzmann's constant, and T is the temperature. A current density exponent $n = 1$ is consistent with void growth limited failure since the rate of unconstrained void growth is proportional to the current density [13]. An exponent $n = 2$ indicates void nucleation limited failure in the model based on the Korhonen analysis [6]. Although variations of the original Black's model have been proposed in [14, 15], Black's equation is still generally used to estimate the lifetime of an interconnect due to electromigration.

1.2 Interconnect Architecture Schemes

Al and Cu are the primary choices of metal for interconnects in modern day ICs. IBM introduced Cu technology in commercial ICs in 1998. Since then, Cu has been replacing Al as the material of choice for interconnects due to its lower sheet resistance. While the electromigration behavior of Al interconnects is quite well understood [16, 17], the reliability of Cu is still under active study. Due to the differences in chemical properties between Cu and Al, the fabrication process for Cu interconnects is drastically different from that for Al interconnects. The differences in processing and architecture schemes lead to the differences in electromigration reliability.

1.2.1 Aluminum Metallization Technology

As illustrated in figure 1-1, metal interconnects are surrounded by ILD materials most commonly SiO_2. Al chemically reacts with SiO_2 to form alumina, which eliminates atomic diffusion of Al into the surrounding ILD and along the interface between Al and SiO_2. As a result, Al metallization is processed by a subtractive etching method in which the patterned lines are formed by etching the deposited blanket Al film. Architecturally, Al interconnects have thick, highly electromigration-resistant refractory metal layers, which are usually made of Titanium Nitride (TiN), serving as anti-reflection coatings at the top of the lines (see figure 1-5(a)). Similar under layers have also been included, which serve as seed layers for the via-fill process. Tungsten (W) filled vias are used to connect layers of Al metallization. Metal under and over layers serve as shunt for electron flow and W-filled vias serve as fully blocking boundaries for electromigration.

1.2.2 Copper Metallization Technology

Cu does not chemically reduce SiO_2 like Al does. Furthermore, a suitable etchant for Cu thin films is unavailable for commercial use. Consequently, Cu interconnects are fabricated by the damascene method, in which a trench is first etched into a blanket layer of ILD before filling it with Cu by electroplating. Since Cu undergoes field

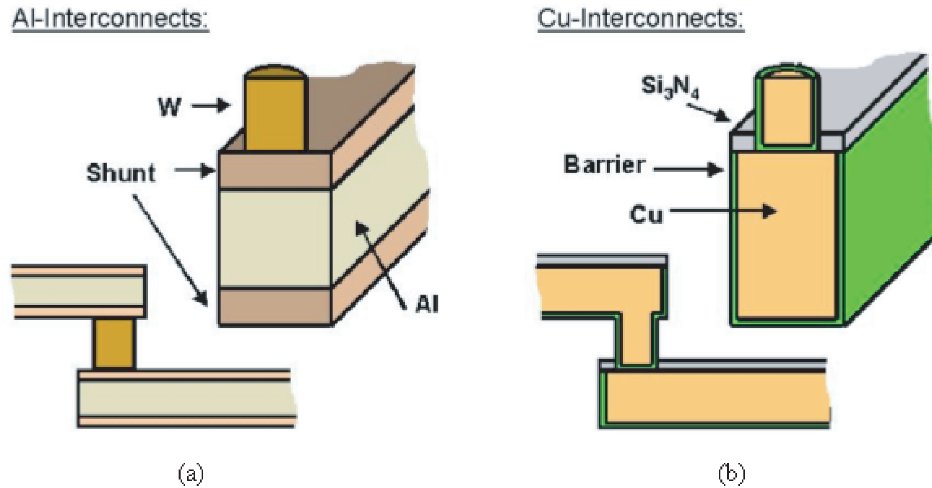

Figure 1-5: *Interconnect architecture schemes; (a) Al interconnect, with W-filled vias and conducting shunt layers at the top and bottom of the interconnect line. (b) Dual-damascene Cu interconnect, with Cu-filled vias, thin refractory liners at the sides and bottom of the line, and a dielectric capping layer at the top of the line.*

enhanced diffusion in most dielectric materials (including SiO_2), in order to prevent Cu atoms diffusing into the device layer, thin refractory metal layers consisting of Tantalum (Ta) or Tantalum Nitride (TaN) are placed at the sides and bottom of the Cu interconnect lines (See figure 1-5(b)). The Cu lines are capped with a dielectric diffusion barrier, which is usually made of Silicon Nitride (Si_3N_4). Cu-filled vias are used to connect multiple layers of metallization.

1.3 Fundamental Reliability Unit

An important concept in hierarchical reliability analysis is the classification of Fundamental Reliability Units (FRU). By definition, an FRU is a component in reliability analysis, contributing to the underlying failure mechanism, which can be treated independently from other such components. During a hierarchical reliability analysis, FRUs are first identified and extracted from a circuit layout. FRU based reliability analysis reduces the complexity of the problem and provides the flexibility of adding new failure units for different reliability phenomena.

Straight line via-to-via structures are used for most electromigration experimental

work. However, in real circuits, multiple straight line segments are connected at junctions and many such junctions are connected within the same layer of metallization. An "interconnect tree" is a unit of continuously connected high-conductivity metal lying within one layer of metallization terminating at the vias or diffusion barriers [18]. An example of an interconnect tree is shown in figure 1-6.

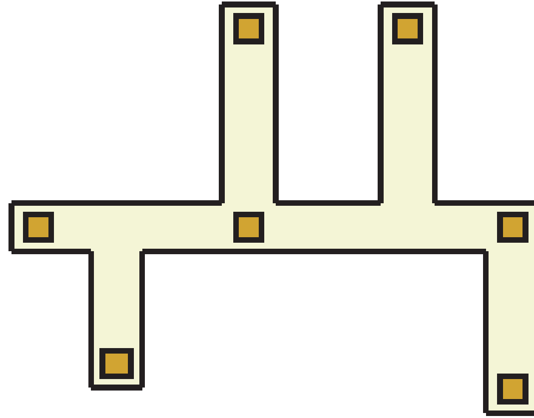

Figure 1-6: *Example of an interconnect tree, the fundamental reliability unit for electromigration analysis.*

It has been established that an interconnect tree is the appropriate fundamental reliability unit for the circuit-level reliability assessment of Al-based metallization [18]. The fully blocking boundaries formed at the W-filled vias confine atomic diffusion within an interconnect. Materials within an interconnect tree can diffuse freely between the segments and the stress evolutions in different segments of a tree are coupled.

On the other hand, Cu-filled vias with thin liner material at the bottom in Cu metallization technology may or may not act as fully blocking boundary for atomic diffusion [10]. Liner ruptures have been observed in some experimental work which allow materials to freely diffuse between different layers of metallization. In such cases an interconnect tree cannot be treated as the fundamental reliability unit for Cu-based technology.

1.4 Three-Dimensional (3D) Integrated Circuit Technology

Technology scaling has posed limitations on overall system performance by degrading the interconnect delay, increasing the power consumption due to interconnect capacitance, and increasing the number of longer global and semi-global lines in a chip. The need for a long-term solution to enhance performance in successive technology generations is apparent. One such long-term solution is the three-dimensional (3D) Integrated Circuit technology.

1.4.1 Technology Concept

The main idea behind 3D integration is to form multiple device layers along the third axis (z axis) and lower the interconnect lengths by connecting the devices in these layers vertically. This has been accomplished by bonding multiple wafers fabricated with different or similar technologies [19, 20] as well as by fabricating multiple device layers on the same wafer [21, 22] using the epitaxial growth of Si. In a wafer bonding technology, each device-interconnect layer is fabricated separately on different wafers with the same or different technologies, and then the wafers are bonded with each other using a bonding layer of Cu, polymeric adhesives, or plain oxide-to-oxide (ILD material) bond.

The wafers in a 3D stack are electrically connected using high aspect ratio vias or contacts. When bonding is complete, 3D ICs have vertical interconnects of significantly longer length than vias or contacts in conventional or 2D ICs. Moreover, the 3D circuits fabricated with a wafer bonding technology have two different types of vertical interconnects as shown in figure 1-7. The Cu-filled inter-wafer vias connect multiple interconnect trees from different wafers. At the bonding surface, the adjacent metallization layers from two wafers can also be connected with vertical Cu lines. The vertical Cu lines create a new type of trees, referred to as a “3D tree”, which expands between two different wafers.

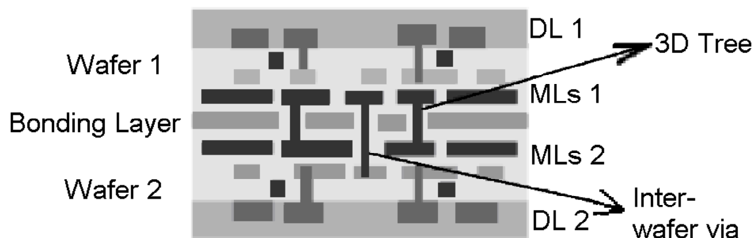

Figure 1-7: *Cross-section of a 3D IC with 3D trees and inter-wafer vias. Here DL and MLs correspond to device and metal layers, respectively.*

Although the concept of 3D integration emerged as early as 1979 [23], significant research work was done only after the early 90's as the limitations of technology scaling became apparent. 3D integration is an attractive solution as it can significantly reduce the number of long wires by mapping a 2D circuit into different layers [24]. Moreover, the total number of repeaters for long and intermediate wires will also decrease, resulting in a higher density and lower interconnect-limited chip area. Consequently, high-performance microprocessors and programmable logic devices are attractive applications for 3D integration.

Another promising advantage of 3D integration is its ability to integrate heterogeneous technologies into a single 3D chip. Future system-on-chip (SOC) applications will consist of digital, analog, RF, and optical components on the same die [25]. Using 3D integration, each unit can be fabricated on separate wafers with its own optimized process technology, and then integrated vertically to form a 3D SOC [19, 20].

1.4.2 Layout Methodology for 3D Circuits

We developed a comprehensive layout methodology for 3D ICs and implemented it in 3D-Magic [26, 27]. In order to facilitate the layout of 3D ICs, all inter-wafer vias or contacts are generalized into three major categories. The three categories of vias are sufficient for defining almost all types of interconnection between wafers in a 3D IC. Figure 1-8 shows the three categories of vias; connected-to-top, connected-to-bottom, and through-wafer vias. Categorizing vias in such a way allows the layout methodology to be flexible enough to support different types of bonding schemes,

such as face-to-face, face-to-back, or back-to-back[1].

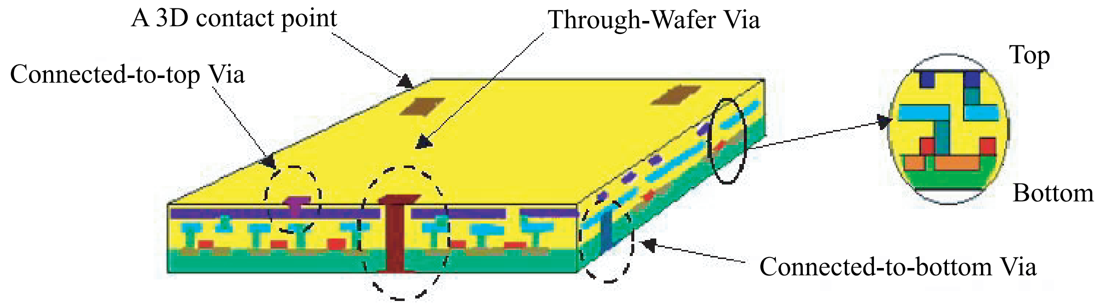

Figure 1-8: *Different types of vias/contacts for 3D ICs.*

The availability of 3D-Magic has led to interesting research with a wide range of layout-specific circuit analyses, from performance comparison of 2D and 3D circuits to layout-specific reliability analyses in 3D circuits. Using 3D-Magic, researchers have designed and simulated an 8-bit encryption processor mapped into 2D and 3D FPGA layouts [28]. The layout methodology has been adopted for 3D CAD tool research at MIT [29] . It has built the bridge between synthesis and layout tools for 3D circuits and allowed the demonstration of physical layouts of circuits produced by the 3D place and route tool. Moreover, it is an essential element of a novel Reliability Computer Aided Design (RCAD) tool, ERNI-3D [26].

1.4.3 3D IC Technology with Thermal Management

While 3D IC technology has its advantages, extracting heat generated by power dissipation in inner wafers is quite a challenge. Conventional packaging technology allows heat extraction from one side, typically the Si-substrate side, by attaching a Cu heat spreader and fin heatsink. When using such a method for heat extraction, heat flux from wafers further away from the heatsink would flow through the 3D stack leading

[1]Face-to-face bonding refers to bonding metal interconnect side of a wafer to the metal interconnect side of another wafer while forming the 3D IC stack. Back-to-back bonding refers to bonding Si-substrate side of a wafer to the Si-substrate side of another wafer in forming the 3D IC stack.

to an increased power density on a smaller footprint. Therefore, temperature rise in 3D circuits can be significantly higher than that in conventional ICs [30].

Flexibility in bonding schemes, such as back-to-back and face-to-face, can be exploited to address thermal management in 3D ICs. We proposed a vision for 3D IC technology that efficiently incorporates thermal management [31]. Figure 1-9 presents the main idea in such a 3D IC technology.

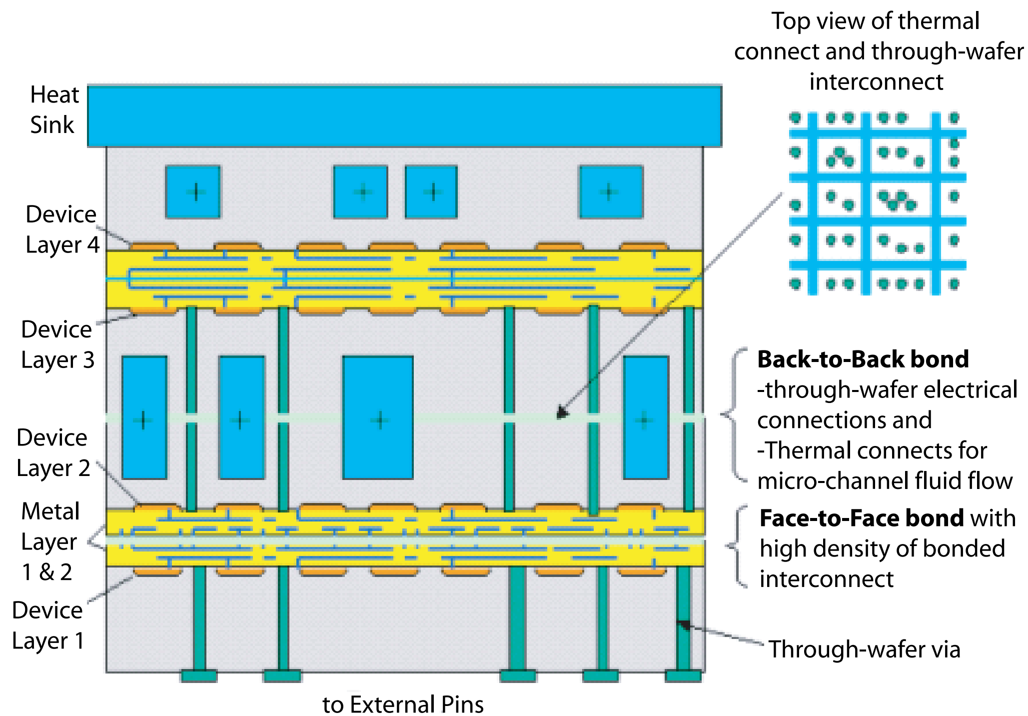

Figure 1-9: *Thermal management using optimal placement of microfluidic thermal connects in a face-to-face and back-to-back bonded 3D IC.*

Starting at the bottom of the stack, there is a face-to-face bond between the first two wafers. This face-to-face bond allows high-density inter-wafer connection and eliminates area trade-off between through-wafer via and active devices that is present in back-to-back (substrate-to-substrate) or face-to-back bonding. In the figure, there are through-wafer vias through the substrate of the first wafer to get signals out for the connections to external pins. The number of such through-wafer vias is comparable to the number of pins and significantly lower than that of inter-wafer vias at a high-

density face-to-face bond.

Next, wafer no. 2 is bonded with wafer no. 3 using a back-to-back bond. This back-to-back or substrate-to-substrate bond results in a bonded silicon interface which can be used to contain micro-channels, termed *thermal connects*, for heat removal with fluid flow. The thermal connects or micro-channels are wet-etched in the back of the individual substrates before bonding. The thermal connects can be criss-crossed among through-wafer vias, as shown in figure 1-9, as well as laid out manually with a higher density at the chip hotspots.

The 3D stack can grow further upward using a face-to-face bond with wafer no. 3. In figure 1-9, the topmost wafer is wafer no. 4 and its substrate is connected to a heat sink/heat removal device. As it is apparent from the structure, every wafer in the 3D stack except wafer no. 1, has a heat removal device connected to its substrate. Therefore, thermal connects can also be placed in the substrate of wafer no. 1 while bonding to the package. Thus, every wafer in a 3D stack has heat removed through its substrate and the heat removal problem can be reduced to that of a conventional 2D IC.

1.4.4 Interconnect Reliability in 3D ICs

Electromigration phenomenon and reliability concerns described in section 1.1 are equally applicable to 3D circuits. In addition to treating interconnect trees as the fundamental reliability unit, special attention is required for the through-wafer and inter-wafer vias in 3D circuits. Through-wafer vias can be significantly longer in height than conventional vias as illustrated in figure 1-9. The reliability impact due to increased height as well as the presence of a bonded interface is under active investigation [32]. Depending on the failure characteristics observed in through-wafer and inter-wafer vias, it may be necessary to treat them as additional fundamental reliability units in 3D ICs.

As discussed earlier, the stack effect in 3D circuits can potentially increase the service temperature of the interconnects. According to equation 1.10, MTTF of an interconnect tree is exponentially dependent on its temperature. Due to high

power dissipation, high temperature rise is already of great concern in modern day conventional ICs. Consequently, the 3D counterparts of such chips are expected to pose a greater reliability challenge due to higher temperature.

1.5 Circuit-Level Reliability Analysis

The conventional approach to meet reliability goals in an integrated circuit has been to use simple and conservative design rules based on current density in a wire segment. However, this simplicity and conservatism lead to limited performance in newer technology generations [18]. Models and techniques have been developed to make realistic reliability assessments of interconnects during the design and layout process (Reliability Computer Aided Design, RCAD), so that the reliability data can be fed back, and changes can be made promptly before the fabrication process to achieve optimum reliability and performance. BERT (BErkeley Reliability Tool) [33], iTEM [34], ERNI (Electromigration Reliability in Networked Interconnects) [35], and ERNI-3D [27] are examples of reliability analysis tools for Al-based interconnect technology that have been previously developed.

BERT and iTEM calculate the overall reliability of a given layout based on the reliability estimations from individual straight line segments. However, the reliability of a segment depends strongly on the activities in the linked segments. In other words, segments can not be treated as independent reliability units. Interconnect trees need to be considered as the fundamental reliability units as discussed in section 1.3. Another limitation in BERT and iTEM is the lack of immortality checking in interconnect segments. As illustrated in section 1.1.3, not all interconnect trees are prone to electromigration failure. Neglecting immortality condition in interconnect trees leads to redundant computation in analysis, and more importantly, leads to overly conservation and incorrect reliability estimates. ERNI and ERNI-3D addresses the limitations of earlier tools.

1.5.1 Electromigration Reliability in Networked Interconnects (ERNI and ERNI-3D)

The RCAD tool ERNI (Electromigration Reliability in Networked Interconnects), developed at MIT [35], allows process-sensitive and layout-specific reliability assessments of a fully or partially laid-out integrated circuit. Figure 1-10 shows the flow diagram of ERNI's operations. First, the interconnect trees are extracted from a

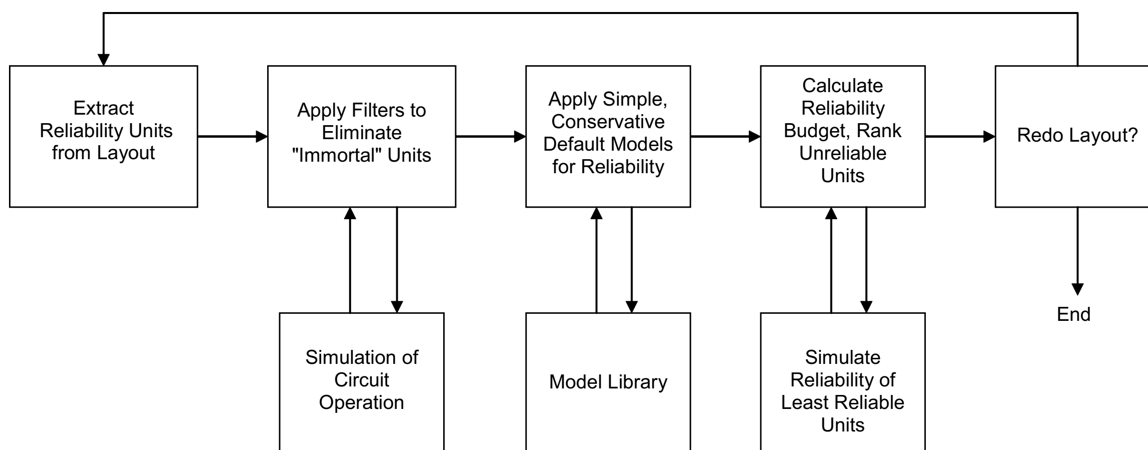

Figure 1-10: *Flow diagram of the hierarchical circuit-level reliability analysis in ERNI.*

circuit layout and categorized into mortal and immortal trees based on the current density and line length product (jL). Further computation goes on with only the mortal trees, and a reliability figure for each tree, in terms of MTTF, is obtained after applying the default electromigration model [18].

The code for ERNI is written in Java 2 (JDK 1.2). It is a client extension to MAJIC, a layout parser and viewer also written entirely in Java [36]. Lots of data-structures and algorithms in MAJIC are based on MAGIC, an IC layout editor developed at UC Berkeley and widely used in academia [37]. Using MAJIC, users can view a circuit layout, and apply reliability analysis by choosing different filtering algorithms from the "ERNI" menu. Figure 1-11 shows a screen-shot of MAJIC with the display of available options in the "ERNI" menu. ERNI operates on 2D IC layouts with multiple metallization layers created using MAGIC.

The reliability analysis concept in ERNI has been extended to develop a distinct

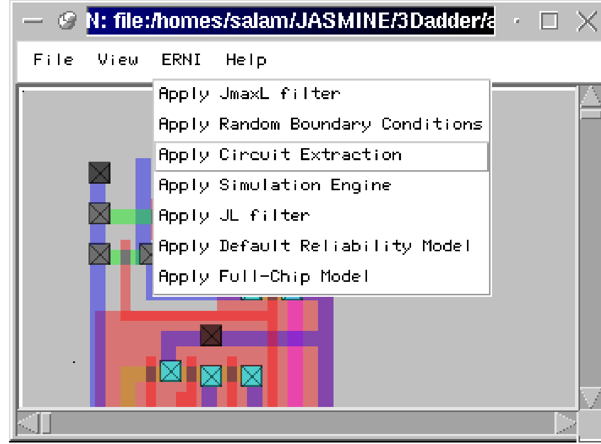

Figure 1-11: *A screen-shot of MAJIC with a view of menu items for reliability analyses in ERNI.*

RCAD tool, ERNI-3D, for the reliability analysis of 3D circuits [27]. Using ERNI-3D, circuit designers can get interactive feedback on the reliability of their circuits associated with electromigration, 3D bonding, and joule heating. ERNI-3D parses 3D circuit layouts from 3D-Magic and extracts both conventional and 3D interconnect trees from the layouts. It employs the hierarchical reliability analysis approach used in ERNI and applies a simplistic joint probability distribution method to report a single reliability figure for the whole chip. The initial version of ERNI-3D treats 3D circuits with two wafers or device-interconnect layers in the stack. Both ERNI and ERNI-3D are capable of reliability analysis of only Al-based interconnect technology.

1.6 Thesis Statement

The goal of this thesis is to develop new methodologies for interconnect reliability analysis and a reliability CAD tool, SysRel (System-level IC Reliability), for circuit-level electromigration analysis with Cu as well as Al metallization technologies in conventional and 3D circuits. Much insight has been gained through electromigration experiments with Cu dual-damascene technology to identify its distinctive behaviors and a set of circuit-level reliability rules [10]. SysRel is capable of reliability assessment and comparison of Cu and Al metallizations in a circuit layout. As experimental

reliability work and lifetime models only deal with individual reliability units, a set of full-chip reliability matrices, based on a joint probability distribution of individual units, has been proposed and implemented in SysRel. Such a system-level view of reliability analysis allows users to identify electromigration critical nets in a circuit layout and quantify their impact on full-chip reliability.

While a comprehensive reliability analysis tool, such as SysRel, is desirable, it is equally important to integrate the tool into existing IC design flows. Large ICs are designed using a cell/module-based hierarchy. The concept of cell/module level reliability characterization has been introduced in SysRel which, in addition to allowing easy integration into existing design flows, significantly reduces the computational load during reliability assessment of a large layout with numerous reliability units. In other design scenarios where layout is fully custom and not necessarily cell-based, SysRel and its methodologies are still applicable.

Reliability is a strong function of chip temperature. Therefore, a cell-based substrate thermal profiling method is developed and implemented in SysRel to estimate the non-uniform layout-level temperature due to cell power dissipations. Using the non-uniform substrate temperature as a boundary condition, interconnect joule heating is taken into account while calculating the lifetimes of mortal reliability units. In addition to a technology-generic feature, such as thermal analysis, SysRel has the capability of non-blocking via analysis specific to Cu metallization technology. Users can investigate the impact of non-blocking vias on full-chip reliability with both stochastic and deterministic assignment of non-blocking vias in a layout. This novel analysis has provided valuable insight into electromigration test structure design in Cu dual-damascene technology.

1.7 Organization of Thesis

The thesis consists of ten chapters and four appendices. Chapter 2 describes the hierarchical electromigration reliability analysis flow with fundamental reliability units in the context of Cu and Al metallization technologies. After extracting fundamental

reliability units from a circuit layout, (jL) product filtering algorithms with specific properties for Cu and Al metallizations are applied to identify immortal trees. A default model is applied to compute the lifetimes of individual mortal units. Chapter 3 describes the concepts in reliability mathematics that are applied to derive the proposed full-chip reliability metrics using the lifetimes of individual failure units. The reliability mathematics coupled with the full-chip metrics give rise to the concept of reliability budget in circuit layout.

In Chapter 4, the electromigration behaviors of Cu and Al metallization technologies are compared. Using the current density and temperature projections from the ITRS, relative reliability of the two metallization technologies is presented in detail. Chapter 5 introduces the RCAD tool, SysRel, and presents the reliability simulation results from a 2D and 3D 32-bit comparator circuit analysis.

Chapter 6 describes the methodology for cell-based reliability analysis in SysRel. The advantages of cell-based reliability analysis are demonstrated using the simulation results from the 32-bit comparator circuit's hierarchical layout. Chapter 7 introduces the non-blocking analysis in Cu metallization technology, and describes the related capabilities and underlying algorithms in SysRel. Simulation results with the 32-bit comparator layout with varying degrees of non-blocking vias are discussed.

Chapter 8 outlines a layout-level thermal profiling technique implemented in SysRel for estimating the temperature rise in 2D circuits. Device-level thermal simulation work with ANSYS[2], leading to insights for the proposed technique, is also discussed. Chapter 9 describes the design of a significantly large circuit, 64-bit Arithmetic and Logic Unit (ALU). The ALU circuit has been simulated to investigate the reliability issues with various metallization technologies and the impact of temperature on full-chip reliability. Future reliability issues with Cu/low-k interconnect technology are also explored.

Finally, Chapter 10 summarizes the results of the thesis and outlines future research directions in interconnect reliability CAD area. The appendices include relevant Matlab code, information on Java classes in SysRel, and SysRel software release.

[2]ANSYS is a comprehensive engineering simulation tool available from ANSYS Inc.

Chapter 2

Hierarchical Reliability Analysis Flow

In actual integrated circuits, various types of interconnect trees exist and their complexities can be greater than the test structures for electromigration experiments. While it is impossible to fabricate and test all the possible interconnect trees found in an IC, a set of rules and specifications can be developed to make realistic reliability assessments of interconnects during the design and layout process. Based on insights gained from the experimental work with various Cu interconnect trees, a hierarchical reliability analysis approach for circuit-level and layout-specific reliability assessment of any Cu interconnect network has been proposed in [10, 38].

This chapter briefly presents the experimental work for electromigration analysis and summarizes the distinct reliability characteristics in Cu metallization technology. The hierarchical reliability assessment flow is also discussed step-by-step and compared with that of Al metallization technology.

2.1 Electromigration Experiments

Gan *et al.* conducted electromigration experiments with Cu dual-damascene technology [10, 39]. The experimental procedure for Cu electromigration was divided into multiple stages: test structure design, sample fabrication, packaging, testing

and failure analyses. Figure 2-1 shows a sample of test structures: straight via-to-via interconnects ('I'), straight via-to-via lines with an additional via at the center ('dotted-I'), straight via-to-via lines with an additional metal limb at the center ('T'), straight via-to-via lines with two additional metal limbs at the center ('+'), straight via-to-via lines with a transition in width along the interconnect ('width-transition') and wider straight via-to-via lines with extra metal limbs nearer to one side of the lines (asymmetrical 'T' transition). The test lines are either in the first (M1) or second (M2) level of metallization.

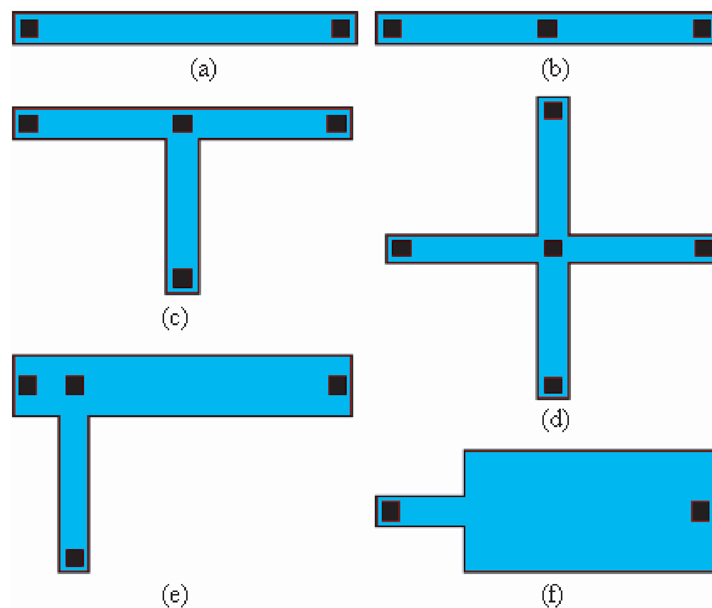

Figure 2-1: *Schematic diagrams of designed interconnect trees for electromigration testing (a)'I' (b) 'dotted-I' (c) 'T' (d) '+' (e) asymmetric 'T' transition (f) width-transition.*

Test samples were fabricated using a Cu dual-damascene process in IME and SEMATECH. Sixteen samples of each test structure were stressed in a package-level electromigration test system with a temperature range between 50^oC and 400^oC. To reduce the possibility of temperature-induced failures and variation in diffusivity due to joule heating, current densities were chosen to limit joule heating below 10^oC through out the experiments. The range of current density in the test structures was between 0.5 to 5.0 MA/cm^2.

In addition to recording median-time-to-failure (t_{50}) (defined as time to 30% increase in resistance), resistance trends were observed for different structures. Various failure analysis tools, such as Optical Microscopy, SEM, FIB and TEM, were used to characterize the type of failure modes. In the experiments, all failures detected were due to open-circuit failure from formation of voids in a line or via. Experimental results and failure characteristics are presented in detail in [39, 40, 41, 42].

2.2 Via Asymmetric Failure Characteristics in Copper Technology

In Cu interconnects, it has been widely reported that the Cu/Si_3N_4 interface acts both as the dominant diffusion pathway for atoms and as the likely site for void nucleation. Due to this fact, the lifetimes of M1 type interconnects are different from that of M2 type interconnects [40]. Gan *et al.* experimentally demonstrated that the lifetimes of the M2 test structures were always higher than those of the M1 test structures, provided that both types of interconnects had the same length, width, and number of vias at each end. The underlying phenomenon applies to all metal layers and can be generalized using via-above and via-below definitions. The via-above test structures refer to interconnect segments which lie in the first level of metallization, and the vias are located above the test lines. Conversely, the via-below test structures are in the second metallization level while the vias are located below the test lines (see figure 2-2).

During electromigration in Cu interconnects, a tensile stress develops at the cathode end, where the Ta liner underneath a via forms a blocking boundary to the diffusing Cu atoms. For current Cu technology, the critical tensile stress for void nucleation has been estimated to be 41 MPa or less [43]. In via-below structures, voids preferentially nucleate at the Cu/Si_3N_4 interface due to the low critical stress required for that interface. An open-circuit failure would occur only when the void grows to span the whole thickness of the metal line, resulting in a very large void

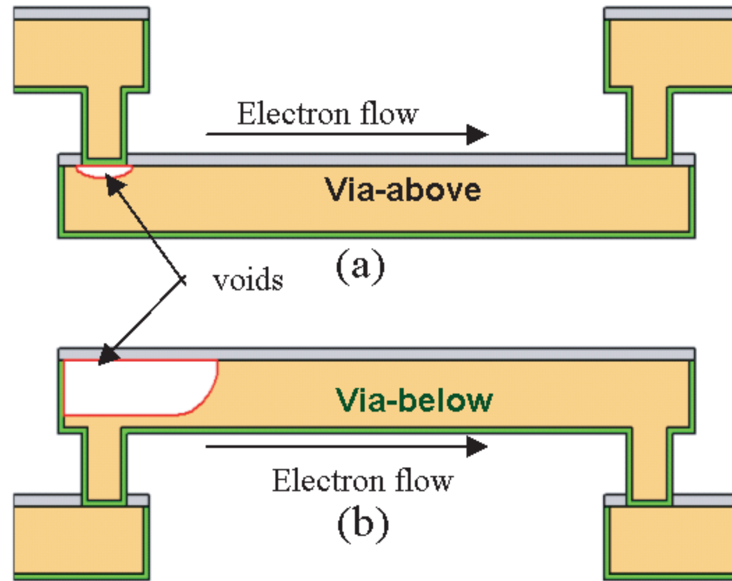

Figure 2-2: *Side-view schematic of void formation in via-above and via-below interconnects. (a) A small-volume fatal void in via-above interconnects; (b) A large-volume partially-spanning non-fatal void in via-below interconnects.*

volume as shown in figure 2-2(b). On the other hand, in via-above structures the maximum tensile stress develops at the Cu/Si_3N_4 interface near the cathode via. Therefore, an open-circuit failure would occur if a small-volume void forms below the via, such that the pathway for electron flow is blocked (see Figure 2-2(a)).

This asymmetry in the void volume required for failure not only accounts for the asymmetry in lifetime but also contributes to the different $(jL)_{crit}$ products in immortality conditions for via-above and via-below type of interconnects. The $(jL)_{crit}$ values are directly related to the critical stresses required for interconnect failure. The values are reported by Ho *et al.* to be 3700 A/cm for via-below [44, 45], and by Hau-Riege to be 2100 A/cm for via-above type interconnects [43]. Therefore, in Cu metallization technology, if the electrons are flowing into the line from a via on top, a shorter lifetime and a smaller $(jL)_{crit}$ is expected compared to the condition in which the electrons flow into the line from below.

2.3 Steps in Hierarchical Reliability Analysis for Copper Metallization

The hierarchical reliability analysis is based on extracting interconnect trees, the fundamental reliability units; applying various stages of filtering conditions to identify mortal units; and then applying a default model to estimate lifetimes of the mortal units. The input to this flow is a circuit layout with mask definitions. The steps for hierarchical reliability analysis in Cu metallization technology are as follows.

(i) Extract interconnect trees from a layout: As defined in section 1.3, an interconnect tree is a unit of continuously connected high-conductivity metal lying within one layer of metallization. In addition to geometric properties, the locations of the vias/contacts are also identified in each interconnect tree. Given an interconnect tree in Cu metallization, every via is classified as either "via-above" or "via-below" depending on whether it is located above or below the interconnect line, respectively.

(ii) Determine the longest terminating via-to-via distance, L_{max}: When an interconnect tree is constructed from multiple segments as shown in figure 2-3, multiple terminating vias exist due to vias at the end of each segment. To find the longest distance between two terminating vias in an extracted tree, an interconnected graph needs to be created where vias and via-to-via distances are represented as nodes and edges, respectively. The shortest path between the vias are calculated using the Minimum Spanning Tree algorithm [46]. Then terminating via-to-via distances are just the addition of via-to-via distances along the path. L_{max} is chosen to be the longest terminating via-to-via distance.

(iii) Filter interconnect trees using via-above $(jL)_{crit_nuc}$ failure criterion: The first step of the filtering algorithm assumes the worst case scenario. The maximum current density allowed by the design rule, j_{max}, can be obtained from the International Technology Roadmap for Semiconductors [2]. Using the worst case threshold of $(jL)_{crit_nuc} = 1500 A/cm$ [47] (i.e. the via is above the line at the

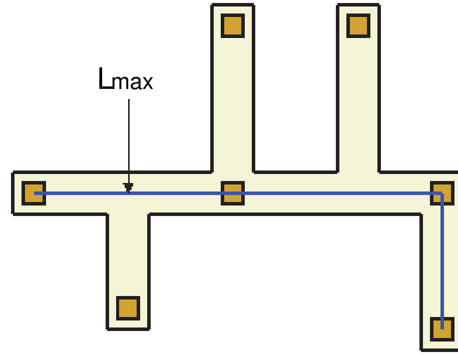

Figure 2-3: *Calculating L_{max} in an interconnect tree.*

cathode end as in a M1 test structure), it is checked whether $(jL)_{crit_nuc}/j_{max} > L_{max}$ for every interconnect tree. If the inequality is true, the tree is considered immortal and is ignored during further analysis.

(iv) Filter mortal trees using via-below $(jL)_{crit_sat}$ failure criterion: Determine if either of the vias contributing to L_{max} is a via-above. If at least one of the vias happens to be via-above, then the interconnect tree may fail with a via-above $(jL)_{crit_nuc}$ failure criterion as determined in the previous step and the hierarchical flow proceeds directly to step (v). However, if both of the vias are via-below, the via-below $(jL)_{crit_sat}$ failure criterion of $(jL)_{crit_sat} = 3700A/cm$ [44, 45] is applied. The condition $(jL)_{crit_sat}/j_{max} > L_{max}$ is checked for such interconnect trees. However, even if a tree passes this test, all the other vias in the tree must be considered before classifying it as immortal. This is because via-above nodes have a much smaller immortality value and thus shorter lengths may fail. The longest distance in the graph from any via-above node to all other vias, L_{max_va}, is determined. We again apply the test $(jL)_{crit_nuc}/j_{max} > L_{max_va}$ to determine whether the tree might fail.

(v) Estimate the current density, j_i, in each segment: The Vdd and Gnd lines in the circuit layout need to be identified as they have unidirectional current flow and are most susceptible to electromigration failure. Most local interconnects transmit signals between devices in the form of bidirectional or alternating cur-

rent (AC). In these cases, an equivalent direct current (DC) which produces the same electromigration damage [48], such as the root-mean-square (RMS) of the AC, is assumed. On the other hand, clock signals usually operate with pulsed DC. Experiments [48] and modeling [49] have shown that the equivalent DC is given by the average of the pulsed DC, I_{avg}, which is given by the expression

$$I_{avg} = rI = \left(\frac{t_{on}}{t_{on} + t_{off}} \right) I \tag{2.1}$$

where I is the current during the "on" time and r is the duty ratio given by t_{on} and t_{off} where t_{on} is the pulse width and the sum of t_{on} and t_{off} equals the clock period. The worst case loading of Vdd and Gnd lines can be identified using power consumption reports from a circuit simulation tool. However, estimating current flow at each segment in local interconnects is complicated and requires detailed circuit-level simulation.

(vi) Filter mortal trees by detailed calculation of steady-state stresses: The concept of steady-state stresses in interconnect trees is an extension of the immortality condition in stud-to-stud straight lines. The maximum stress difference in an interconnect tree, $\Delta\sigma_{max}$, is given by the path that has the highest sum of the (jL) products, summing over the limbs/segments in the path [18]. This is expressed by

$$\Delta\sigma_{max} = \frac{\rho e Z^*}{\Omega} (jL)_{eff} \tag{2.2}$$

where

$$(jL)_{eff} = \max_{\text{all junction pairs } i,\, j} \left(\sum_k j_k L_k \right) \tag{2.3}$$

where $\Delta\sigma_{max}$ is the stress difference between the anode and the cathode, Z^* is the effective charge number, e is the elementary charge, ρ is the electrical resistivity of the metal, Ω is the atomic volume, i and j are two terminating vias, and j_k and L_k are the current density in, and length of, segment k along the path between i and j, respectively. To filter the immune interconnect trees using $(jL)_{eff}$, repeat steps (iii) and (iv) by replacing $(jL)_{max}$ with $(jL)_{eff}$.

(vii) Analyze mortal trees with electromigration failure models: A conservative default electromigration model based on the analysis of individual nodes (vias or contacts) in a tree is used to estimate the lifetime of a mortal tree. The output of the model is taken to be the median-time-to-failure of a tree. The default model is discussed in a later section. Furthermore, numerical methods can be applied for detailed calculation of the time-dependent stress analysis. Software programs, such as MIT/EmSim [9, 50] and CuEmSim [51] , allow the calculation of stress evolution in multi-terminal interconnect trees.

(viii) Apply full chip stochastic reliability model: The median-time-to-failure (MTTF) or t_{50} of each mortal interconnect tree is estimated using the default model or a micro-structure level electromigration simulation program (e.g. MIT/EmSim). The full-chip reliability model combines the t_{50} of multiple interconnect trees from the same layout and provides a set of reliability metrics for the overall chip. The stochastic model and underlying reliability mathematics are discussed in the next chapter.

2.4 Contrast with Hierarchical Reliability Analysis in Aluminum Metallization

Failure mechanisms in Cu and Al interconnects are significantly different due to their different architectural schemes as described in section 1.2. In Al metallization structures, the Al line has refractory metal layers above and below, and tungsten (W) filled vias are used to connect interconnects from different levels. The W-filled vias act as a perfect and symmetric diffusion barrier irrespective of the electron flow direction. Therefore, lifetime differences were not observed for M1 and M2 type structures in Al metallization.

Via classification is not required in hierarchical analysis flow for Al technology. Steps (iii) and (iv) in the hierarchical reliability analysis are merged into one where a single (jL) product threshold, $(jL)_{sat} = 4000A/cm$, is applied to check for immortal-

ity [11]. The reported (jL) product threshold for Al is also slightly higher than that of Cu via-below type interconnects. This is because Al interconnect architecture allows for top and bottom conducting shunt layers to divert electron flow around voids for both M1 and M2 type of geometry.

2.5 Electromigration Default Model

While microstructure level simulation of electromigration in complex interconnect trees is possible, less computationally-intensive analytical models are needed for circuit-level reliability assessment.

A conservative model for the analysis of interconnect trees has been previously proposed by Hau-Riege and Thompson [18] and verified for Al-based test structures. Such a model is extended, by Gan, for Cu interconnect analysis [41]. The stress evolution at a node ($x = 0$) in an interconnect tree has been conservatively estimated as

$$\sigma(t) = \sqrt{\frac{4t}{\pi}}\frac{\rho e Z^*}{\Omega}\sqrt{\frac{B\Omega}{kT}}\frac{\sum_i D_i j_i}{\sum_i \sqrt{D_i}} + \sigma_o \tag{2.4}$$

where B is the bulk modulus of the material surrounding the migrating material, σ_o is the initial hydrostatic stress in the tree, and D_i and j_i are the atomic diffusivity and current density in segment i, respectively. Assuming a constant and time-independent atomic diffusivity along the segment and initial hydrostatic stress, $\sigma_o = 0$, we can estimate the time, t_{nucl}, when the tensile stress is equal to the critical stress for void nucleation, σ_{nucl}, at the node.

$$t_{nucl} = \left(\frac{\sigma_{nucl}\Omega}{\rho e Z^*}\sqrt{\frac{\pi}{4}}\sqrt{\frac{kT}{B\Omega}}\frac{\sum_i \sqrt{D_i}}{\sum_i D_i j_i}\right)^2 \tag{2.5}$$

The above equation can be used to calculate time to extrusion, t_{extru}, using the critical stress for extrusion, σ_{extru}, at the node.

$$t_{extru} = \left(\frac{\sigma_{extru}\Omega}{\rho e Z^*}\sqrt{\frac{\pi}{4}}\sqrt{\frac{kT}{B\Omega}}\frac{\sum_i \sqrt{D_i}}{\sum_i D_i j_i}\right)^2 \tag{2.6}$$

At each via, the stress evolution of the whole subtree connected to the node is considered. Each subtree is then replaced with a semi-infinite limb with an effective diffusivity and current density. Once a void nucleates, it starts to grow and leads to a resistance increase in one of the limbs. Assuming that the void spans the whole width and thickness of the interconnect, the void length is written as a function of time, using which, time to void growth, t_{grow}, to length L_v can be expressed as,

$$t_{grow} = \frac{L_v kT}{\rho e Z^*} \frac{1}{\sum_i D_i j_i} \tag{2.7}$$

Equations 2.5, 2.6, and 2.7 are described in detail in [41]. All the vias in an interconnect tree are evaluated individually using the equations. The methodologies for estimating the lifetime of an interconnect tree, using the vias' t_{nucl}, t_{grow}, or t_{extru}, for Cu and Al technologies are as follows.

Derivation of tree's time-to-failure (TTF): Cu technology

1. For each via in a tree
2. if (via-above)
3. TTF $= t_{nucl}$
4. if (via-below)
5. TTF = min($t_{nucl} + t_{grow}$, t_{extru})
6. Tree's TTF = min(all via TTFs)

Derivation of tree's time-to-failure (TTF): Al technology

1. For each via in a tree
2. TTF = min($t_{nucl} + t_{grow}$, t_{extru})
3. Tree's TTF = min(all via TTFs)

It is important to note that extrusion has not been observed to be the primary cause of failure for both circuit operating and accelerated testing conditions ($j < 5MA/cm^2$). Therefore, $(t_{nucl} + t_{grow})$ is calculated as the TTF of via-below type nodes in Cu technology and for all nodes in Al technology.

2.6 Summary

We have described a new hierarchical approach for predicting the reliability of Cu-based interconnects in circuit layouts. Based on the differences in electromigration failure mechanisms from Al technology observed in experimental work, a (jL) product filtering algorithm with a classification of separate via-above and via-below treatments are applied to Cu interconnects. Such a via classification is not required in Al metallization technology. After the filtering of immortal trees, a default model is applied to the remaining trees to compute a reliability figure, in terms of time-to-failure, for individual units.

Chapter 3

Electromigration Reliability Mathematics

Time-to-failure of an interconnect tree needs to be modelled as a stochastic variable. The operating stress conditions leading to failure and degradation of physical environment, such as grain structure and texture, are not deterministic. In this chapter basic probability terms essential for understanding stochastic reliability analysis are first defined. Electromigration failure times are found in many studies to approximate a lognormal distribution. While models and experiments aid in predicting the lifetime of a single unit, it is more important to quantify the impact on full-chip reliability as a large number of failure units can exist in a circuit. Therefore, probability theories in series and parallel systems are reviewed and brought into the perspective of IC reliability analysis. Finally, a set of reliability metrics, based on the desired lifetime of the chip, has been proposed and demonstrated for full-chip reliability analysis.

3.1 Basic Definitions

There are four basic probability functions that are the fundamental building blocks of reliability mathematics. These are cumulative distribution function, reliability function, probability density function, and failure rate.

If time-to-failure, TTF, of a failure unit is a continuous time random variable, the

probability that the unit will fail within time t is called the *Cumulative Distribution Function* (CDF) and denoted by $F(t)$. Mathematically,

$$F(t) = \mathcal{P}(TTF \leq t) \tag{3.1}$$

Conversely, the probability that the unit will survive beyond time t is called the *Reliability Function* and denoted by $R(t)$. Mathematically,

$$R(t) = 1 - F(t) \tag{3.2}$$

It is important to understand the physical meaning of the CDF from the perspective of reliability experiments. Suppose that N identical components are put into operation at the same instant ($t = 0$) and operate for a time t. If the number of components failing is $N_f(t)$ at time t, the fraction of population failing is $N_f(t)/N$. This ratio is regarded as an attribution of a single such component by defining it as the CDF, i.e.

$$F(t) = \frac{N_f(t)}{N} \tag{3.3}$$

The above statement can be made with the number of surviving components beyond a time t to define the Reliability Function as equation 3.2. Both $R(t)$ and $F(t)$ are dimensionless quantities between 0 and 1.

For a continuous random variable the *Probability Distribution Function* (PDF) is defined as

$$f(t) = \frac{dF(t)}{dt} \tag{3.4}$$

A TTF distribution model can be any probability density function defined over the range of time from $t = 0$ to $t = \infty$. According to equation 3.4, $f(t)dt$ is the probability that a component will fail between time t and $t + dt$; and CDF $F(t)$ can be derived by integrating $f(t)$ until a time t. Mathematically,

$$F(t) = \int_0^t f(x)dx \tag{3.5}$$

and therefore,

$$\int_0^\infty f(t)dt = 1 \tag{3.6}$$

In light of the above discussion, the CDF $F(t)$ can be described in the following three ways.

Definition 1 $F(t) =$ the area under the PDF $f(t)$ from 0 to t.

Definition 2 $F(t) =$ the probability that a single randomly chosen new unit will fail by time t.

Definition 3 $F(t) =$ the proportion of the entire population that fails by time t.

The latter two definitions are particularly useful in interpreting the joint CDF in the presence of multiple failure units as we will see in a later section.

While Cumulative Distribution Function, Probability Distribution Function, and Reliability Function are the fundamental building blocks of reliability mathematics, most literature makes the greatest use of the Failure Rate. The *Failure Rate* is defined for non repairable populations as the (instantaneous) rate of failure for the survivors to time t during the next instant of time [52]. It is a rate per unit of time similar in meaning to reading a car speedometer at a particular instant and seeing 45 mph. The next instant the Failure Rate may change and the units that have already failed play no further role since only the survivors count.

The Failure Rate is denoted by $\lambda(t)$ and calculated as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)} \tag{3.7}$$

As seen from the expression, it is the conditional probability that a t hour-old unit will fail in an additional time dt. $\lambda(t)$ has a unit of $[\text{time}]^{-1}$. For a well-designed system, Failure Rate, in terms of number of units per hour, can be very small. As a result, special units are commonly used. In the IC industry, Failure Rates are most commonly expressed in FITs, which is an abbreviation for *Failure unIT* [53]. By definition, FIT is parts per million per thousand hours, also referred to as PPM/K.

Therefore,

$$1 \text{ FIT} = 1 \text{ in } 1,000,000 \text{ failures per } 1000 \text{ hours}$$
$$1 \text{ FIT} = 1 \text{ in } 1,000,000,000 \text{ failures per hour}$$

When $\lambda(t)$ is expressed in failures per hour, the Failure Rate in terms of FIT can be obtained by

$$FIT \equiv \lambda(t) \times 10^9 \tag{3.8}$$

The typical Failure Rate of an integrated circuit generally varies as a function of time in the manner illustrated in figure 3-1 [52, 53, 54]. During the early life of an

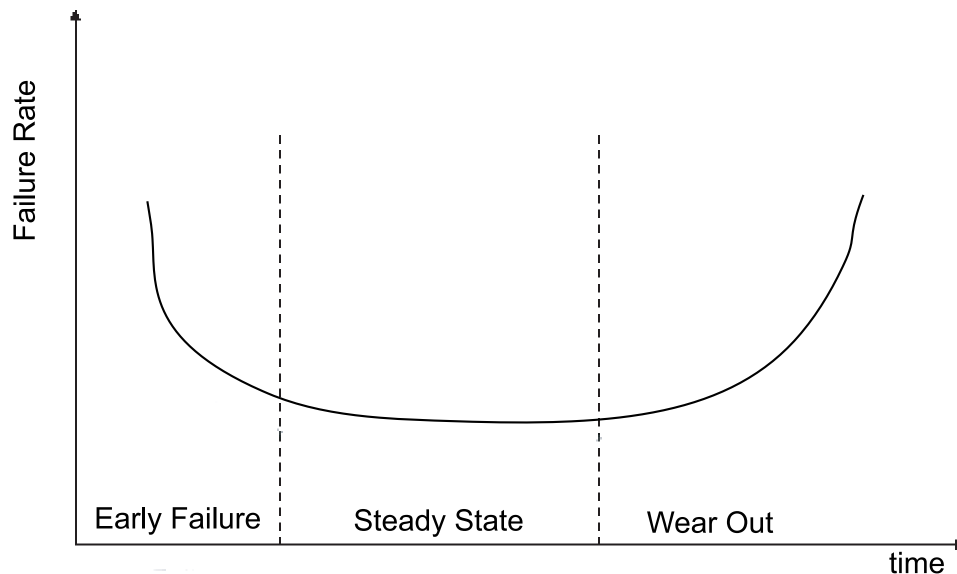

Figure 3-1: *Failure rate versus time for typical Integrated Circuits (bathtub curve).*

IC the failure rate is high and decreasing over time. This period is often referred to as "early failure" or "infant mortality". The cause of early failures are generally manufacturing defects. Rigorous tests (burn-in tests) are performed on ICs right after manufacturing to weed-out early failures and produce high quality chips. During the steady-state period, the Failure Rate is fairly low and constant over time as a result of large number of unrelated causes. Eventually ICs enter the wearout period where

the Failure Rate is monotonically increasing. The electromigration induced failure of interconnects is an example of a wearout process.

3.2 Lifetime Distribution Model

Several probability distribution functions have been successfully used to characterize lifetimes arising from a wide range of failure mechanisms. Following are some widely used lifetime distribution models.

- Exponential
- Weibull
- Extreme Value
- Lognormal

The lifetime distribution model can be chosen based on the physics of failure mechanism and empirical success in fitting actual failure data. However, choosing a model by fit alone is not sufficient as many distributions are flexible enough to fit a wide range of failure analysis data.

Electromigration failure times are found in many studies to approximate a lognormal distribution. Moreover, lognormal distribution is appropriate for a failure mechanism with the multiplicative degradation property. If at any instant in time a degradation process undergoes a small increase in the total amount of degradation that is proportional to the current total amount of degradation, then it is reasonable to expect the time to failure (i.e. reaching a critical amount of degradation) to follow a lognormal distribution [55].

Let x_1, x_2, $\cdots x_n$ be measurements of the amount of degradation for a particular failure process taken at successive discrete instants of time as the process moves towards failure. Assume the following relationships exist between the xs:

$$x_i = (1 + \epsilon_i)x_{i-1} \tag{3.9}$$

where ϵ_i are small, independent random perturbations to the system that move the failure process along. Equation 3.9 is referred to as the multiplicative degradation in a failure mechanism. The total amount of degradation at the *n-th* instant of time can be expressed by

$$x_n = \left(\prod_{i=1}^{n} (1 + \epsilon_i) \right) x_0 \tag{3.10}$$

where x_0 is initial state and the ϵ_i are small random perturbations. Taking the natural logarithm of both sides,

$$\ln x_n = \sum_{i=1}^{n} \ln(1 + \epsilon_i) + \ln x_0 = \sum_{i=1}^{n} \epsilon_i + \ln x_0 \tag{3.11}$$

Using the Central Limit Theorem argument we can conclude that $\ln x_n$ has a normal distribution, and therefore, x_n will follow a lognormal distribution for any n or at any t in continuous time [52]. Diffusion or migration of atoms are expected to follow the multiplicative degradation property which in turn justifies the wide use of lognormal distribution for electromigration.

3.2.1 Lognormal Distribution

When the natural logarithm of a random variable X is normally distributed with mean μ and variance σ^2, X is said to have a lognormal distribution. The relationship between normal and lognormal distribution allows us to analyze lognormally distributed data with methods for normal distributions. A lognormal random variable X can be transformed to a standard normal variable Z where $Z = (\ln X - \mu)/\sigma$. The Cumulative Distribution Function of the lognormal distribution, $F(t)$, is then

$$F(t) = \Phi \left(\frac{\ln t - \mu}{\sigma} \right) \tag{3.12}$$

where Φ is the Cumulative Distribution Function of the standard normal distribution defined by

$$\Phi(t) = \frac{1}{2\pi} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du \tag{3.13}$$

The Probability Density Function, $f(t)$, of the lognormal distribution is given by

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left(\ln\frac{t}{t_{50}}\right)^2} \tag{3.14}$$

and

$$t_{50} = e^{\mu} \tag{3.15}$$

where t_{50} is the median-time-to-failure (MTTF). Therefore, two parameters are required to fully define a lifetime model with lognormal distribution: the t_{50} or median-time-to-failure (MTTF) and σ standard deviation (STD).

3.3 Reliability Mathematics in Series and Parallel Systems

Multiple failure units exist in a circuit layout each with its own lifetime distribution. To derive a single lifetime distribution applicable to the whole layout, system level reliability mathematics is required.

A series system is shown in figure 3-2(a). Each block represents an individual failure unit. The following three assumptions are needed to apply a series system reliability model.

1. Each component operates or fails independently of every other one, at least until the first component failure occurs.

2. The system fails when the first component failure occurs.

3. Each of the n (possibly different) components in the system has a known lifetime distribution model $F_i(t)$.

Assumption no. 2 above suggests the time-to-failure of the system, t_f, is the minimum value of the individual times-to-failure, t_i.

$$t_f = \min(t_1, t_2, t_3, \cdots) \tag{3.16}$$

When the series model assumption holds, system Reliability Function, $R_s(t)$, Cumulative Distribution Function, $F_s(t)$, and Failure Rate, $\lambda_s(t)$ are computed as follows.

$$R_s(t) = \prod_{i=1}^{n} R_i(t) \tag{3.17}$$

$$F_s(t) = 1 - \prod_{i=1}^{n}(1 - F_i(t)) \tag{3.18}$$

$$\lambda_s(t) = \sum_{i=1}^{n} \lambda_i(t) \tag{3.19}$$

The subscript i refers to the *i-th* component in a series system with total n components. If $F_s(t)$ is the same type of function as $F_i(t)$, it is said to be self-reproducing. The lognormal distribution is not self-reproducing, so that if $F_i(t)$ are lognormal, then $F_s(t)$ cannot be lognormal, and vice versa [54]. The distribution of the series system with lognormal unit distribution is often called the multi-lognormal distribution [56].

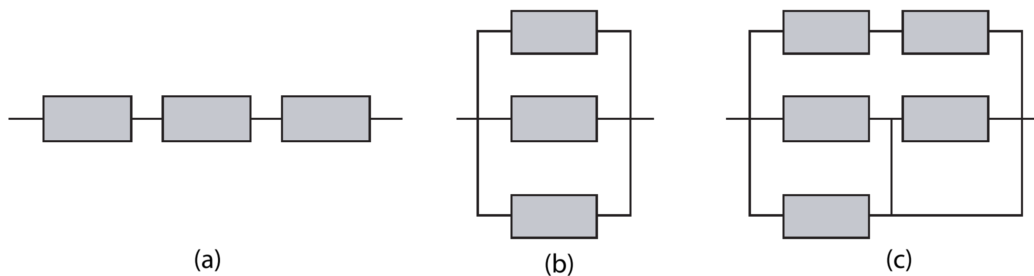

Figure 3-2: *Schematic illustration of various system reliability models; (a) series model; (b) parallel model; (c) complex model.*

In contrast to a series model, for which the first component failure causes the system to fail, is a parallel model for which all the components have to fail before the system fails. If there are n components, any $(n - 1)$ of them may be considered redundant to the remaining one (even if the components are all different). Figure 3-2(b) shows a diagram of a parallel system. The following three assumptions are needed to apply parallel system reliability model.

1. Each component operates or fails independently of every other one.

2. The system operates as long as at least one component is still operating. System

failure occurs at the time of the last component failure.

3. Each of the n (possibly different) components in the system has a known lifetime distribution model $F_i(t)$.

Assumption no. 2 above suggests the time-to-failure of the system, t_f, is the maximum value of the individual times-to-failure, t_i.

$$t_f = \max(t_1, t_2, t_3, \cdots) \tag{3.20}$$

When the parallel model assumptions hold, the system Cumulative Distribution Function, $F_p(t)$ is calculated as follows.

$$F_p(t) = \prod_{i=1}^{n} F_i(t) \tag{3.21}$$

with the subscript i referring to the $i\text{-}th$ component. The system Reliability Function, $R_p(t)$, and Failure Rate, $\lambda_p(t)$, can be evaluated using basic definitions, once we have $F_p(t)$.

Systems encountered in real life are not often as simple as just a series or parallel model. Rather they can be a combination of series and parallel models an example of which is shown in 3-2(c). Such systems are analyzed with a bottom up approach applying series or parallel model on components and sub-components.

3.4 Reliability Metrics for Full-chip Analysis

Electromigration failure dictates the long range lifetime of a chip as the underlying mechanism is a wear out process. While the lifetimes of individual failure units are assumed to have lognormal probability distribution, it is important to derive a set of full-chip stochastic metrics useful to designers. Moreover, different chips would have different reliability goals vis-a-vis their applications. For example, the reliability target for a chip inside a mobile phone, typically used for 3 to 4 years, might be very different than that of an ear implant chip.

We propose a set of full-chip metrics based on the desired lifetime of a chip [38]. Given a target lifetime T_l, reliability metrics useful to designers are: probability of no failure, FIT along lifetime, maximum FIT, and time to cumulative % failure. Taking the time-to-failure estimated using the default electromigration model (described in section 2.5) as median-time-to-failure, t_{50}, and standard deviation, σ, from experimental results, we model the lifetimes of individual interconnect trees using the lognormal distribution. For full chip reliability analysis, it is conservatively assumed that all the failure units are in series, i.e. the chip will fail when any one of the units fails. Lognormal distribution is not self-reproducing for a series system and the proposed metrics do not require any probability distribution assumption on the full-chip lifetime. The derivation procedures for the metrics are described next.

3.4.1 Probability of no failure

The probability of no failure until lifetime T_l is essentially the value of the full-chip Reliability Function $R_f(t)$ at $t = T_l$. As a series system model is assumed, the full-chip Reliability Function can be calculated as

$$R_f(t) = \prod_{i=1}^{n} R_i(t) \tag{3.22}$$

where $R_i(t)$ is the Reliability Function of the *i-th* component in a circuit layout with n mortal interconnect trees. $R_i(t)$ of each component is derived using

$$R_i(t) = 1 - F_i(t) \tag{3.23}$$

where the lognormal Cumulative Distribution Function $F_i(t)$ of *i-th* component is calculated using equation 3.12.

3.4.2 Failure Rate along chip lifetime

According to the series system model, the full-chip Failure Rate, $\lambda_f(t)$, is the summation of each mortal interconnect tree's Failure Rate, $\lambda_i(t)$. The Failure Rate of each

unit is calculated using the definition,

$$\lambda_i(t) = \frac{f_i(t)}{1 - F_i(t)} \tag{3.24}$$

where $f_i(t)$ and $F_i(t)$ are the lognormal Probability Density and Cumulative Distribution Functions, respectively. The full-chip failure rate is converted into FIT and plotted against the desired lifetime of the chip. The maximum FIT along the lifetime is also reported.

3.4.3 Time to cumulative % failure

Unlike the Failure Rate and Reliability Function, time to cumulative % failure does not have a closed form expression particularly for a series system with lognormal probability units. According to the definition no. 3 of the Cumulative Distribution Function, $F(t)$ is the proportion of the entire population that fails by time t. Therefore, once the full-chip Cumulative Distribution Function $F_f(t)$ is derived, time to any cumulative % failure can be read from the plot where the x-axis range of 0 to 1 represents 0% to 100% of the population and y-axis is the time-to-failure.

3.5 Reliability Metrics in Use

Let us assume that a circuit layout has two types of trees, type 1 and type 2, with median-time-to-failures (t_{50}) and standard deviations (σ) for lognormal distribution as defined in table 3.1. Given that the target lifetime, T_l, of the chip is 30 years, probability of no failure $R_i(T_l)$ and maximum FIT of each type of trees are shown as well in the table.

The median-times-to-failure of the two types of trees are significantly larger in comparison to the target lifetime of 30 years. However, the impact of an individual unit becomes obvious only when full-chip reliability is considered. Full-chip probability of no failure within target lifetime, $R_f(T_l)$, is the product of $R_i(T_l)$ (equation 3.22). $R_f(T_l)$ can be low with multiple type 2 trees in a layout. Full-chip FIT is the sum-

mation of individual unit's FIT which is also a large value with few type 1 trees in a layout.

Table 3.1: *Tree types with lognormal lifetime distribution.*

	t_{50} **(years)**	σ	$R_i(T_l)$	max **FIT**
Tree Type 1	145	1.59	0.8391	747.40
Tree Type 2	2000	1.6	0.9957	30.4075

The two techniques for improving full-chip reliability, from a mathematical perspective, are increasing t_{50} and reducing σ. While t_{50} of individual trees can be increased with circuit design and layout techniques, such as reducing current flow and increasing line width, σ is often dictated by process variations and micro-structure of materials. Circuit level reliability enhancement techniques will be further discussed in chapter 5. Assuming that the t_{50} of type 1 tree is improved to 300 years and σ of type 2 tree is reduced to 0.80, the new $R_i(T_l)$ and max FITs are shown in table 3.2. Now let us consider full-chip reliability of a circuit layout with a total of seven mortal

Table 3.2: *Improved tree types with lognormal lifetime distribution.*

	t_{50} **(years)**	σ	$R_i(T_l)$	max **FIT**
Improved Tree Type 1	300	1.59	0.9262	361.224
Improved Tree Type 2	2000	0.8	1.00	0.0020

units with combinations of three type 1, four type 2 trees and three improved type 1, four improved type 2 trees. The full-chip reliability metrics are reported in table 3.3 for each combination. The improvement in full-chip reliability is significant.

The above example demonstrates the significance of the proposed full-chip metrics

Table 3.3: *Full-chip reliability metrics of circuit layout with various types of mortal trees. Here target lifetime of the chip is* 30 *years.*

Mortal interconnect tree combination	**Probability of no failure**	max **FIT**	**time to** 50% **failure**	**time to** 0.2% **failure**
3 Type 1 + 4 Type 2	0.5807	2315.3	37.88 years	0.875 years
3 Improved Type 1 + 4 Improved Type 2	0.7946	1083.7	81.5 years	1.82 years

in identifying the impact of individual unit based on desired lifetime of the chip. Moreover, a designer has the flexibility to choose any particular metric, such as probability of no failure, max FIT, or time to any cumulative % failure, as the figure of merit depending on the chip's application. FIT is more widely used in the ITRS predictions while time to 0.02% failure is the figure of merit for selected applications [57].

3.6 Summary

This chapter gives an introduction to probability mathematics used in reliability analysis and describes the lognormal probability distribution widely accepted for an electromigration lifetime model. While models and experiments aid in predicting the lifetime of a single unit, it is more important to quantify the impact on full-chip reliability as a large number of failure units can exist in a circuit. Therefore, probability theories in series and parallel systems are reviewed. To understand the impact of an individual failure unit on full-chip reliability, it is important to have a set of reliability metrics. Electromigration failure essentially dictates the long range life-time of a chip. We propose to conduct full-chip reliability analysis based on a target lifetime, T_l. Given a target lifetime T_l, reliability metrics useful to designers are: probability of survival, FIT along lifetime, maximum FIT, and time to cumulative % failed. These metrics are calculated using the series system model. Lognormal distribution is not self-reproducing for a series system and the proposed metrics do not require any assumption on the system's failure distribution. The proposed metrics successfully identify the impact of individual unit on full-chip reliability. Moreover, a designer has the flexibility to choose any particular metric as the figure of merit depending on the chip's application.

Chapter 4

Electromigration Reliability Comparison of Aluminum and Copper Interconnects

Under similar test conditions, electromigration reliability of Al and Cu metallization interconnect trees demonstrate significant differences because of the differences in interconnect architectural schemes presented in section 1.2. The low critical stress for void nucleation at the Cu and inter-level diffusion barrier, such as Si_3N_4, interface leads to asymmetric failure characteristics depending on the via position in a line. Unlike Al technology, a (jL) product filtering algorithm with a classification of separate via-above and via-below treatments is required for Cu interconnect trees. In this chapter, electromigration reliabilities of Al and Cu dual-damascene interconnect lines are compared using the best estimates of material parameters and the default model. Differences in atomic diffusivity mechanisms in Al and Cu metallization technologies are also described. Lifetimes of a straight line interconnect are calculated using operating conditions from the ITRS projections for future technology nodes.

4.1 Immortality Condition Filters

The (jL) product thresholds for immortality conditions are already mentioned in earlier chapters. Table 4.1 summarizes reported (jL) products for Al and Cu technologies [44, 45, 47, 11].

Table 4.1: *Reported (jL) product thresholds for immortality conditions in Al and Cu technologies.*

Metallization technology	**Critical (jL) product threshold**	
Al interconnects	$(jL)_{sat} = 4000\ A/cm$	
Cu interconnects	$(jL)_{crit_nuc} = 1500\ A/cm$ (via-above)	$(jL)_{crit_sat} = 3700\ A/cm$ (via-below)

As seen in the table, Al has the highest (jL) product threshold for immortality. The interconnect architecture allows top and bottom conducting shunt layers to divert electron flow around voids in both M1 and M2 type lines. As a result, the tolerable non-fatal void volume in Al M1 and M2 structures is higher than that in Cu interconnects. The asymmetry in void volume for failure in Cu metallization technology accounts for the different (jL) products for the via-above and via-below types of interconnects. As discussed in section 2.2, void volume in via-above type interconnect is smallest for an open circuit or resistance failure, and hence the lowest (jL) product for immortality is observed.

The International Technology Roadmap for Semiconductors projects maximum current density ($j_{\max}$), among other interconnect related parameters, for each technology generation in the future. Using the $j_{\max}$ from the 2003 edition, we computed the immortality length (length below which an interconnect tree is immortal) in various technology generations as illustrated in table 4.2. As technology scaling progresses, the immortality length is consistently decreasing for both Cu and Al technologies. This suggests more and more interconnects trees will be prone to electromigration failure when stressed at their maximum current densities. Al metallization technology has the longest immortality lengths due to the highest (jL) product threshold.

Table 4.2: *Immortality length in various technology generations (ITRS 2003) with Al and Cu metallizations.*

Technology Node	$j_{\max}$ MA/cm²	Immortality Length (um)		
		Cu via-above	**Cu via-below**	**Al**
90nm	0.5	30.0	74.0	80.0
65nm	1.0	15.0	37.0	40.0
45nm	3.0	5.0	12.3	13.3
32nm	4.3	3.49	8.60	9.30
22nm	5.8	2.59	6.38	6.90

4.2 Diffusivity Mechanisms and Models

The electromigration-induced diffusion of metallic atoms in an interconnect can occur through different pathways. Due to the differences in architecture schemes and material properties, diffusion pathways and values are significantly different for *Al* and *Cu* metallization technologies.

4.2.1 Diffusion in Aluminum Interconnects

The dominant diffusion paths in *Al* interconnects can be represented by equation 4.1 following the convention presented by Hu *et al.* [58]

$$(DZ^*)_{eff} = D_S Z_S^* \delta_S \left(\frac{2}{h}\right) + D_{GB} Z_{GB}^* \frac{\delta_{GB}}{d}\left(1 - \frac{d}{w}\right) \tag{4.1}$$

Here, D is the diffusivity, which bears an Arrhenius or exponential form, Z^* is the effective charge, δ is the diffusion interface width, and d, h, and w, are the grain size, line thickness, and line width, respectively. Subscript S denotes the Al/refractory metal liner interfaces at the top and bottom of a line. Subscript GB represents grain boundary inside a line. The other diffusion pathways, grain bulk and side wall interfaces, are assumed to have negligible contributions.

When line width of an *Al* interconnect is larger than the *Al* grain size, for example $0.5um$, multiple grain boundaries exist in the line as illustrated with a cross-section view in figure 4-1(a). Then the dominant diffusion pathways come from the grain boundaries. The microstructure in a wide line, where one of the grain boundary

planes lie in the direction of the current flow, is referred to as polygranular. Such lines are often termed Al polygranular type. The diffusivity, D_{poly}, in Al polygranular type lines can be expressed as

$$D_{poly} = D_{o,GB} \times e^{-\frac{Ea_{GB}}{kT}} \times \frac{\delta_{GB}}{d}\left(1 - \frac{d}{w}\right) \tag{4.2}$$

where $D_{o,GB}$ is the diffusion coefficient of polygranular type lines, Ea_{GB} is the activation energy, k is the Boltzmann's constant, and T is the temperature.

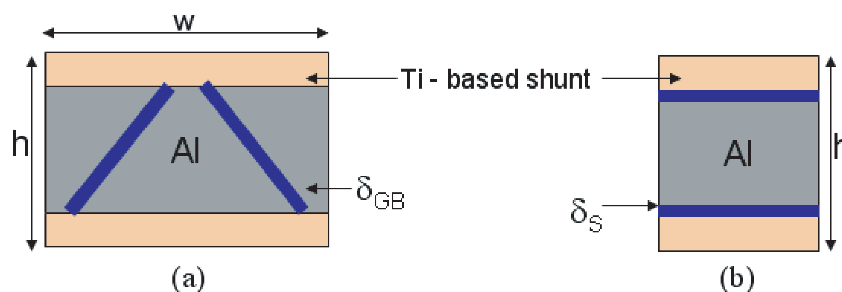

Figure 4-1: *Cross-section of an Al interconnect with schematic illustration of diffusion pathways; (a) wide lines are referred to as polygranular type lines where dominant diffusion paths are grain boundaries (δ_{GB}); (b) narrow lines are referred to bamboo type lines where dominant diffusion pathways are top and bottom interfaces (δ_S).*

Table 4.3 lists the parameter values used in equation 4.2 for calculating diffusivity in Al polygranular type lines [59].

Table 4.3: *Material parameters used for calculating diffusivity of Al polygranular type lines.*

Name	**Symbol**	**Value**
Diffusion coefficient	$D_{o,GB}$	$1.9 \times 10^{-5}\ m^2/s$
Activation energy	Ea_{GB}	$0.8\ eV$
Grain boundary thickness	δ_{GB}	$0.5\ nm$
Grain size	d	$0.5\ um$

As the line width decreases, the microstructure of the line changes from polygranular to bamboo type where grain boundary planes are normal to the direction of current flow. In bamboo type Al interconnects, the dominating diffusion pathways are the top and bottom interfaces between Al and Ti-based shunt layers as illustrated

in figure 4-1(b). The diffusivity, D_{bam}, in Al bamboo type lines can be expressed as

$$D_{bam} = D_{o,S} \times e^{-\frac{Ea_S}{kT}} \times \frac{2\delta_S}{h} \tag{4.3}$$

where $D_{o,S}$ is the diffusion coefficient of bamboo type lines, Ea_S is the activation energy, k is the Boltzmann's constant, T is the temperature, δ_S is the interface thickness, and h is the line height. Diffusivity experiments were carried out by Srikar *et al.* on 0.4 um thick lines, i.e. $h_{ref} = h = 0.4um$ [60]. Reported material parameters are listed in table 4.4.

Table 4.4: *Material parameters used for calculating diffusivity of Al bamboo type lines.*

Name	**Symbol**	**Value**
Diffusion coefficient	$D_{o,S}$	$1.49 \times 10^{-4}\ m^2/s$
Activation energy	Ea_S	$0.9\ eV$
Interface thickness-height ratio	$\frac{2\delta_S}{h_{ref}}$	4×10^{-5}

4.2.2 Diffusion in Copper Interconnects

In Cu interconnects, the Cu/Si_3N_4 interface is the dominant diffusion pathway [40] as illustrated in figure 4-2.

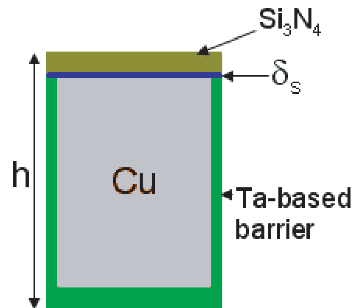

Figure 4-2: *Cross-section of a Cu interconnect with schematic illustration of diffusion pathway (δ_S).*

The Arrhenius model to calculate diffusivity in Cu is

$$D = D_{o,S} \times e^{-\frac{Ea}{kT}} \times \frac{\delta_S}{h} \tag{4.4}$$

It is important to note that atomic diffusivity has been modelled as independent of Cu line width similar to that of Al bamboo type lines. In Cu interconnects, Cu diffusion along the Cu/Si_3N_4 interface dominates even for polygranular microstructure [42].

The kinetic parameters associated with the effective diffusivity has been determined by Wei *et al.* through conventional package-level electromigration stress experiments using fully-processed dual-damascene Cu interconnects (both via-above and via-below structures) under constant current densities [61]. A significant fraction of the test population exhibited steady resistance increase over time prior to failure. This gradual resistance increase results from void growth and the rate of resistance increase has been correlated to the drift velocity for electromigration [10]. Among the results, the activation energy Ea for Cu electromigration has been determined to be $0.80 \pm 0.06eV$, which is in very close agreement with the value found from lifetime analysis for the same structures. With test structure thickness of $h_{ref} = h = 0.24um$, $D_{o,S}\frac{\delta_S}{h_{ref}}$ of equation 4.4 has been determined to be $1.3229 \times 10^{-9} m^2/s$. The material parameters used for calculating diffusivity in Cu metallization are summarized in table 4.5.

Table 4.5: *Material parameters used for calculating diffusivity in Cu metallization technology.*

Name	**Symbol**	**Value**
Diffusion factor	$D_{o,S}\left(\frac{\delta_S}{h_{ref}}\right)$	$1.3229 \times 10^{-9}\ m^2/s$
Activation energy	Ea	$0.8\ eV$

4.2.3 Diffusivity Comparison

Electromigration is the electron-flow induced atomic diffusion, and therefore, comparing the diffusivity of Al and Cu technologies would give us insight on relative electromigration reliability. Al metallization has two regions for consideration, bam-

boo and polygranular. For line width less than the grain size ($0.5um$), we observe bamboo region's diffusivity to be independent of line width. For line width greater than the grain size, Al polygranular region's diffusivity is a function of line width and is higher than that of bamboo region. Although there is discontinuity in slope at the point of conversion (line width=$0.5um$) from bamboo to polygranular region, this composite model is a continuous step model in which polygranular diffusivity is higher than that of bamboo region. On the other hand, atomic diffusivity in Cu is independent of line width and is smaller than Al polygranular diffusivity. We have compared atomic diffusivity for Al and Cu metallization technologies taking a straight interconnect line in TSMC $0.18um$ technology with height $h = 0.51um$. Figure 4-3 shows diffusivity versus interconnect line width for Al and Cu technologies at the maximum interconnect temperature $T = 105^oC$ as reported in the ITRS.

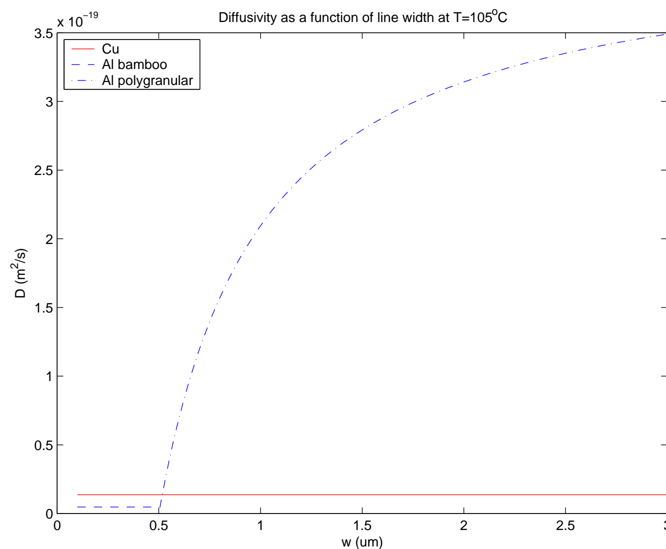

Figure 4-3: *A composite model of atomic diffusivity versus interconnect line width.*

Figure 4-4 shows a plot of diffusivity versus temperature for different metallization technologies. The interconnect line width is assumed to be $1.0um$, twice the grain size, while calculating Al polygranular diffusivity. As seen in the plot, Al polygranular type line has the highest diffusivity. Comparing Al and Cu diffusivities, for a given line width, Cu diffusivity is always about $1/15$ of that for polygranular Al, due to the

similarity in activation energies. Al bamboo diffusivity is lower than that of Cu only for temperatures lower than 302^oC, because the activation energy in Cu diffusivity is lower than that in Al bamboo diffusivity.

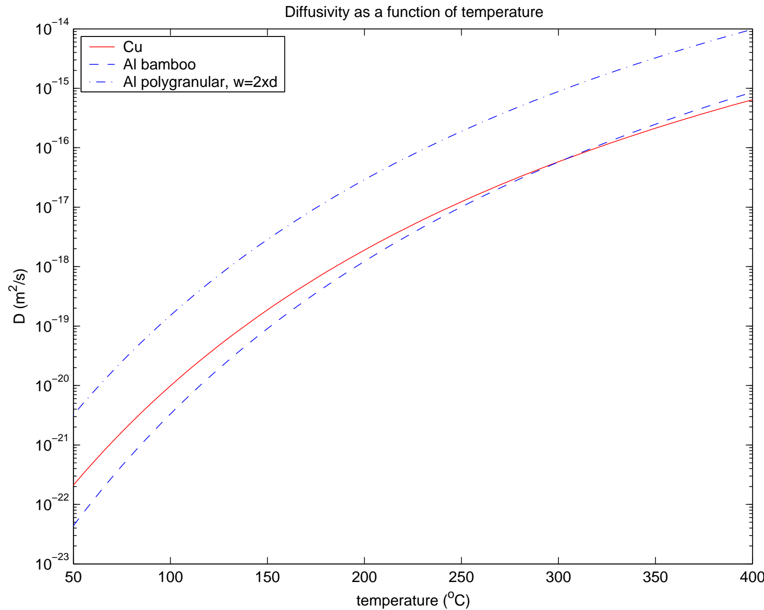

Figure 4-4: *Atomic diffusivity of Al and Cu metallization technologies at different temperatures.*

During the introduction of Cu metallization, Cu interconnects were assumed to be greater electromigration resistive [62]. Lower atomic diffusivity in Cu might have led to such perception. However, figure 4-4 shows that Cu diffusivity is only significantly lower than that of Al polygranular type lines; the difference is not significant when compared to Al bamboo type lines. Therefore, a detailed analysis of lifetimes in different technologies is required to quantify relative electromigration reliability.

4.3 Lifetime Comparison of Straight Line Interconnects

A default model that is applicable to both Cu and Al metallization to predict electromigration lifetime has been discussed in section 2.5. The model estimates the lifetime

of a via considering current densities and diffusivity of connected segments. For a straight line via-to-via interconnect, the equations for time-to-failure due to void nucleation, extrusion, and time to void growth to length L_v at any of the two vias are stated below.

$$t_{nucl} = \left(\frac{\sigma_{nucl}\Omega}{\rho e Z^*} \sqrt{\frac{\pi}{4}} \sqrt{\frac{kT}{B\Omega}} \frac{\sqrt{D}}{Dj} \right)^2 \tag{4.5}$$

$$t_{extru} = \left(\frac{\sigma_{extru}\Omega}{\rho e Z^*} \sqrt{\frac{\pi}{4}} \sqrt{\frac{kT}{B\Omega}} \frac{\sqrt{D}}{Dj} \right)^2 \tag{4.6}$$

$$t_{grow} = \frac{L_v kT}{\rho e Z^*} \frac{1}{Dj} \tag{4.7}$$

where D is the atomic diffusivity in the straight line and j is the current density. Current density j is fixed at 0.5 MA/cm^2 for this analysis. Diffusivity, D, values for Cu, Al polygranular, and Al bamboo types are calculated as described in section 4.2.3. Table 4.6 lists the other parameter values used to estimate the time to void nucleation and time to void growth in Cu and Al metallization technologies.

Table 4.6: *Technology-specific parameter description and values for lifetime calculation.*

Name	**Symbol**	**Cu**	**Al**
Stress for void nucleation	σ_{nucl}	40 MPa	500 MPa
Effective charge number	Z^*	1	4.3
Bulk modulus	B	28 GPa	50 GPa
Atomic volume	Ω	1.18×10^{-29} m^{-3}	1.66×10^{-29} m^{-3}
Electrical resistivity	ρ	1.95 $\Omega\mu - cm$	2.67 $\Omega\mu - cm$
Void length	L_v	0.2 um	0.2 um

Straight line interconnect's lifetimes associated with Cu via-above, Cu via-below, Al polygranular, and Al bamboo types are calculated from t_{nucl}, t_{grow}, and t_{extru} using the algorithms in 2.5. Figure 4-5 plots the lifetimes versus temperature for each type of interconnects.

An Al polygranular type line (using width=1um) has the lowest lifetime. Via-above type Cu interconnect has better lifetime than Al polygranular. Via-below type Cu and Al bamboo type lines compete for the best lifetime depending on the

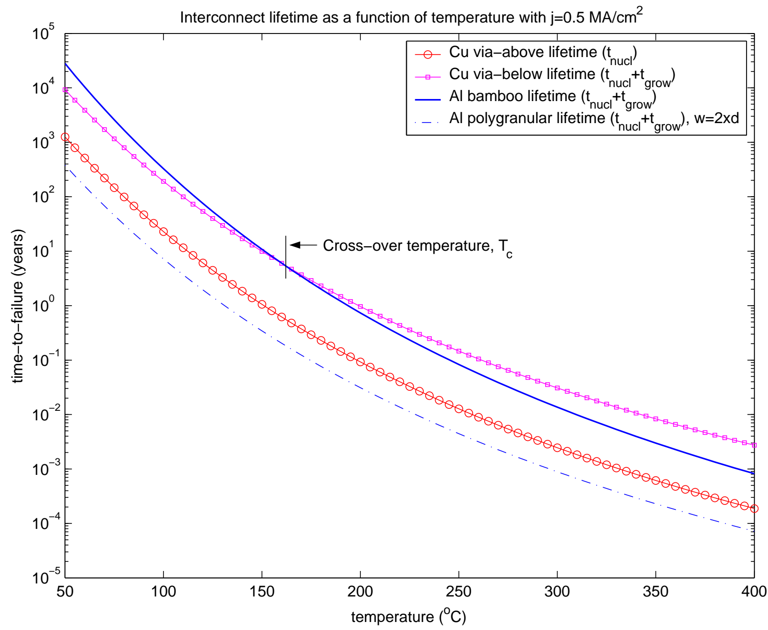

Figure 4-5: *Interconnect lifetimes of various types of lines in Cu and Al metallization technologies.*

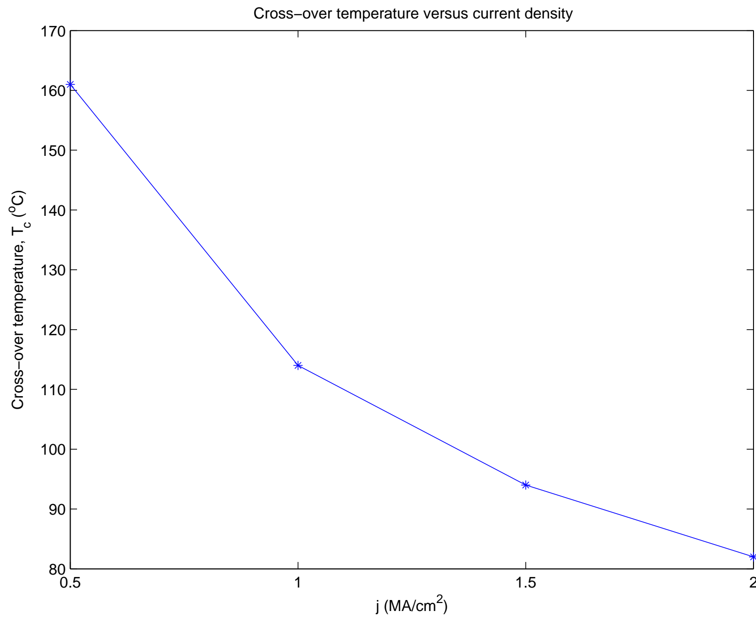

Figure 4-6: *Cross-over temperature, defined as the temperature after which a via-below type Cu line has better lifetime than an Al bamboo line, as a function of current density.*

temperature. An Al bamboo type line has better lifetime than via-below type Cu interconnect for temperatures less than 161^oC. Defining this particular temperature point to be the cross-over temperature, T_c, after which via-below Cu has the best lifetime of all, we observe a decrease in T_c with increasing current density as illustrated in figure 4-6.

We have defined a typical operating condition of interconnects to be current densities below 0.5 MA/cm^2 ($j \leq 0.5MA/cm^2$) and temperature range below 105^oC ($T \leq 105^oC$). Figure 4-7 shows the lifetimes of various types of interconnect lines operating at the above condition. Al bamboo type lines have the best lifetime followed by Cu via-below, Cu via-above, and Al polygranular type lines.

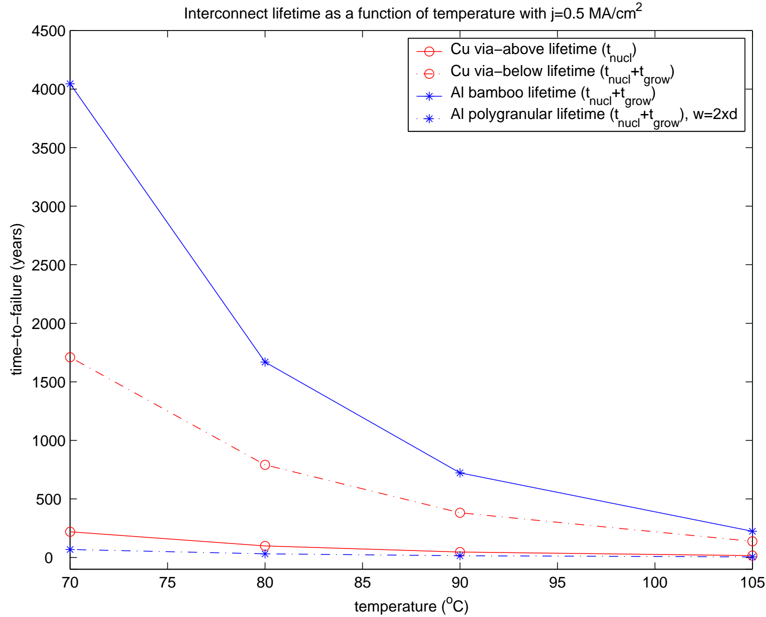

Figure 4-7: *Interconnect lifetimes of various types of lines in Cu and Al metallization technologies in typical operating condition ($j \leq 0.5MA/cm^2$ and $T \leq 105^oC$).*

4.4 Reliability Prediction Using the ITRS

The characteristics of cross-over temperature can be applied to assess interconnect reliability direction in future technology generations. As the maximum current density increases in successive technology generations, according to figure 4-6, decreasing

Table 4.7: *Worst case interconnect lifetimes of various types of Cu and Al lines. Technology generation data is from the ITRS 2003 edition [2].*

Tech.	$j_{\max}$	max	**lifetime of Cu (years)**		**lifetime of Al (years)**	
Node	$\mathbf{MA/cm^2}$	T (^oC)	**via-above**	**via-below**	**bamboo**	**polygran.**
90nm	0.5	105	16.19	137.58	223.43	5.12
65nm	1.0	105	4.05	64.74	70.6	1.62
45nm	3.0	105	0.45	20.68	14.40	0.33
32nm	4.3	105	0.22	14.33	9.08	0.21
22nm	5.8	105	0.12	10.58	6.31	0.14

cross-over temperature indicates better lifetime for Cu via-below type lines. Table 4.7 lists the lifetimes of various types of interconnect lines in Al and Cu metallizations using reported maximum current density and operating temperature values from the ITRS 2003 edition [2]. Al polygranular type or wide Al lines have the worst lifetimes in almost all technology generations. In the 90 and 65 technology nodes, Al bamboo type or narrow Al lines have the best lifetimes. Via-below type Cu line eventually emerges as the longest lifetime interconnect as current density increases in later technology nodes. Table 4.7 also points out that Cu interconnects, particularly with dual-damascene process with Si_3N_4 capping layer, are as susceptible to electromigration failure as Al interconnects considering the small differences in relative lifetimes.

4.5 Summary

The differences in interconnect architecture between Al and Cu metallization technologies lead to different electromigration failure mechanisms. Unlike Al technology, a (jL) product filtering algorithm with a classification of separate via-above and via-below treatments is required for Cu interconnect trees. Different diffusivity mechanisms also lead to the differences in electromigration lifetimes. Al lines wider than the grain size ($d = 0.5um$) have polygranular microstructure with the highest atomic diffusivity. On the other hand, narrow Al lines with bamboo type microstructure have lower diffusivity comparable to that of Cu metallization. Diffusivity in Cu technology

is dominant at the Cu/Si_3N_4 interface and is modelled as independent of line width.

Using the best estimates of material parameters and the default model, electromigration lifetimes of a via-to-via straight line interconnect are compared for four configurations: Cu via-above, Cu via-below, Al bamboo, and Al polygranular. In typical operating condition ($j \leq 0.5MA/cm^2$ and $T \leq 105^oC$), an Al bamboo type line has the longest time-to-failure followed by Cu via-below, Cu via-above, and Al polygranular type line. As the maximum current density in interconnect lines increases in future technology generations according to the ITRS projections, a Cu via-below type line emerges as the longest lifetime interconnect. However, the small difference in relative lifetimes of Cu and Al technologies suggest that Cu metallization is as susceptible to electromigration failure as Al. Moreover, the significant differences in failure characteristics outdates traditional techniques and tools for reliability analysis with Al interconnects. A new sophisticated Reliability CAD tool is essential for circuit level reliability analysis of Cu interconnects.

Chapter 5

SysRel: Circuit-Level Reliability CAD Tool

SysRel is a new reliability CAD tool for electromigration reliability analysis and comparison with Al and Cu metallization technologies in 2D and 3D circuit layouts. SysRel utilizes the interconnect tree based hierarchical reliability analysis. The software design approach in SysRel is detailed in this chapter. The chapter also contains references to source code files wherever appropriate. The earlier sections present general algorithms and data-structures for layout parsing and extracting interconnect trees. Then the graphical user interface and reliability analysis flow in SysRel are discussed. SysRel utilizes a set of joint stochastic reliability metrics based on the desired lifetime of a chip and combines reliability figures from individual fundamental reliability units. Simulation results with a 32-bit comparator circuit layout are discussed to demonstrate the functionality and significance of the tool.

5.1 Software Architecture

SysRel is written in Java for the Java 2 platform. The Java platform consists of three components; Java Language, Java Virtual Machine (JVM), and Java Application Program Interface (API). Java Language defines the semantics and syntax of writing a program. JVM executes Java byte codes produced by compiling a Java program.

The Java API is a set of predefined classes and methods[1] provided with the Java Development Environment. The Java API includes everything from data-structure classes, such as Vector, ArrayList, and Hashtable, to Graphical User Interface (GUI) development classes. SysRel makes extensive use of the APIs from Java Development Kit version 1.3.1 [63]. SysRel consists of 32 classes which can be grouped into the following six categories.

- Main application classes
- Graphical interface classes
- File parser classes
- Corner-stitched data-structure classes
- Layout and tree representation classes
- Core reliability engine

Appendix A defines each category listed above and briefly describes the major classes in them. The code length of SysRel is approximately 9000 lines.

5.2 Graphical User Interface

SysRel is a stand-alone application with its own desktop. The GUI is developed using the Java Swing[2] package. Figure 5-1 shows the graphical interface with a layout opened in the Layout Window. Menu items invoke selected operations in SysRel and reports outcome in the Output Window. Table 5.1 lists the menu bars and corresponding operations through available menu items.

[1]Functions defined inside classes for specific operations, such as construction of the class and data-manipulation.

[2]Swing is the new package in Java 2 with GUI control classes such as window frames, dialog boxes, menu items, check boxes, and buttons.

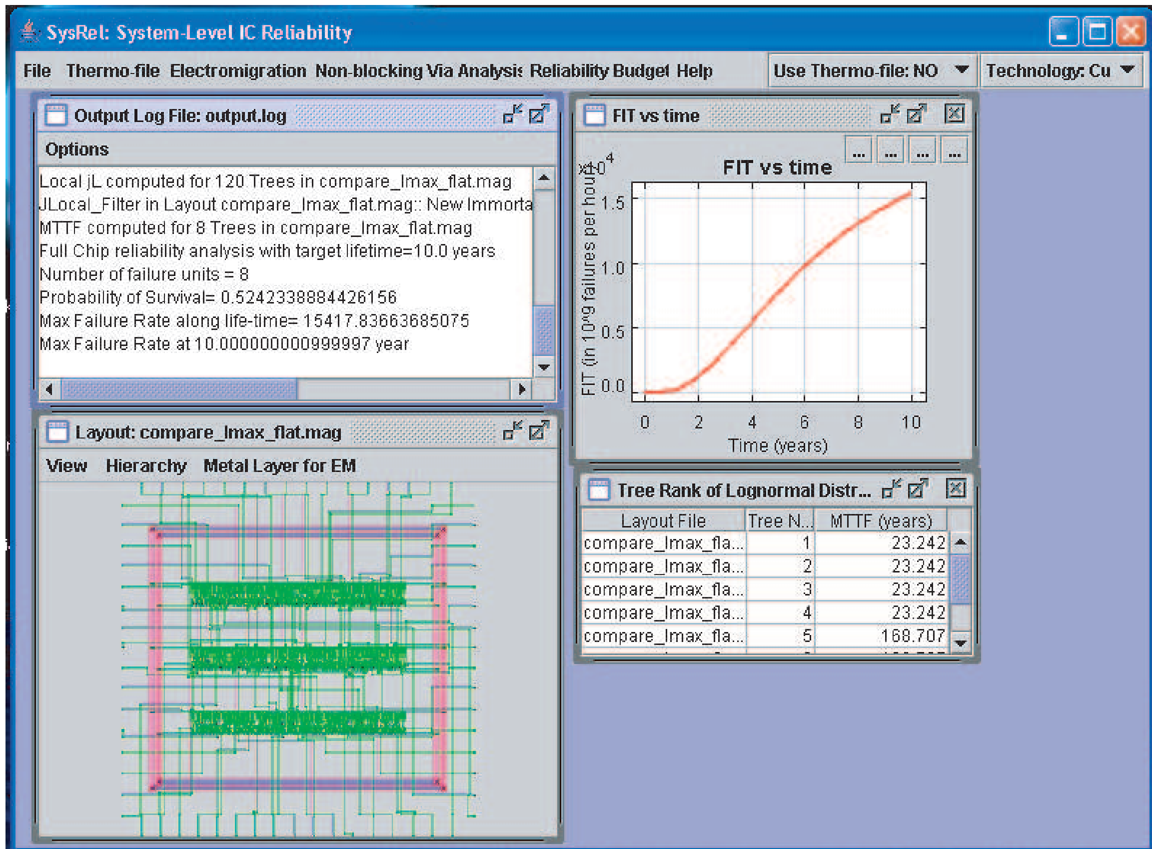

Figure 5-1: *Screen shot of SysRel's desktop pane. Menu items and options are described in table 5.1. At the top right corner are selection tabs to enable thermal analysis and select either Cu or Al metallization. Inside the desktop pane, there are Output Window, Layout Window, Plot Window, and Tree Rank Table.*

Table 5.1: *Menu bars and corresponding operations in SysRel.*

Menu	**Menu Items**
File	Open Magic layout file, Read EM parameters (.emparam) file, Exit
Thermo-file	Read Hierarchical Magic file, Read Power Distribution (.cpower) file, Write Power Density Matrix (currentpden.m) file, Read Temperature Profile Matrix (tprofile.dat) file
Electromigration	List EM parameters, Update selected EM parameters, Extract Failure Units, Apply JmaxL Filter, Compute local jL, Apply jlocalL filter, Apply Default Model
Non-blocking Via Analysis	List via types, Assign non-blocking vias, Merge trees
Reliability Budget	Read Cell Reliability Characterization, Full Chip Metric, Plot % failure, Rank Failure Units
Help	Documentation, About

5.3 Circuit Layout Parsing

SysRel reads layout files in Magic format, referred to as *.mag* files, created with Magic and 3D-Magic. A *.mag* file provides an ASCII file representation of a circuit layout with coordinates of rectangular tiles representing mask layers [37]. A set of coordinates has a header tag << layer >>, where "layer" defines the mask type of following rectangles. Table 5.2 shows a sample *.mag* file. While parsing such a file, SysRel decomposes rectangular elements into mask layers and their positions (source code: *ParseMAGFile.java*). The mask information is stored in a data-structure known as corner-stitching introduced in Magic [64]. According to the corner-stitching data-structure, multiple planes, such as active, oxide, poly, and metal1, are superimposed on each other to represent a layout. Each plane contains different types of non-overlapping rectangular tiles with stitches at the four corners. Figure 5-2 illustrates a representation of three corner-stitched solid tiles in a single plane.

The Java source files defining the corner-stitching data-structure in SysRel are *Plane.java, PlaneType.java, Tile.java*, and *TileType.java*. *Layout.java* implements the class that puts together multiple *Tile* and *Plane* objects to provide an internal

Table 5.2: *A sample* .mag *layout file.*

```
magic
tech scmos
timestamp 962374816
<< polysilicon >>
rect -21 -4 -12 -1
rect -8 -4 10 -1
<< metal1 >>
rect -27 35 -21 36
rect 8 35 20 50
<< polycontact >>
rect -12 -4 -8 0
<< labels >>
rlabel polysilicon -1 -3 -1 -3 1 opa
rlabel metal1 -10 2 -10 2 1 m1th
<< end >>
```

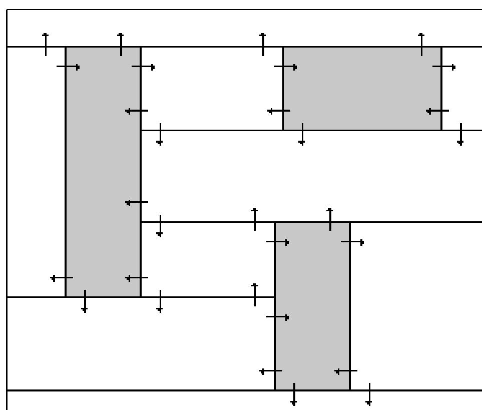

Figure 5-2: *Corner-stitched representation of a single plane with multiple tiles in a Magic layout. The gray areas are solid tiles that are corner-stitched to neighboring space tiles.*

representation of a layout. While plotting the layout on a window, the *Planes* are accessed in a serial fashion, and the corner-stitched *Tiles* in each *Plane* are painted using an appropriate color (source code: *LayoutPanel.java*).

A *.mag* file only provides the coordinates of mask layers in a layout. Therefore, further information on the connectivity and type of mask layers is retrieved from the technology file, scmos3D.tech27. While parsing a *.mag* file, the tiles that also represent contacts and inter-wafer vias are specially tagged. A contact tile has multiple representations, one at every plane that it connects. For example, m2contact connects metal interconnects from metal1 and metal2 planes, and has one tile representation at the metal1 plane and another at the metal2 plane. The Java source file, *TechDB.java*, parses the technology file to retrieve such crucial connectivity information for contacts. It also identifies the inter-wafer vias specific to 3D integration technology.

5.4 Interconnect Tree Extraction

The corner-stitching data-structure enables several key operations of a CAD layout tool in a very efficient manner. Two critical operations are locating a tile in a given area and searching for its neighboring tiles in a plane. Both operations are at the heart of extracting interconnect geometries from a layout. Given a particular tile, all adjacent tiles are retrieved from the corner-stitches implemented as an ArrayList[3]. The adjacent tiles are then stored in a Vector[4] (source code: *ITree.java*) [63]. Using a depth-first search algorithm, the top left-most tile is identified, and the whole tree is built via a depth-first walk on adjacent tiles [46]. Figure 5-3 illustrates an interconnect tree extraction from corner-stitched representation of tiles.

Setting the top left-most tile as a starting point facilitates the computation of several useful interconnect parameters including via-to-via paths in an interconnect tree. SysRel calculates the length of all possible via-to-via paths using the coordinates

[3]The ArrayList class in Java implements a resizable array.

[4]The Vector class is similar to ArrayList except that this implementation is synchronized in case of simultaneous element accesses.

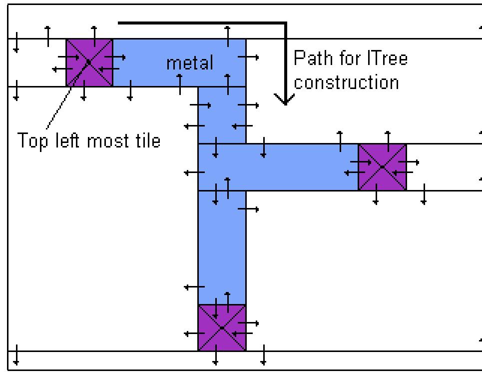

Figure 5-3: *Extraction of an interconnect tree from corner-stitched representation of tiles. Tiles marked with a cross are contact tiles and are the starting and terminating points for this particular tree.*

of rectangular tiles. Moreover, the parameters are calculated at the same time when a tree is being built with the depth-first walk algorithm. An *ITree* object, the internal representation of an interconnect tree, stores the parameters in a path-table for future uses. For the interconnect tree shown in figure 5-3, the path-table contains data for two paths, starting at the top left-most contact tile and ending at the two terminating contact tiles.

5.5 Reliability Analysis Flow

The reliability analysis flow diagram in SysRel is based on the hierarchical reliability analysis approach discussed in sections 2.3 and 2.4 for Cu and Al metallizations, respectively. Figure 5-4 shows the flow diagram for reliability analysis in SysRel. In addition to a layout file in Magic format, SysRel requires input for process parameters and critical stress numbers for use in the default model. Current estimates in interconnects trees are based on simulation results from commercial power analysis tools, such as Spice or PowerMill. While Vdd and Gnd lines are stressed at DC currents computed from total average power dissipation of the circuit, current flow in signal networks are assigned from average cell power dissipations. This gives conservative current estimates in the interconnect trees. Detail current flow information including current directions in different segments of an interconnect tree is not available from

any power analysis tool, and more importantly, associated computation is often not feasible in a large circuit layout. The electromigration parameters, *.emparam*, file defines necessary parameters for immortality filtering thresholds, metallization specific parameters in the default model, and power dissipation in the circuit.

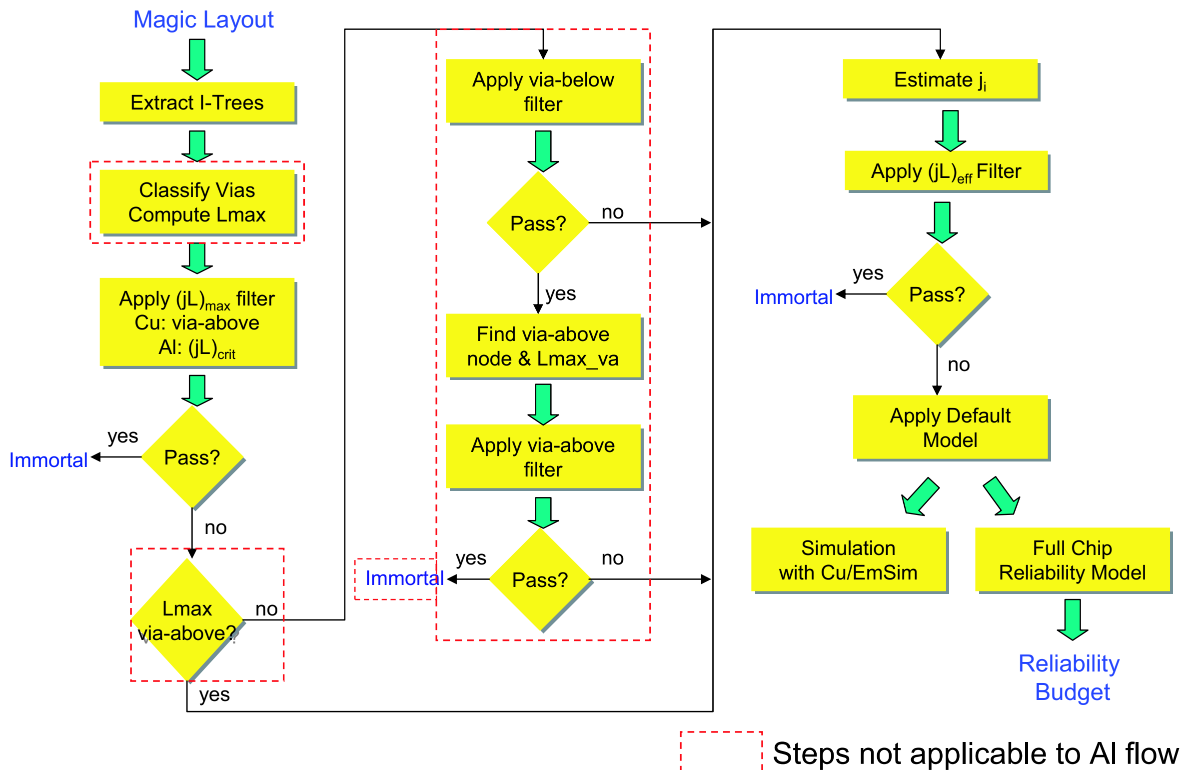

Figure 5-4: *Reliability analysis flow diagram in SysRel. The steps boxed in dashed rectangles apply only to reliability analysis in Cu metallization technology.*

SysRel is a GUI based interactive tool requiring user input before initiating a process in the flow diagram. A technology selection tab allows a user to choose either Cu or Al metallization technology for current analysis. A user can select all or any particular metallization layer in a layout for tree extraction. Extracted trees are high-lighted along a display of statistics. While a single (jL) product threshold is used for the immortality condition in Al metallization technology, via-based immortality filtering with separate treatments for via-above and via-below types in Cu metallization requires extra steps as marked in the flow diagram. After each filtering stages, the immortal trees are highlighted in green and discarded from further

analysis. The maximum current density, $j_{\max}$ is read from the emparam file. Local current density j_{local} is conservatively calculated using current flow estimation from average cell power dissipations. As mentioned earlier, segment-based current density values and directions[5] are unavailable. The worse case reliability occurs when currents from all connected segments are flowing into a via-above junction and flowing out of a via-below junction in Cu metallization, and either flowing into or out of a W-filled via junction in Al metallization [10]. We conservatively assume the worst case current directions during the $(jL)_{eff}$ filtering stage and time-to-failure calculations. Time-to-failures (t_{50}) of mortal trees are computed using the default model discussed in section 2.5. Finally, times-to-failure from different trees are combined using a joint stochastic process of series elements of lognormal lifetime distribution with standard deviation (σ) as reported in experimental work. Given a target lifetime of the chip, the full-chip reliability metrics for output are probability of no failure, maximum failure rate in FIT, and time to any cumulative % failure.

As illustrated in figure 5-4, there exists a link to MIT-Emsim or Cu-Emsim which is in parallel to full-chip reliability analysis. MIT-Emsim and Cu-Emsim are microstructure level electromigration simulators currently under development [50, 51]. Microstructure level simulations involve detail numerical analysis. Although microstructure level simulations are computation intensive, a more accurate estimate of stress evolution with various current directions in interconnect trees and electromigration lifetimes are derived using such simulations. Only the least reliable trees, as indicated by SysRel, would be passed to MIT-Emsim or Cu-Emsim.

5.6 Simulation with 32-bit Comparator Layout

A 32-bit comparator circuit layout is used to verify functionality and accuracy of the reliability flow in SysRel. The 32-bit unsigned integer comparator is designed using a design flow from behavioral description with VHDL to physical layout. The tools used in the flow are Design Analyzer (analysis and synthesis tool), Silicon Ensemble

[5] All signal networks, except for Vdd, Gnd, and clock, carry bidirectional currents.

(place and route tool), and Magic (layout editor). An IIT standard cell library for TSMC 0.18um process is used for synthesis [65]. Total dynamic power reported in Design Analyzer using the library cell characterization data is 26.6mW at $Vdd = 5V$. It is also possible to use a circuit simulator, such as HSpice, for estimating power dissipation. For this reliability analysis, we used the power dissipation as calculated by Design Analyzer. The layout has up to five layers of metallization; metal1 and metal2 for intra-cell routing, and all metal layers for inter-cell routing. Power delivery lines are in metal1 and metal2 in a ring format (figure 5-5). The layout size is approximately

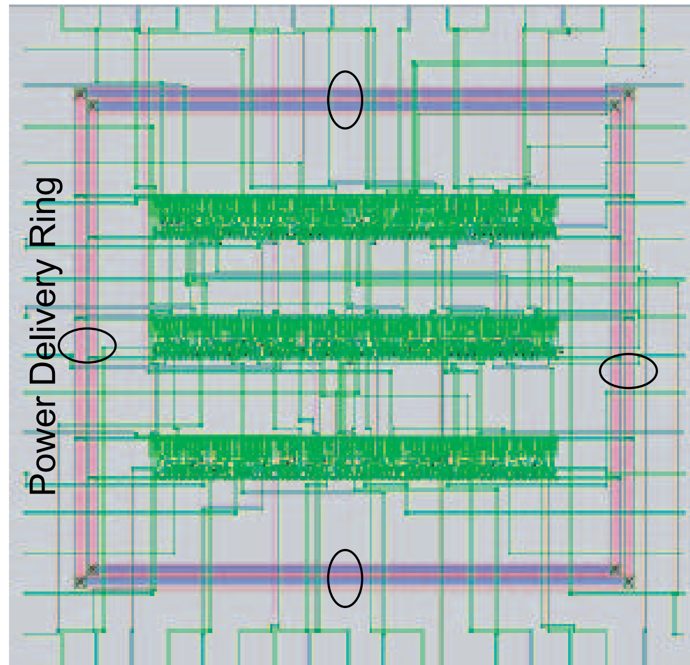

Figure 5-5: *The 32-bit comparator circuit layout analysis in SysRel. Eight mortal trees come from the power delivery rings as marked in the layout.*

164um × 164um. According to the ITRS, a $j_{\max}$ value of 0.96 MA/cm^2 and worst case interconnect temperature of 105^oC are used. While the signal lines in the design are stressed with bidirectional currents, the Vdd and Gnd lines have the worst case unidirectional current density of 0.42 MA/cm^2.

5.6.1 Copper Metallization Technology

Table 5.3 shows the results of hierarchical reliability analysis with Cu metallization technology in the 32-bit comparator circuit layout. After applying the first immortality filtering with via-based $j_{\max}L$ thresholds, 1023 out of 1143 trees are found to be immortal. The second filtering step using the tree's $j_{local}L$ products finds 112 more trees to be immortal. Figure 5-5 shows the layout after the final simulation step when the mortal trees are marked. According to the analysis, only eight interconnect trees in Vdd and Gnd lines in the power delivery rings are prone to electromigration failure. Four metal1 trees are via-above type with lifetimes of 23.2 years. The other four trees are metal2 via-below type lines with lifetimes of 168.7 years. The resulting full-chip reliability metrics are shown in step 5 of table 5.3.

Table 5.3: *SysRel simulation results of hierarchical reliability analysis with* Cu *metallization in the* 32*-bit comparator circuit layout.*

Step 1	Layout Extraction (total # of interconnect trees=1143)	
	Metal Plane	# of interconnects
	metal1	580
	metal2	438
	metal3	102
	metal4	20
	metal5	3
Step 2	Via-based ($j_{\max}L$) filter	
	Number of immortal trees identified	1023
Step 3	Via-based ($j_{local}L$) filter	
	Number of immortal trees identified	112
Step 4	Default Model (with $\sigma_{nucl} = 40MPa$) on 8 mortal trees	
	t_{50} of 4 mortal tees in metal1 (via-above)	23.2 years
	t_{50} of 4 mortal tees in metal2 (via-below)	168.7 years
Step 5	Full chip stochastic analysis ($\sigma = 0.81$, lognormal)	
	Target chip lifetime	10 years
	Probability of no failure	0.524
	max FIT	$15.4k$ 10th year
	t_{50} for full chip	10.35 years

5.6.2 Aluminum Metallization Technology

Table 5.4 shows the results of hierarchical reliability analysis with Al metallization technology in the 32-bit comparator circuit layout. After applying the first immortality filtering with a single (jL) product threshold, 1093 out of 1143 trees are identified as immortal. The second filtering step using the tree's $j_{local}L$ products finds 42 more trees to be immortal. The default model is applied to the same eight mortal trees in power delivery lines shown in figure 5-5. As no via-classification is required in Al technology, all the trees have lifetimes of 4.39 years. The lifetimes are very low as the $2.5um$ wide lines, wider than an Al grain size of $0.5um$, have Al polygranular microstructure. The full-chip metrics are significantly lower than those from Cu metallization analysis.

Table 5.4: *SysRel simulation results of hierarchical reliability analysis with Al metallization in the* 32*-bit comparator circuit layout.*

Step	Description	Value
Step 1	Layout Extraction (total # of interconnect trees=1143)	
Step 2	$(j_{\max}L)$ filter	
	Number of immortal trees identified	1093
Step 3	$(j_{local}L)$ filter	
	Number of immortal trees identified	42
Step 4	Default Model (with $\sigma_{nucl} = 500MPa$) on 8 mortal trees	
	t_{50} of mortal tees	4.39 years
Step 5	Full chip stochastic analysis ($\sigma = 0.81$, lognormal)	
	Target chip lifetime	10 years
	Probability of no failure	3.309×10^{-7}
	max FIT	$205k$
	t_{50} for full chip	1.43 years

5.7 Implications of SysRel

Reliability simulations with the 32-bit comparator layout clearly demonstrate the basic functionality of SysRel. More importantly, SysRel identifies electromigration critical (mortal) trees and allows a designer to focus on those trees while ignoring numerous other immortal trees. "What-if" analysis is possible on selected mortal

trees to observe the impact on full-chip reliability. In reliability simulation with Cu metallization in the comparator circuit layout, the worst lifetimes of 23.2 years in four mortal trees are due to via-above type Cu line in metal1. Increasing only those line widths from $2.5um$ to $5um$ and running the reliability simulation again, SysRel predicts lifetime improvement in the worst case lines from 23.2 years to 92.9 years and the new full-chip reliability metrics are $t_{50} = 37.5$ years, FIT 563, and probability of survival=0.98.

In reliability simulation with Al metallization, all eight mortal trees are Al polygranular type lines due to their $2.5um$ widths. As discussed in section 4.3, narrow Al lines (width $\leq$ grain size=$0.5um$) have bamboo type microstructure and higher electromigration lifetimes due to lower diffusivity. Using a track layout method with narrow ($0.5um$) metal lines, as illustrated in figure 5-6, we can convert the mortal lines in the power supply rings from polygranular type to bamboo type. Using such a technique, the lifetime of each metal line can be improved to 306.82 years and the new full-chip reliability metrics are $t_{50} = 99.8$ years, FIT 5.9, probability of survival= 0.99.

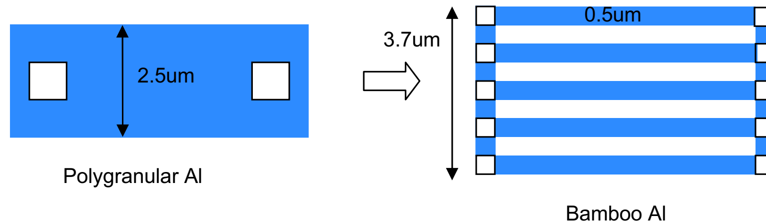

Figure 5-6: *Layout technique for Al polygranular to bamboo type conversion in an interconnect line. Let current through the interconnect line is I A. Then current densities,* $j_{polygranular} = I/(2.5 \times h)$, $j_{bamboo} = (I/5)/((2.5/5) \times h) = I/(2.5 \times h)$. *Current density remains the same in this conversion. We can treat the bamboo Al track line as a single tree.*

5.8 Three-Dimensional IC Analysis in SysRel

A 3D 32-bit comparator circuit is laid out from the 2D version using data-path folding technique. The 2D circuit is first synthesized with four rows of cells. Then the data-path is folded to create two wafers in the 3D circuit such that each wafer contains

two rows. 3D-Magic is used to place inter-wafer vias and power delivery rings in each wafer. The layout sizes corresponding to top and bottom wafers are $141.6um \times 75.5um$ and $141.6um \times 79.7um$. Figure 5-7 shows the layout view in SysRel.

Table 5.5 lists the reliability simulation results of the 3D 32-bit comparator layout with Cu metallization technology. Out of 1115 trees from two wafers, a total of eight trees, four from each wafer, are mortal as marked in figure 5-7. The mortal trees are in metal1 in the power delivery rings. The default model predicts lifetimes of each tree to be 92.97 years. The full-chip reliability metrics as shown in the table are better than those of the 2D layout. As total power dissipation is distributed in two wafers in the 3D circuit, current densities in the mortal lines are approximately scaled by half. In 2D layout, metal2 lines in the power delivery rings were mortal. However, due to data-path folding, those metal2 lines in the 3D circuit layout have lower $L_{\max}$ and get filtered out. Thus the 3D circuit demonstrates reliability improvement due to wire length reduction and power distribution in two wafers.

Table 5.5: *SysRel simulation results of hierarchical reliability analysis with Cu metallization in the 3D 32-bit comparator circuit layout.*

Step 1-3	Layout Extraction (total # of interconnect trees=1115, # of mortal trees=8)	
Step 4	Default Model (with $\sigma_{nucl} = 40MPa$) on 8 mortal trees	
	t_{50} of mortal trees	92.97 years
Step 5	Full chip stochastic analysis ($\sigma = 0.81$, lognormal)	
	Target chip lifetime	10 years
	Probability of no failure	0.976
	max FIT	1020
	t_{50} for full chip	30 years

During the reliability simulation of the 3D circuit, a worst case interconnect temperature of 105^oC, as reported by the ITRS, was used for both wafers. Temperature in 3D circuits are expected to be higher than that in 2D circuits due to higher power density in a smaller foot-print [30]. Electromigration lifetime is exponentially dependent on temperature. Therefore, reliability improvement illustrated in this analysis is not necessarily conclusive and is highly dependent on thermal management techniques used in the 3D circuit. In addition, reliability of inter-wafer vias in the presence of

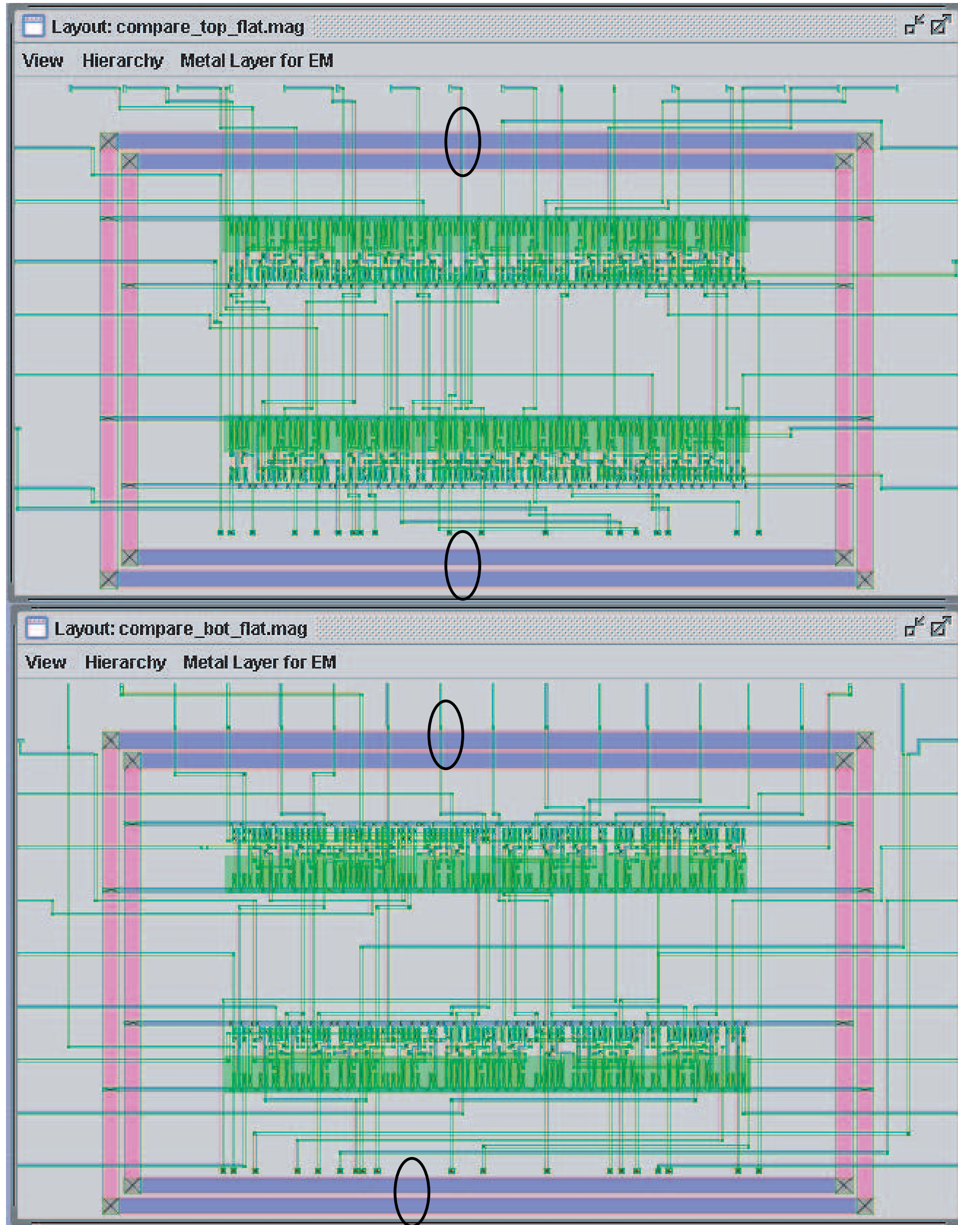

Figure 5-7: *The 3D 32-bit comparator circuit layout for SysRel reliability simulation. Mortal trees in each wafer are marked in the layout.*

bonded interface are under active study. No lifetime models for inter-wafer vias are yet available for application to SysRel.

5.9 Summary

A new reliability CAD tool, SysRel, has been developed for circuit-level electromigration analysis with either Cu or Al metallization technology in conventional and 3D circuits. SysRel is written on the Java 2 platform. SysRel is an interactive graphical user interface based RCAD tool that treats circuit layouts from Magic and 3D-Magic. The basic reliability analysis in SysRel is based on hierarchical reliability analysis with interconnect trees as the fundamental reliability units. Based on the differences in electromigration failure mechanisms observed in experimental work, we present a (jL) product filtering algorithm with a classification of separate via-above and via-below treatments in Cu metallization. After the filtering of immortal trees, a default model is applied to the remaining trees to compute reliability figures for individual units. SysRel utilizes new reliability metrics based on the desired lifetime of a chip and combines reliability figures from individual fundamental reliability units. A 32-bit comparator circuit layout in 2D and 3D technology is analyzed in SysRel to demonstrate its functionality and significance. SysRel identifies electromigration critical (mortal) trees and allows a designer to focus on those trees while ignoring numerous other immortal trees. "What-if" analysis is possible on selected mortal trees to investigate the impact on full-chip reliability.

Chapter 6

Cell-Based Reliability Analysis in SysRel

Large digital circuits, such as microprocessors and application specific ICs (ASICs), are often designed and laid out using a hierarchical scheme with blocks or modules. The hierarchy is crucial in managing the design complexity as well as facilitating parallel design efforts in the design team. Each block or module is thus designed and laid out taking the requirements into account from neighboring blocks and overall product specification. Characterization for power and timing are often done block-by-block before the full-chip is integrated. To integrate electromigration analysis into such an IC design flow, we have introduced the concept of cell-based reliability analysis in SysRel. According to this concept, each cell or module layout is characterized for electromigration reliability using SysRel. When a large cell-based layout is analyzed that uses pre-characterized cells, SysRel reuses reliability results of those particular cells.

Cell-based reliability analysis is ideal for integrating electromigration analysis into a conventional IC design flow. A block or module can be characterized for reliability while it is characterized for power and timing. Moreover, reusing cell characterization data reduces computational load while analyzing a very large circuit. Treating a large flattened layout, such as that of a microprocessor, is impractical in most CAD tools. Block level abstraction is already in use to manage the computation load

of CAD applications. This chapter outlines the infrastructure in SysRel for cell-based electromigration reliability analysis. The accuracy and advantages of cell-based analysis with SysRel are demonstrated using adder and comparator circuits.

6.1 Cell Characterization in SysRel

According to the cell-based reliability analysis, first, a cell or module layout is analyzed for electromigration reliability using SysRel. The output is formatted to record the number of mortal units in the cell and corresponding lifetimes. The outputs from each cell are stored in a cell characterization file, *sysrelcell.char* file, in a format shown in table 6.1.

Table 6.1: *Cell reliability characterization data format in* sysrelcell.char *file*

```
# example cell data for immortal cell
Cellname xor2x1
Number-of-mortal-FRU 0

# example cell data for mortal cell with 4 FRUs
Cellname or2x1
Number-of-mortal-FRU 4 20 21 22.3 54 years
```

A standard cell library developed at Illinois Institute of Technology (IIT) with 27 basic cells, such as nand, nor, inv, and aoi gates, is used in this work to synthesize cell-based circuit layouts [65]. All the cells have been characterized using SysRel to create the *sysrelcell.char* file. Due to short line length in cell layouts (maximum cell size is approximately $13um \times 4um$), all the cells are immortal after the $(j_{\max}L)$ product filter with both Al and Cu metallizations. Figure 6-1 shows immortality after $(j_{\max}L)$ product filter in the largest cell, FAX1[1], in the cell library.

[1] FAX1 implements a binary adder with $YC = ((A\&B)|(B\&C)|(C\&A))$ and $YS = (A \oplus B \oplus C)$ with inputs A, B, and C.

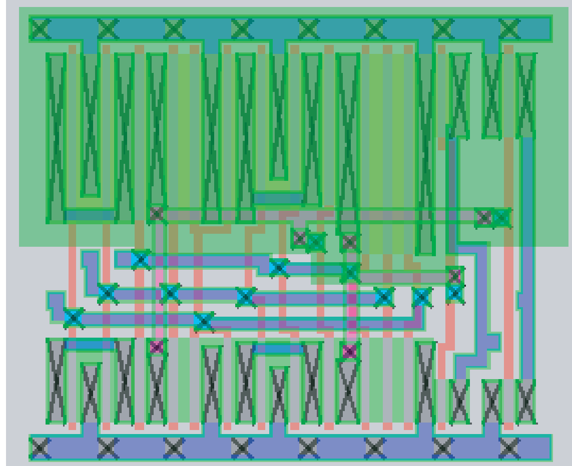

Figure 6-1: *The largest cell, FAX1, characterization with SysRel. All trees highlighted in green indicates immortality after the ($j_{\max}L$) product filter.*

6.2 Hierarchical Layout Support in SysRel

A hierarchical Magic layout file instantiates different types of cells in the layout with references to separate layout files corresponding to those cells. SysRel identifies cell instantiations in a hierarchical root layout, and then reads the corresponding layout file. In the case of multiple instantiations of the same cell, cell layout is read only once. SysRel Output Window lists the statistics on the types and number of cells used in a layout. The hierarchy levels are preserved after reading the root layout. The number of levels is shown in the "Hierarchy" menu of the Layout Window. Figure 6-2 shows a cell-based layout with three levels of hierarchy.

Using the radio buttons in the "Cell View" menu, users can select any particular hierarchy level for display. The flat root level displays the metal layers from cells in all levels as shown in figure 6-2. When any particular level is selected via a radio button, all lower levels including the selected one get hidden in the layout window. Hidden cells are outlined and their instance names appear at the center.

While extracting interconnect trees from a hierarchical layout, SysRel extracts trees only from the metal layers that are visible in the Layout Window. Thus, extracting trees at the flat root level would extract trees from metal layers in all cells including the topmost level. If the cells are already characterized for reliability, a user

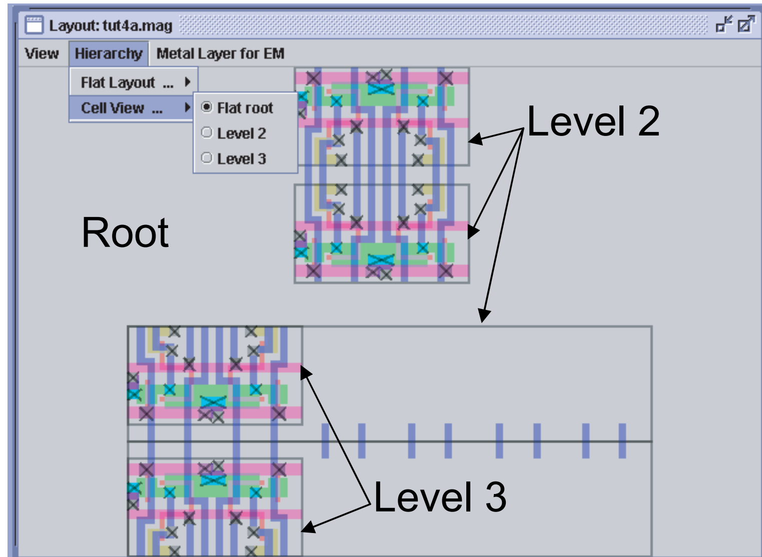

Figure 6-2: *Hierarchical layout view in SysRel.*

can exclude those from tree extraction by selecting their hierarchy level. In that case, interconnect trees are extracted from higher levels and fed into the basic reliability analysis flow diagram described in section 5.5. After the lifetimes of mortal trees are computed, SysRel reads the cell characterization file to include lifetimes for any mortal units in the hidden cells. The “Read Cell Characterization (.cellparam) file...” menu item in the “Reliability Budget” menu invokes a file chooser dialog box to select the *.cellparam* file. The full-chip reliability is computed using a series system model of all mortal units, as described in section 3.4.

6.3 Cell-based Reliability Simulation

A simple 4-bit adder, synthesized using the IIT cell library, is first used to debug and verify accurate cell-based reliability analysis in SysRel. The adder layout has two levels of hierarchy: root level consisting of inter-cell routing and power delivery lines, and level 2 consisting of 18 cells from the library. Figure 6-3 illustrates level 2 view of the layout in SysRel. Extracting interconnect trees from this view will extract trees from inter-cell routing and power lines only. Table 6.2 lists the statistics of interconnect trees and cells in the adder circuit layout.

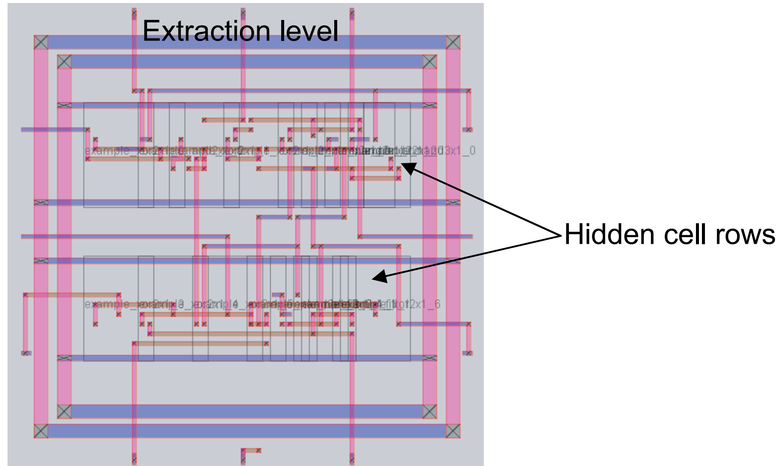

Figure 6-3: *Cell-based adder layout's level 2 view in SysRel.*

Table 6.2: *Statistics of Fundamental Reliability Units in the 4-bit adder layout.*

Total # of trees in flat layout	317
# of trees from inter-cell routing	116
% of trees from cells	63.4
Total number of cells	18
Cell Report	
Cell name	# of instances
nand3x1	1
and2x1	1
nand2x1	5
or2x1	2
fill	2
xor2x1	7

Data in table 6.2 illustrate that 63.4% of the interconnect trees come from cells. Thus the cell-based full-chip reliability analysis is expected to increase computational efficiency in SysRel. Moreover, the cell report also suggests that only a few types of cells are used multiple times in the circuit layout. SysRel full-chip reliability analysis with the 4-bit adder circuit predicts immortality in both cell-based and fully flattened layout analyses.

Section 5.6 presented SysRel reliability simulations with a fully-flattened 32-bit comparator circuit layout. The hierarchical version of the same layout is analyzed here using the cell-based approach. Table 6.3 lists the statistics of interconnect trees

and cells in the comparator circuit layout.

Table 6.3: *Statistics of Fundamental Reliability Units in the* 32*-bit comparator circuit layout.*

Total # of trees in flat layout	1143
# of trees from inter-cell routing	678
% of trees from cells	40.7
Total number of cells	132
Cell Report	
Cell name	# of instances
oai21x1	48
aoi21x1	14
nand2x1	1
invx1	55
fill	14

Table 6.3 illustrates that 40.7% of the interconnect trees come from cells in this circuit. Similar to the adder circuit layout, only few types of cells are used multiple times in the comparator circuit. SysRel reliability analysis with the 32-bit comparator circuit predicts the same eight trees from power delivery lines to be mortal in cell-based analysis with both Al and Cu metallizations. Therefore, the full-chip reliability results are the same as presented in section 5.6.

6.4 Computational Efficiency

The two metrics used to investigate the computational efficiency with cell-based reliability analysis are the times required for layout drawing and interconnect tree extraction. Layout drawing in the Layout Window entails painting a collection of rectangles using the Java Graphics2D class. Rectangles are drawn in different colors to represent different mask layers. When a cell-based layout is viewed in any hierarchy level other than root level (figure 6-3), cell layouts are hidden in outlined boxes. Table 6.4 lists the time required for layout drawing with the 4-bit adder and 32-bit comparator circuit layouts. The layouts are drawn while zoomed to fit the Layout Window of approximately $3inch \times 3inch$. The time recorded is the average time computed using a few trials.

Table 6.4: *Time required for layout drawing in SysRel.*

4-bit adder circuit		32-bit comparator circuit	
Analysis mode	time (ms)	Analysis mode	time (ms)
Cell-based	125	Cell-based	563
Flattened layout	313	Flattened layout	1250

Interconnect trees are extracted from the layout using only the visible tiles in the Layout Window. During a cell-based analysis, interconnect trees inside a cell are not extracted as the cell layouts are just outlined. Table 6.5 lists the time required for extracting interconnect trees in the 4-bit adder and 32-bit comparator circuit layouts during different modes of analysis.

Table 6.5: *Time required for extracting interconnect trees in SysRel.*

4-bit adder circuit		32-bit comparator circuit	
Analysis mode	time (ms)	Analysis mode	time (ms)
Cell-based	94	Cell-based	265
Flattened layout	172	Flattened layout	485

We observe significant reductions in times required for layout drawing and interconnect tree extraction in tables 6.4 and 6.5, respectively. Table 6.6 summarizes the percentage improvement using the two metrics to quantify computational efficiency in cell-based reliability analysis in SysRel.

Table 6.6: *Computational efficiency with cell-based reliability analysis in SysRel.*

	% improvement	
	Layout drawing	tree extraction
4-bit adder	60	45
32-bit comparator	55	45

6.5 Summary

Large circuits are often designed using a hierarchical flow where individual blocks or cells are first designed and characterized. In this chapter, we have introduced cell-level reliability characterization and computationally efficient full-chip reliability analysis

in SysRel. Cell layouts can be characterized for reliability while they are characterized for power and timing. The outcome of reliability characterization is the number of mortal trees and the lifetimes. An IIT standard cell library has been characterized using SysRel to create the reliability characterization file *sysrelcell.char*. SysRel reads and retains the hierarchy levels in a hierarchical Magic layout. Reliability simulation results with two hierarchical Magic layouts, a 4-bit adder and 32-bit comparator, demonstrate the proper functionality and significance of cell-based reliability analysis in SysRel.

Statistics from the hierarchical layouts suggest that only a few cells are used multiple times and a large percentage (minimum 40% found in the 32-bit comparator layout) of total extracted interconnect trees are internal to cells. Therefore, cell-based reliability analysis, where cells are pre-characterized and hidden in a hierarchical layout, significantly reduces simulation time in SysRel. The two metrics used to investigate the computational efficiency with cell-based reliability analysis are the times required for layout drawing and interconnect tree extraction. The maximum improvement in times required for layout drawing and interconnect tree extraction of 60% and 45%, respectively, have been demonstrated with the 4-bit adder and 32-bit comparator layouts. Both cell-based and flattened layout analyses produce the same full-chip reliability results.

Chapter 7

Non-blocking Via Analysis in Copper Metallization Technology

An interconnect tree is the fundamental reliability unit for circuit-level reliability assessments for metallization schemes with fully-blocking boundaries at the vias. A tree is composed of linked segments in a single layer of metallization, in which metal atoms can freely diffuse and lead to coupled stress evolution in its segments. Because of blocking boundaries at vias, stress in a tree is not affected by stress in other trees. The tree-based hierarchical reliability analysis flow described in section 5.5 applies to circuit-level reliability assessments with fully-blocking vias. In Al metallization technology, tungsten-filled vias always block electromigration. In Cu metallization technology, refractory liners at the bottom of vias generally block electromigration. However, some experimental studies suggest that this is not always the case, and as liner thicknesses are decreased, fully-blocking liners at vias become less certain due to liner ruptures. When Cu-filled vias are not fully blocking, an interconnect tree is connected to other metallization layers. While liner ruptures can lead to increased lifetimes in test structures, the impact of tree linking on circuit-level reliability is not clear. Therefore, we have incorporated in SysRel a capability for making reliability analysis with non-blocking vias in a layout [66].

7.1 Non-blocking Via Due to Liner Ruptures

Cu interconnects are fabricated by the dual-damascene method, in which a trench is first etched into a blanket layer of dielectric material (usually SiO_2) before filling it with Cu by electroplating. Since Cu undergoes field enhanced diffusion in most dielectric materials (including SiO_2), in order to prevent Cu atoms from diffusing into the device layer, thin refractory metal layers, referred to as liners, consisting of Tantalum (Ta) or Tantalum Nitride (TaN) are placed at the sides and bottom of the Cu line. When vias are filled with Cu, there exists a liner at the bottom of a via.

Hu *et al.* carried out Cu electromigration studies with different liner thicknesses and blocking materials at the cathode and anode ends of the lines [67]. Figure 7-1 illustrates three layers of Cu metal interconnects, used in their experiments, connected by vias V1 and V2. The liner thicknesses at the bottom of a via are $3nm$ and $10nm$ for thin and thick liners, respectively. The study demonstrated that a $3nm$ thick liner can lead to an increase in mean-time-to-failure by more than an order of magnitude.

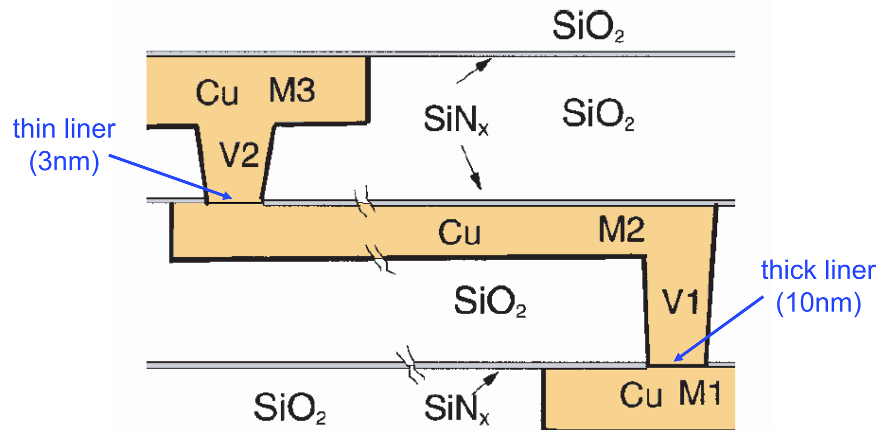

Figure 7-1: *Multi-level Cu interconnects with different liner thicknesses [67].*

Wei *et al.* conducted electromigration experiments with straight line via-to-via Cu dual-damascene test structures as shown in figure 7-2 [68]. Few $1000um$ long lines tested at $j = 2.0MA/cm^2$, and $800um$ long lines tested at $j = 2.5MA/cm^2$ never failed even though analytically computed mechanical stress values in the lines were high enough for void nucleation. The apparent immortality of a sub-population of

long lines is postulated to be the result of stress-induced ruptures of the Ta liners at the vias. In a test structure in figure 7-2, liner ruptures would allow continuous flow of Cu to and from the lead lines and contact pads, which serve as large sinks and reservoirs for Cu atoms. Therefore, the liner ruptures would cause immortality in those types of lines.

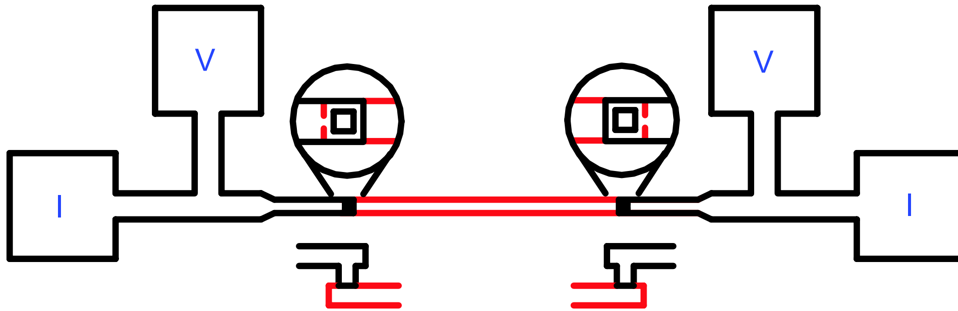

Figure 7-2: *Straight line via-to-via Cu dual-damascene test structure for electromigration experiments.*

7.2 Non-blocking Via Assignment in Circuit Layout

In Cu metallization technology, metal1 to poly-silicon gate, diffusion, such as source/drain, substrate, and well contacts are tungsten (W) filled similar to vias in Al metallization. Those vias always block electromigration. On the other hand, Cu-filled vias connecting two metal layers have Ta-based liners at the bottom. Vias with liners may or may not block electromigration depending on the liner thickness, stress build-up, and other material properties. Therefore, non-blocking via assignment in SysRel is designed to be stochastic in nature. Table 7.1 illustrates the input parameters for non-blocking via analysis defined in *sysrel.emparam* file.

The first line in table 7.1 is "stochastic 90" indicating that 90% of potentially non-blocking vias in the circuit layout will be assigned to be non-blocking. The stochastic value needs to be greater than 0% and can be as high as 100%. The 100% value would indicate a deterministic assignment of non-blocking boundaries in all metal-to-metal

Table 7.1: *Input parameters for non-blocking via analysis defined in* sysrel.emparam *file.*

Non-blocking via analysis in Cu
Stochastic 90
random_seed 123412342
force_blocking 5 polycontact ndcontact pdcontact
psubstratepcontact nsubstratencontact

vias in the layout. The second parameter in table 7.1 is a random seed that needs to be a maximum of 9 digit number for the pseudo-random number generator in SysRel. During stochastic non-blocking via assignment, SysRel tags each potentially non-blocking via with a random number. The random numbers are sorted, and then a cut-off value is picked from the sorted list that would assign the desired percentage of vias to be non-blocking. The same random_seed value during multiple SysRel sessions will assign the same set of vias in a circuit layout to be non-blocking. The final parameter for non-blocking via analysis allows a user to list a set of vias to be treated as fully blocking. The entry in table 7.1 lists five W-filled gate, source/drain, and substrate contacts that need to be treated as fully blocking in Cu metallization technology. A user can also list any particular metal-to-metal vias, such as metal2 to metal3 via, to be fully-blocking. Such flexibility would allow a user to selectively investigate the impact of non-blocking boundaries at different metal-to-metal vias.

7.3 Steps for Non-blocking Via Analysis in SysRel

The "Non-Blocking Via Analysis" menu bar in SysRel provides three menu items for the core operations in non-blocking via analysis (figure 7-3). The "List via types" menu item displays the number and types of vias used in the circuit layout. The "Assign Non-blocking Vias" assigns non-blocking vias in the layout. The "Merge trees" links the trees from different metallization layers connected by non-blocking vias. The tree merging process is described in detail in the next section.

The steps for reliability analysis in SysRel with the presence of non-blocking vias in a layout are as follows.

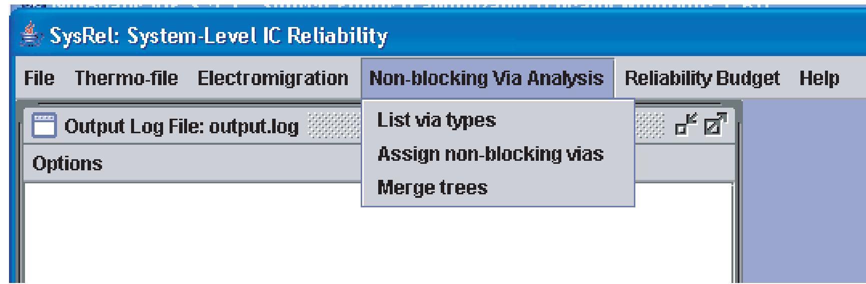

Figure 7-3: *Menu options for reliability analysis in SysRel with non-blocking vias in a circuit layout.*

1. Open a circuit layout in SysRel using "File→ Open Magic layout file" menu option. The Magic layout needs to be fully flattened. Non-blocking via assignment and analysis are not possible with a cell-based hierarchical Magic layout.

2. Read the *sysrel.emparam* file into SysRel using the menu option "File→ Read EM parameters file". The technology selection tab in SysRel needs to be set to Cu metallization.

3. Extract interconnect trees from the layout using the menu option "Electromigration→ Extract Failure Units".

4. List and investigate different types of vias used in the circuit layout using the menu option "Non-blocking Via Analysis→ List via types".

5. Assign non-blocking vias in the circuit layout using the menu option "Non-blocking Via Analysis→ Assign non-blocking vias". The Output Window will list the number of vias assigned to be non-blocking.

6. Merge the interconnect trees connected by non-blocking vias using the menu option "Non-blocking Via Analysis→ Merge trees". The Output Window will list the number of trees in the layout after the tree linking process.

7. Continue conventional reliability analysis flow (section 5.5) with the $(j_{\max}L)$ product filter, $(j_{local}L)$ product filter, lifetime calculation of mortal trees, and then full-chip reliability budget.

7.4 Interconnect Tree Merging

The interconnect trees connected by a non-blocking via are merged to create a single interconnect tree spanning multiple layers of metallization. Interconnect tree merging is a fairly complicated process as all via-to-via paths need to be updated and new via-to-via paths need to be extracted in a merged tree. Figure 7-4 illustrates the tree merging algorithm with two simple straight line trees A and B.

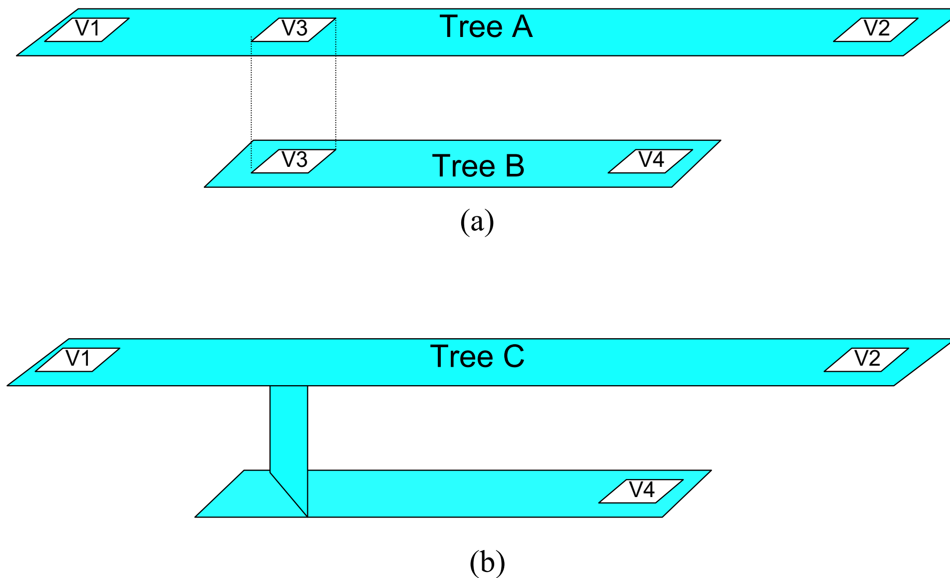

Figure 7-4: *Schematic illustration of interconnect tree merging with a non-blocking via. (a) Tree A and B are linked by non-blocking via V3. (b) Tree C is created after the two trees are merged and V3 is removed. The longest path in tree C, used as the effective length, L_{max}, in the immortality condition filter, is now from V2 to V4.*

The effective length used in the (jL) product filters, in tree A and B are from V1-to-V2 and V3-V4, respectively. Via-to-via paths in tree A are V1-to-V3, and V3-to-V2. Via-to-via path in tree B is V3-V4. Assuming that V3 is a non-blocking via, V3 is removed from both trees. New paths from V1-to-V4, V2-to-V4, and V1-V2 are created. Tree A and B are combined to create a single tree C with the new paths and tile representations from two layers of metallization (figure 7-4(b)). The effective length of tree C is from V2-V4 which is longer than those of tree A or B. The effective length in a merged tree either increases or remains unchanged.

Interconnect trees are merged recursively until all assigned non-blocking vias are

exhausted in a layout. During this process if a non-blocking via is found in a single via tree[1], the via is converted to be fully blocking. For example, both vias in a serpentine ring structure in metal2 and metal1 in figure 7-5(a) can be assigned as non-blocking. However, when the trees are merged, at least one via is required to apply the via-based default model in the merged tree (figure 7-5(b)).

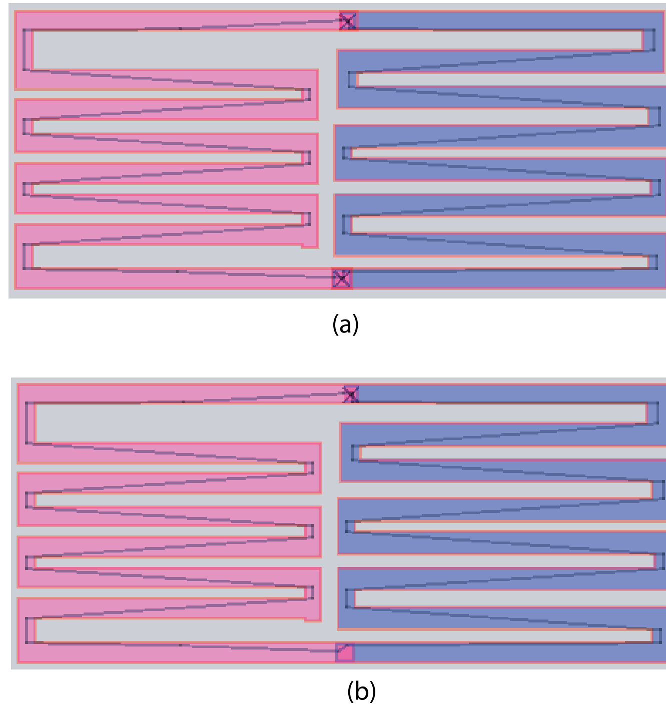

Figure 7-5: *Dynamic conversion of a non-blocking via to fully blocking via in a merged tree. (a) A serpentine ring in metal1 and metal2 with two vias assigned to be non-blocking. The longest paths are drawn in the trees. (b) One of the vias is converted to fully blocking in the merged tree. The non-blocking via is marked in 'X'.*

The proper functionality of interconnect tree merging has been carefully traced in SysRel using simple circuit layouts. Figure 7-6 illustrates interconnect tree merging and its effect in a 3-input nand gate layout with interconnect routing connected to its output. The three inputs are in poly-silicon lines. When all vias are fully blocking, there are eight trees extracted from the layout; one from Vdd, one from Gnd, and the rests from interconnect routing at the output (figure 7-6(a)). After assigning all

[1]A single via tree is defined as continuous metal segments in one metallization layer with just one via.

possible vias to be non-blocking and merging the trees accordingly, there are three trees in the layout; one from Vdd, one from Gnd, and one from output routing. The three trees and their longest paths are shown in figure 7-6(b).

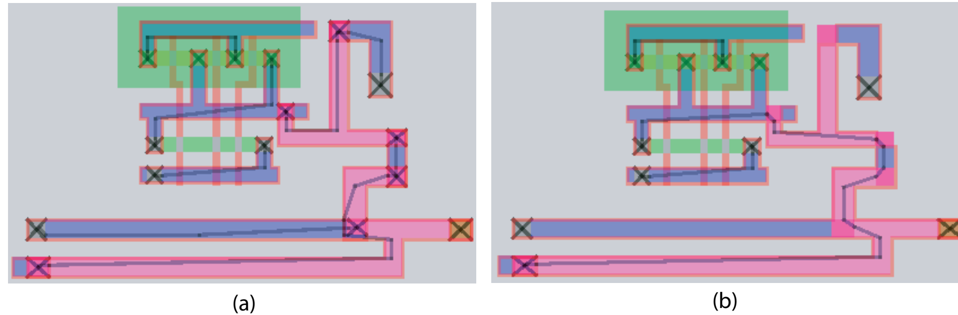

Figure 7-6: *Illustration of interconnect tree merging in a* 3*-input nand gate layout. (a) There are eight trees in the layout. (b) With all possible vias assigned as non-blocking, the number of tree goes down to three. The longest paths are marked.*

7.5 Line-length Effect on Reliability

As illustrated in the previous examples, merging interconnect trees linked with a non-blocking via either increases the effective length in the merged tree or keeps the effective length unchanged. The increase in effective line length can influence the outcome of immortality condition filters based on (jL) products. Multiple trees that would be immortal with blocking vias can be linked to form mortal trees. However, two or more mortal trees can be linked to form a single mortal tree with a longer effective length. Therefore, the presence of non-blocking vias and interconnect tree merging can lead to either an increase or decrease in the number of mortal trees.

In the hierarchical reliability analysis flow, a default model is applied to compute the lifetime of mortal trees. As described in section 2.5, the lifetime of each via in a tree is derived using the equations for time to void nucleation, extrusion, and void growth. The equations in the default model do not take into account the segment-lengths connected to a via; rather a semi-infinite length is assumed that leads to conservative lifetime estimates. Therefore, when the default model is applied to

compute the lifetime of a merged tree, the lifetime is unchanged despite an increase in the effective line length.

The effect of interconnect length on lifetime is obvious in experimental work with Al metallization technology. The lifetime of a straight via-to-via Al line as a function of its length has a trend as shown in figure 7-7(a). At a fixed stress condition, its lifetime (time-to-failure) increases as the length decreases. This is due to the fact that the compressive stress developed at the anode end of the line can interact with, and thus slow down the rate of tensile stress increase at the cathode end over a shorter distance. Figure 7-7(b) shows the time-to-failure of a Cu M1 line as observed in experimental work reported in [10]. Cu interconnects have a minimum lifetime for intermediate lengths, as the effects of backstress and non-fully blocking boundaries in vias are assumed to increase the lifetime in short and long lines respectively.

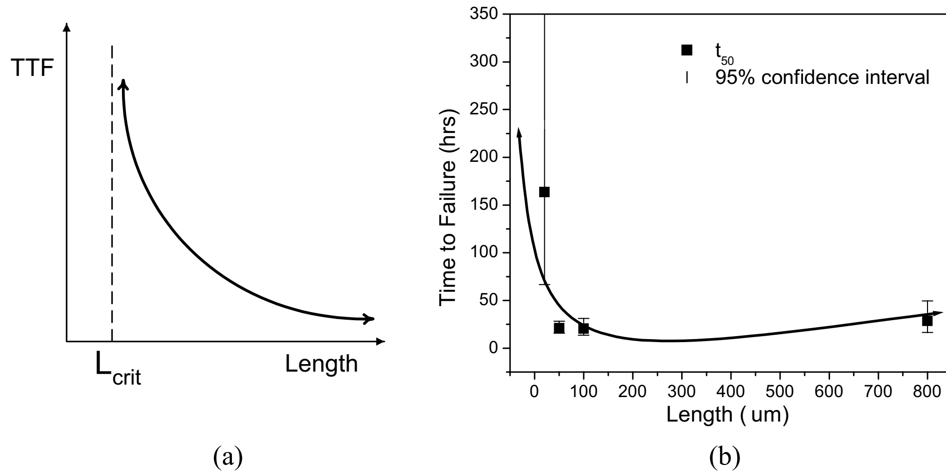

Figure 7-7: *Time-to-failure dependence on interconnect length. (a) Al interconnects have a minimum TTF for long lines due to absence of the backstress effect; (b) M1 Cu interconnects have a minimum TTF for intermediate lengths, as the effects of backstress and non-fully blocking boundaries in vias increase the TTF in short and long lines, respectively. Source: [10].*

To incorporate one possible impact of increased interconnect length during non-blocking via analysis in SysRel, we have assumed that the time-to-failure in Cu interconnect is inversely proportional to line length i.e. $TTF \propto (1/L)$. If a via at the end of an L_{ref} long segment produces a lifetime of TTF_{ref} using the default model,

the same via when connected to an L_{new} long segment (due to tree merging) would have a lifetime TTF_{new} computed as

$$TTF_{new} = \left(\frac{L_{ref}}{L_{new}}\right) \times TTF_{ref} \tag{7.1}$$

Thus a longer interconnect after tree merging would decrease the lifetime according to the increase in line length.

7.6 32-bit Comparator Layout Simulation

The 32-bit comparator circuit layout discussed in section 5.6 is simulated here to investigate circuit-level reliability impact with non-blocking vias in the layout. Full-chip reliability figures, both including and excluding line length dependence on lifetime of the mortal trees, are computed with the percentage of non-blocking via ranging from 0% to 100%. The seed for random number generator during non-blocking via assignment is set to 223123223. The target lifetime of the chip of 10 years and an uniform temperature of 105^oC are used for these simulations. Table 7.2 lists the simulation results.

When all vias are fully-blocking, the 32-bit comparator circuit layout has eight mortal trees; all from the power delivery, Vdd and Gnd, rings. As the percentage of non-blocking vias is gradually increased to 90%, the number of mortal trees decreases. This is because non-blocking vias in the power delivery rings merge multiple trees from metal1 and metal2 planes, and eventually results in one mortal tree for Vdd and another for Gnd ring. If we ignore the line-length effect on interconnect lifetime, full-chip reliability, in terms of probability of no failure and failure rate, improves because of fewer mortal units as illustrated in table 7.2. However, if the line-length effect is taken into account, the time-to-failure of mortal units decreases according to the relation in equation 7.1, and the full-chip reliability degrades.

When non-blocking via percentages in the layout are set to 90% and 100%, in addition to the two mortal trees from the power delivery rings, few immortal trees for

Table 7.2: *Reliability simulation with various degrees of non-blocking vias in the* 32*-bit comparator circuit layout.*

NB via	Total # of trees	# of mortal trees	No line-length dependence			With Line-length dependence		
			TTF (#) years	Prob. of no fail.	max FIT	TTF (#) years	Prob. of no fail.	max FIT
0%	1143	8	23.2 (4) 168.7 (4)	0.5242	15418	23.2 (4) 168.7 (4)	0.5242	15418
10%	1059	8	23.2 (4) 168.7 (4)	0.5242	15418	23.2 (4) 168.7 (4)	0.5242	15418
30%	891	7	23.2 (4) 168.7 (3)	0.5244	15405	12.21 (1) 23.2 (3) 168.7 (3)	0.3680	20694
50%	721	3	23.2 (2) 168.7 (1)	0.7242	7696.1	8.4 (2) 168.7 (1)	0.1720	26785
70%	555	3	23.2 (2) 168.7 (1)	0.7242	7696.1	8.4 (2) 168.7 (1)	0.1720	26785
90%	393	3	23.2 (2) 517.8 (1)	0.7244	7683.3	6.42 (2) 517.8 (1)	0.0854	35037
100%	311	4	23.2 (2) 517.8 (2)	0.7244	7683.4	6.42 (2) 517.8 (2)	0.0854	35037

signal routing get merged and become mortal due to an increase in effective length (figure 7-8). The new mortal trees, however, have fairly high lifetimes due to small current density in the signal lines, and thus do not degrade the full-chip reliability significantly.

7.7 Discussion of Results

The reliability simulation results with the 32-bit comparator layout demonstrate that the presence of non-blocking vias overall decreases the number of mortal units. However, merging interconnect trees linked by non-blocking vias increases the effective line length of the merged trees. If the default model for estimating the lifetime of a mortal tree assumes semi-infinite segment lengths, there is a predicted reliability improvement with non-blocking boundaries. However, if the impact of line length is taken into account according to equation 7.1, the full-chip reliability would degrade. The impact of non-blocking vias depends on how reliability depends on length, and

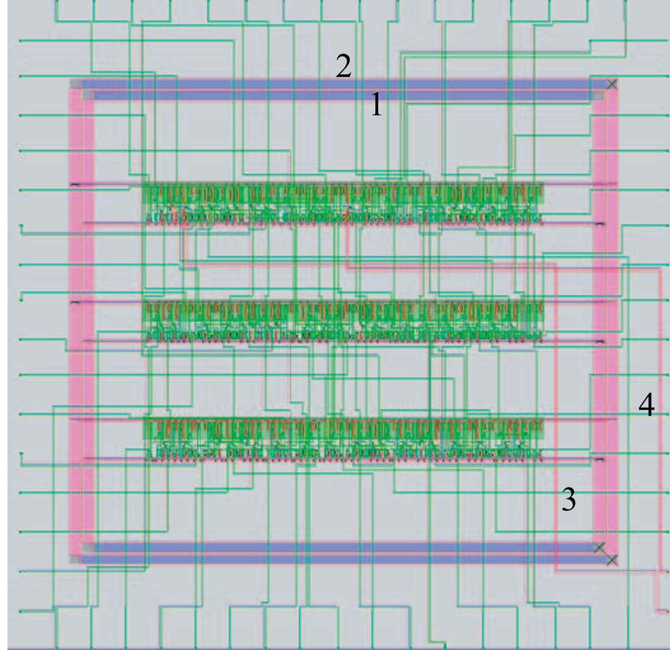

Figure 7-8: *Mortal trees in the* 32*-bit comparator circuit layout when all vias are assigned non-blocking. Tree no.* 1 *and* 2 *comprise power delivery rings. Tree no.* 3 *and* 4 *comprise signal routing nets.*

how non-blocking vias affect the line length dependence of reliability. Figure 7-9 shows two proposed test structures for investigating line-length dependence of reliability in Cu interconnects.

7.8 Summary

Some electromigration experimental studies postulate that Ta-based liners at the bottom of Cu filled vias in dual-damascene Cu technology do not always block electromigration. Moreover, as the liner thicknesses are decreased in future technology generations, fully-blocking liners at vias become less certain due to liner ruptures. We have incorporated into SysRel the capability of reliability simulation with the presence of non-blocking vias in a circuit layout. Users can stochastically or deterministically assign non-blocking vias in the design. Multiple trees, when linked by a non-blocking via, are merged to create a single tree, which is then treated as the fundamental reliability unit. The presence of non-blocking vias leads to either an in-

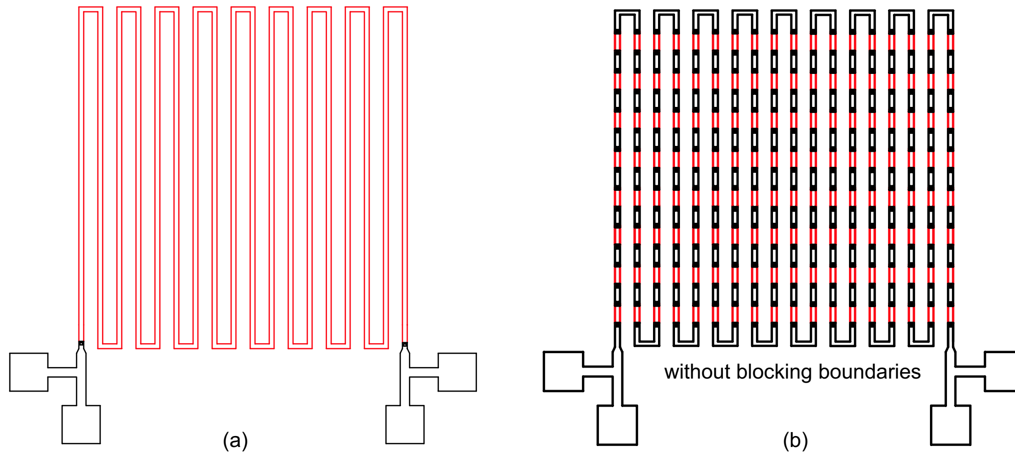

Figure 7-9: *Test structures to investigate the line length dependence of reliability in Cu interconnects. (a) A long serpentine line in one layer of metallization, and (b) a serpentine line in two layers of metallization linked by multiple non-blocking vias. The two end vias are fully blocking. This structure is to investigate line length dependence in the presence of non-blocking vias.*

crease or decrease in the number of mortal trees. Trees that would be immortal with blocking vias can be linked to form mortal trees. However, two or more mortal trees can be linked to form a single mortal tree with a longer effective length. The latter effect results in an overall reduction in the number of mortal trees, as illustrated in SysRel simulation using a 32-bit comparator circuit layout.

If the default model for estimating the lifetime of a mortal tree assumes semi-infinite segment lengths, there is a predicted reliability improvement with non-blocking boundaries. However, if the impact of line length is taken into account (e.g., the lifetime to be inversely proportional to line length), the full-chip reliability would degrade. These results from SysRel demonstrate the critical need for understanding and accurately modeling the line length dependence on reliability. We have proposed the development and use of long multiple metal level test structures with non-blocking vias for inter-metal connections and blocking vias at the ends. The results from such test structures would demonstrate line length dependence in the presence of non-blocking vias and allow accurate circuit-level reliability assessments with *Cu* metallization.

Chapter 8

Thermal-Aware Reliability Analysis in SysRel

Traditionally, thermal analysis in ICs has been confined to chip-level packaging where careful consideration is required for heat sink design [69, 70]. However, as the technology scaling continues, the factors contributing to temperature rise, such as total power consumption, power density, and interconnect current density, are consistently becoming more and more significant [71]. Consequently, thermal effects have become inseparable aspects of IC design and analysis [34, 72]. Interconnect lifetime due to electromigration is exponentially dependent on inverse temperature. Joule heating in an interconnect due to its current density can increase the temperature, and thus reduce interconnect reliability. Using a worst case temperature for reliability analysis, and then setting a reliability goal may lead to excessive conservatism in IC design. Therefore, accurate thermal analysis is a prerequisite to circuit-level reliability analysis and reliability-aware IC design. This chapter describes fundamentals of thermal analysis in the perspective of ICs. Finite element simulations using ANSYS have been conducted to investigate thermal effects in transistors. Based on our findings, a cell-based technique for estimating layout-level temperature profile is described and implemented in SysRel. Using the layout-level temperature profile as a boundary condition, SysRel then incorporates joule heating in interconnects while calculating the lifetimes of mortal trees. The 32-bit comparator circuit has been reanalyzed with

thermal capabilities and the results are discussed here.

8.1 Fundamental Concepts for Thermal Analysis

There are three modes of heat transfer: conduction, convection, and radiation [73].

Conduction is the mode of heat transfer where energy is exchanged within a solid body or between two solid bodies in contact due to temperature gradient.

Convection is the mode of heat transfer where energy is exchanged between a solid body and a surrounding fluid.

Radiation is the mode of heat transfer from a body or between two bodies by electromagnetic waves.

Thermal analysis in ICs typically involves conduction and convection. While conduction is the mode of heat transfer within a die and package, convection is the primary mode for package-level cooling. Fourier's law of heat conduction states heat flux, $\vec{q}$, as following

$$\vec{q} = -k_i \frac{\partial T}{\partial i} \tag{8.1}$$

where k_i is the thermal conductivity and $(\partial T/\partial i)$ the thermal gradient in direction i. The negative sign indicates that heat flows in the opposite direction of the gradient. The SI units for heat flux and thermal conductivity are $Watt/meter^2$ and $Watt/meter \cdot^o C$, respectively. When heat conducts a distance L from a high to low temperature in a body, Fourier's law is written as

$$\vec{q} = k\frac{\triangle T}{L} = k\frac{T_{high} - T_{low}}{L} \tag{8.2}$$

Multiplying both sides with cross-sectional area of the conductor, A,

$$\vec{q} \times A = \frac{kA}{L}\triangle T \tag{8.3}$$

$$Q = \frac{1}{R_{th}}\triangle T \tag{8.4}$$

where Q is the heat flow in $Watts$ and R_{th} is the thermal resistance of the material in $^oC/Watt$. Equation 8.4 forms the basis of electrical and thermal analogy for conduction as illustrated in table 8.1.

Convection is typically applied to heat flow from a surface to neighboring fluid. Newton's law of cooling states the heat flux due to convection as

$$\vec{q} = h_f(T_S - T_B) \tag{8.5}$$

where h_f is the convective film coefficient or heat transfer coefficient in $Watt/meter^2 \cdot^o C$, T_S is the surface temperature, and T_B is the bulk fluid temperature. Considering a surface area, A, for heat flux, we can derive an equation for heat flow, Q, as following

$$\vec{q}A = Ah_f(T_S - T_B) \tag{8.6}$$

$$Q = \frac{1}{R_{th}}(T_S - T_B) \tag{8.7}$$

where R_{th} is the thermal resistance for convection. Table 8.1 outlines the electrical and thermal analogy for both conduction and convection.

Table 8.1: *Electrical and thermal analogy for modeling steady-state heat flow.*

Electrical	Thermal
Current Flow (I) Voltage Drop $\triangle V = (V_{high} - V_{low})$ Electrical Resistance (R)	Heat Flow (Q) Temperature Drop $\triangle T = (T_{high} - T_{low})$ Thermal Resistance (R_{th})
$\triangle V = IR$	$\triangle T = QR_{th}$
$R = \frac{L}{\sigma A}$ σ : electrical conductivity L : length of conductor A : cross-sectional area	Conduction $R_{th} = \frac{L}{kA}$ Convection $R_{th} = \frac{1}{Ah_f}$ k : thermal conductivity L: length of conductor A : cross-sectional area h_f : convective film coefficient

In steady heat transfer problems, thermal resistance is the most significant physical variable. Modeling unsteady or transient heat transfer problems require thermal capacitance of an object in addition to its thermal resistance. The rate at which the temperature of an object rises or decreases is determined by its thermal resistance

and capacitance. The thermal capacitance of an object with volume, V, is expressed as

$$C_{th} = \rho_m c_p V \tag{8.8}$$

where ρ_m is the material's density and c_p is the heat capacity. The units in SI system for C_{th}, ρ_m, c_p, and V are $joule/^oC$, $kg/meter^3$, $joule/kg \cdot^o C$, and m^3, respectively. The thermal capacitance is considered to be analogous to electrical capacitance to complete the electrical and thermal analogy stated in table 8.1.

According to the first law of thermodynamics, energy in an object needs to be conserved at any instant. For thermal energy, we can write the energy conservation law as

$$E_{in} + E_{gen} = E_{out} + \frac{dE_s t}{dt} \tag{8.9}$$

where E_{in} is the thermal energy entering the object, E_{gen} is the thermal energy generated in the object, E_{out} is the thermal energy exiting the object, and $\frac{dE_s t}{dt}$ is the rate of change in the stored thermal energy. Equivalently,

$$q_{out} - q_{in} = q_{gen} - \frac{dE_s t}{dt} \tag{8.10}$$

$$\nabla q = q_{gen} - \rho_m c_p \frac{dT}{dt} \tag{8.11}$$

$$\nabla \cdot (-k \nabla T) = q_{gen} - \rho_m c_p \frac{dT}{dt} \tag{8.12}$$

Assuming the thermal conductivity is uniform in the object

$$-k \nabla^2 T = q_{gen} - \rho_m c_p \frac{dT}{dt} \tag{8.13}$$

where

$$\nabla T = \frac{\partial T}{\partial x} + \frac{\partial T}{\partial y} + \frac{\partial T}{\partial z} \tag{8.14}$$

Equation 8.13 is a simple form of the heat diffusion equation often solved numerically by using finite difference method for thermal analysis in ICs [74]. The heat diffusion equation is the governing equation for heat conduction problems. The following three boundary conditions can be applied to the object depending on the surrounding

environment:

$$\text{Isothermal (Dirichlet)}: \quad T = f(x, y, z, t) \tag{8.15}$$

$$\text{Adiabatic (Neumann)}: \quad \vec{q} = 0 \tag{8.16}$$

$$\text{Convective (Robin)}: \quad k\vec{q} = h(T - T_a) \tag{8.17}$$

where T_a is the ambient temperature.

In Finite Element Method (FEM), the governing differential equation, derived using the conservation law, is first transformed to an equivalent integral equation. The geometric model is discretized into simple shapes, such as rectangles and/or triangles in 2D, and tetrahedrons, pyramids or hexahedrons in 3D, in which the governing equations are solved [75]. Most FEM tools are capable of non-uniform discretization or meshing and evaluation of mesh discretization error. Thus, simulation results from FEM tools are often used as benchmarks for other approaches in IC thermal analysis. In this work, a commercial FEM tool, ANSYS, is used for thermal analysis.

The ability to analyze both steady-state and transient conduction and convection problems using thermal resistors and capacitors has led to Spice-based thermal simulation of ICs [76]. The substrate is discretized into volumes which is replaced with an equivalent thermal RC network. The power dissipations in devices are represented as current sources in the network. The interconnect lines are replaced with thermal RC transmission lines. T. Chiang *et al.* have shown that Spice-based simulation results demonstrate good agreement with Finite Element thermal simulations with ANSYS [76]. However, Spice-based simulation is computation intensive and often impractical for large circuit layouts.

8.2 Thermal Simulation with ANSYS

ANSYS, like any FEM simulator, requires geometric input defined by points, lines, areas, or cubes. IC layouts have numerous elements which make it very difficult to draw 3D geometries in ANSYS. Therefore, a program/script is developed to port a

Cadence circuit layout to ANSYS after adding z-dimensions as illustrated in figure 8-1. A user can specify substrate height as an input to the script. The z-dimensions are derived for $0.18um$ process technology using metal and via aspect ratios reported in the ITRS [2]. The script is written in perl, and named cds2inp. The instructions for using cds2inp and the script are included in Appendix B.

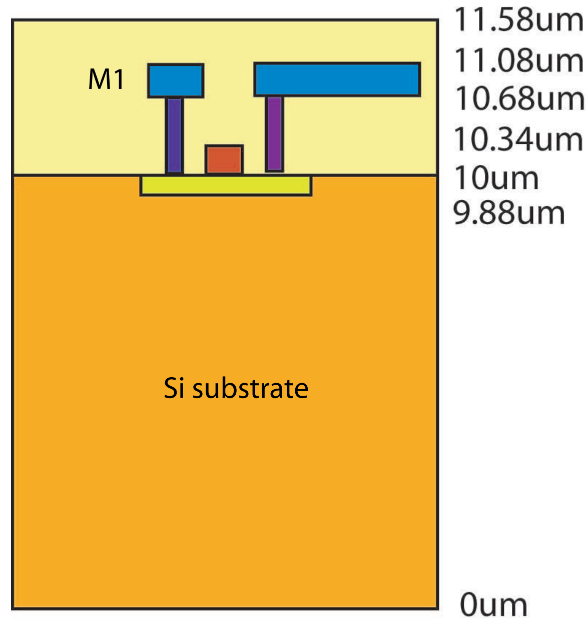

Figure 8-1: *Z-axis dimensions with a substrate height of* $10um$ *in cds2inp.*

In addition to a geometric model, ANSYS requires definitions of several material parameters. Table 8.2 lists the required material parameters used in this work. The parameter values are used from [77, 78]. A heat sink is attached to the substrate in conventional ICs. Therefore, the substrate is assumed to be the sole heat flow path to the ambient. Adiabatic boundary conditions are applied at the four sides and top of the chip, whereas the bottom of the substrate has either convective or isothermal boundary condition. This is a reasonable and widely used assumption as the chip is enclosed in a thermally insulated packaging material with the bottom surface attached to a heat sink.

A CMOS device from $0.18um$ process technology with width $W = 2um$, channel length $L = 0.2um$, and power dissipation $P = 6mW$ has been simulated in ANSYS.

Table 8.2: *Material parameters for thermal simulation in ANSYS.*

Material	Defined properties
ILD SiO_2	$k = 1.4e-6\ W/um-C$ $\rho_m = 2650e-18\ kg/um^3$ $c_p = 753.1\ J/kg-C$
Cu	$k = 390e-6\ W/um-C$ $\rho_m = 8960e-18\ kg/um^3$ $c_p = 385\ J/kg-C$
Bulk Si (for substrate)	$k = 120e-6\ W/um-C$ $\rho_m = 2330e-18\ kg/um^3$ $c_p = 705\ J/kg-C$
Silicide (for PC)	$k = 40e-6\ W/um-C$ $\rho_m = 2330e-18\ Kg/um^3$ $c_p = 753\ J/Kg-C$
Tungsten (W)	$k = 170e-6\ W/um-C$ $\rho_m = 19300e-18\ kg/um^3$ $c_p = 132\ J/kg-C$

Figure 8-2(a) shows the 3D geometry of the layout with diffusion contacts and metal routing in ANSYS. The substrate dimension is $20um \times 20um \times 10um$ with an isothermal (0^oC) boundary condition at the bottom. Adiabatic boundary conditions are applied to the rest of the surfaces. The maximum temperature in the transistor channel, referred to as the junction temperature, is recorded to be 17.3^oC from ANSYS simulation. Figure 8-2(b) and (c) illustrate the temperature profile at the horizontal and vertical cross-sections at the transistor plane, respectively. We see that the temperature profile at the top surface of the substrate (transistor plane) in figure 8-2(b) is axisymmetric despite asymmetric metal routing. This is because Si substrate has approximately 100 times better thermal conductivity than the ILD material on top.

Transistor junction temperature due to self heating is usually reported in terms of $^oC/mW/um$ (^oC per $1mW$ power dissipation per $1um$ width of the transistor) for a fixed channel length. The number is referred to as the thermal resistance of a device. Using the result from ANSYS, we compute the thermal resistance of the simulated device to be $5.76^oC/mW/um$ for $L = 0.2um$. Using reported data from [79], the thermal resistance of a bulk CMOS device with $L = 0.2um$ is calculated to be $4.28^oC/mW/um$ which is in close agreement with our simulation result.

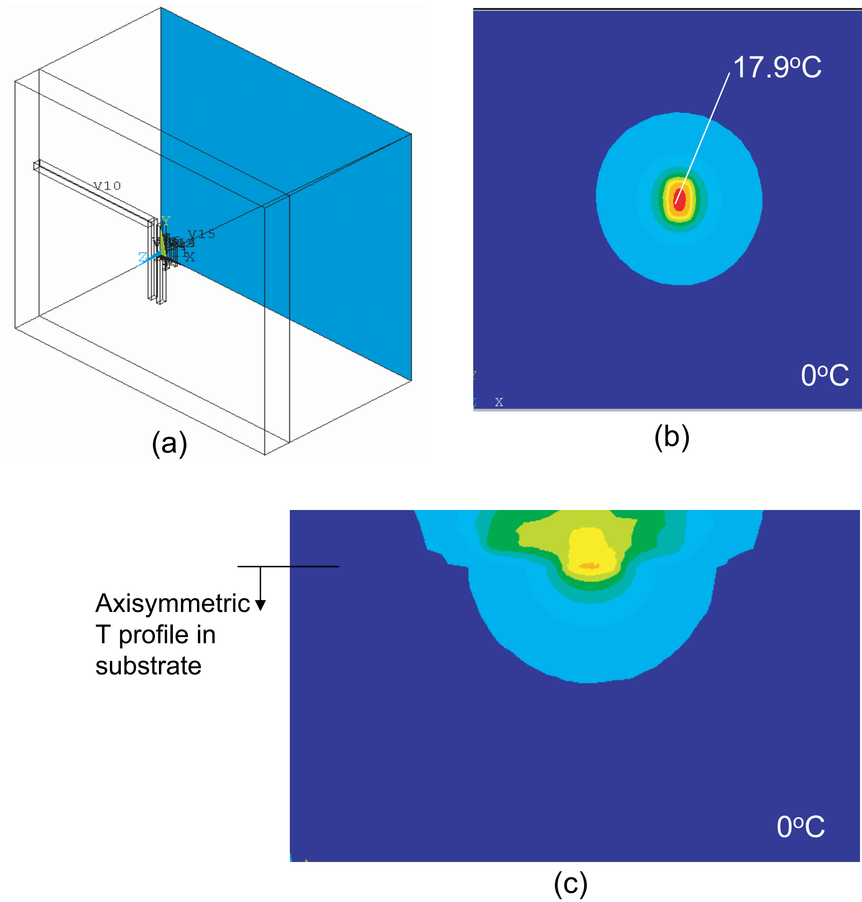

Figure 8-2: *ANSYS thermal simulation with a bulk CMOS transistor; (a) 3D geometry of the transistor with metal routing in ANSYS, temperature profile (b) at the top surface of the substrate (transistor plane), and (c) at the vertical cross-section of the IC.*

8.3 Transient Thermal Behavior in Integrated Circuits

Steady-state thermal analysis represents the temperature during static or DC power dissipation in a circuit. Assuming that the "off" current of a transistor is not high enough to cause self heating, the energy that contributes to a temperature rise is dissipated in digital circuits only when a transistor is switching a node. Figure 8-3 illustrates the power dissipation at the NMOS transistor of an inverter when the output node is switching at different frequencies.

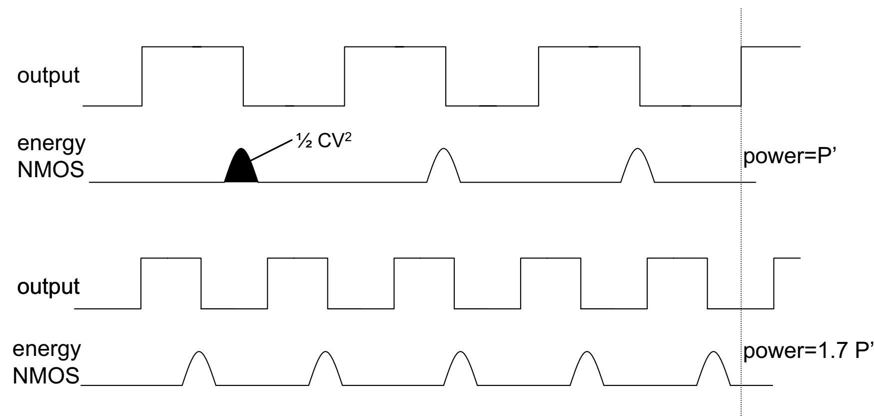

Figure 8-3: *Energy dissipation in the NMOS transistor of an inverter at different output frequencies.*

To simulate the transient thermal behavior of a transistor, a square wave is used to define volumetric heat generation rate over time in the channel. The area under the "on" period in the square wave would represent the amount of energy dissipation. During the "on" period of the square wave, the device temperature increases due to self heating, and conversely, the "off" period permits cooling. Transient thermal behavior can be modelled using a thermal time constant, τ, defined as the time required for the temperature of a device to increase or decrease by 36.78% ($1/e$). From a simple transient cooling simulation in ANSYS, we found the thermal time constant of the particular bulk CMOS device to be $17.55ns$ or approximately $20ns$ used for simplicity in later calculations. When the heat generation square wave in

ANSYS has low enough frequency such that there is enough time for cooling between two pulses, the transient behavior in the device temperature is prominent as shown in figure 8-4. Figure 8-4 illustrates the junction temperature of the device when the input square has a pulse width 20 times the thermal time constant.

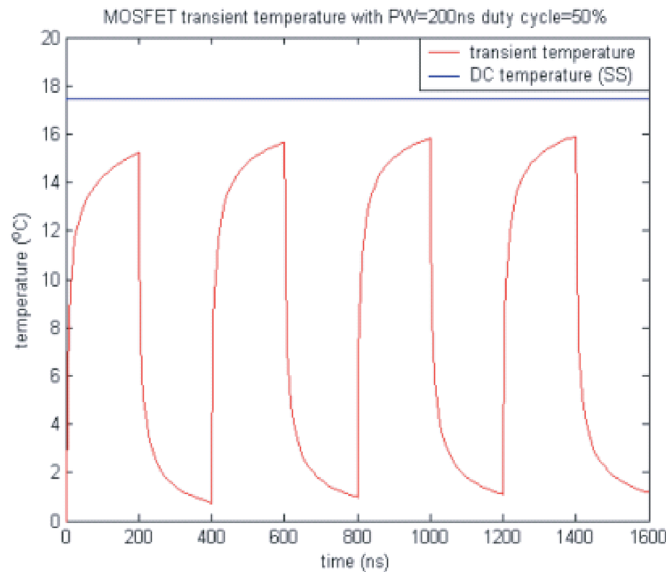

Figure 8-4: *MOSFET transient temperature plot at* $2.5MHz$ *square wave heat generation frequency corresponding to "on" pulse width*$= 10 \times \tau$. *Here,* $\tau = 20ns$.

As we increase frequency of the heat generation square wave in ANSYS, the time between heat generation pulses decreases while the representative energy dissipation (total area under "on" periods) per unit time, i.e. the average power, remains unchanged. When the pulse width is (1/40) of the thermal time constant, corresponding to $1GHz$ frequency of the input square wave, the device temperature reaches a steady state value over time as illustrated in figure 8-5. By changing the duty cycle of the heat generation square wave, we can represent different average power dissipations of the device corresponding to the change in operating frequency of a circuit (figure 8-3). Figure 8-5 shows the device temperature plots at the DC power, average power 1/5 of DC power, and 1/2 of DC power. With 20% duty cycle, the steady-state temperature

is $3.35 \pm 0.68\ ^oC$ that can be approximated by

$$T_{avg} = \frac{P_{avg}}{P_{DC}} T_{DC} = \frac{1}{5} \times 17.3 = 3.49^oC$$

Similarly, with 50% duty cycle,

$$T_{avg} = \frac{P_{avg}}{P_{DC}} T_{DC} = \frac{1}{2} \times 17.3 = 8.65^oC$$

whereas the simulation shows $8 \pm 0.89\ ^oC$.

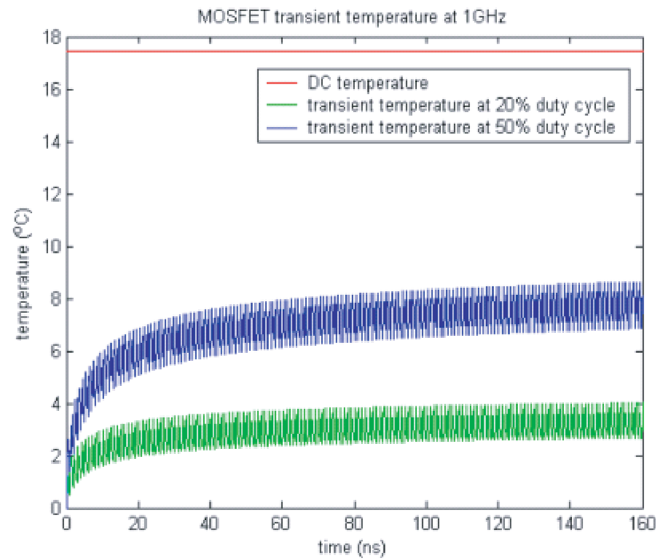

Figure 8-5: *MOSFET transient temperature plot at* $1GHz$ *square wave heat generation frequency corresponding to "on" pulse width=* $\tau/40$*. Here,* $\tau = 20ns$*.*

The above simulation results indicate that it is possible to accurately model transient temperature in digital circuits using temperature rise due to DC/static power. When the circuit is operating at a high enough frequency such that there is minimal cooling between energy dissipation periods in a device, the device temperature reaches a steady-state value proportional to the average power. For the simulated NMOS transistor with thermal time constant of $20ns$, an inverter would need to operate at $2GHz$ to maintain $\tau/40 = 0.5ns$ between two energy dissipation pulses as used in figure 8-5. If the thermal time constant of a device is higher, for example

$50-100\ ns$ reported for silicon-on-insulator devices [80, 81], the circuit operating frequency, for which we can accurately model average temperature from average power, would be even lower than $2GHz$. Thermal effects are often of major concern and need to be modelled for chips during high frequency operations. Therefore, the method described here for estimating the steady average temperature from average power dissipation can by applied to most high performance circuits with excellent accuracy.

8.4 Cell-based Thermal Profiling Technique

FEM simulation or any type of numerical approach for thermal analysis is only applicable to analysis with devices and small circuits. Due to computational cost and limitation on the number of nodes and memory usage in FEM simulation, ANSYS cannot be used for full 3D analysis with any circuit layout analyzed in SysRel. Therefore, a layout-level temperature profiling technique has been developed and implemented in SysRel. The proposed technique first computes temperature at the top of the substrate using the average power dissipation in a cell-based layout, and then, estimates the interconnect temperature using a conservative joule heating model.

Given the total power dissipation of a circuit, the uniform temperature at the bottom of the substrate is first computed using a simple thermal resistance based model for the heat sink. Figure 8-6 illustrates a conventional heat sink attached to the bottom of the substrate. The heat sink is replaced with an equivalent 1D thermal resistance, R_{hs}, connected between the ambient temperature, T_a, and substrate temperature, T_{sub}. The thermal resistance of thermal interface material, used to attach the heat sink to the chip, can also be placed in series. As reported in [82], an R_{hs} of $0.26^oC/W$ is used in this work for a fin heat sink with air volume flow of $0.006m^3/s$. T_{sub} is calculated using equation 8.4 where Q is the total power dissipation of the chip.

The uniform substrate temperature at the bottom can now be applied as the isothermal boundary condition to compute the temperature at the transistor plane on top of the substrate. Based on our observations in ANSYS thermal simulations,

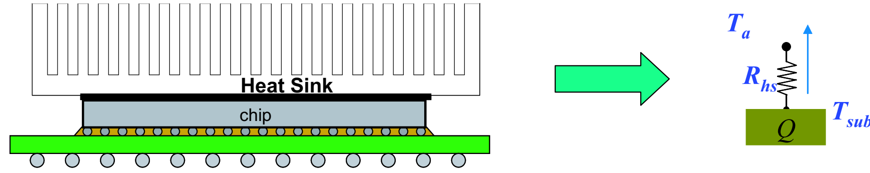

Figure 8-6: *Package-level thermal modeling to compute uniform temperature at the bottom surface of a substrate.*

the temperature profile due to a power source at the top surface of a substrate is axisymmetric and linearly dependent on power. Therefore, temperature can be treated as a linear and spatially independent variable. We define an impulse response as the temperature rise at the top of the substrate (transistor plane) due to $1mW$ of power dissipation in a unit area ($um \times um$). The impulse response has been constructed from ANSYS as shown in figure 8-7. Given a power density profile in a circuit layout, in terms of mW/um^2, we can compute the temperature distribution at the top surface of the substrate by convolving the impulse response with the power density profile. Convolution is an $\mathcal{O}(n^2)$ computational process. Therefore, frequency domain calculation using Fast Fourier Transform (an $\mathcal{O}(n \log_2 n)$ process) can be done to reduce computational load in large circuit layouts.

A cell-based approach, compatible with existing IC design flows, has been adopted to compute the power density profile in a layout. During a large circuit design, cell libraries are typically used where the cells are already characterized for power. Moreover, automated design flows with behavioral level description to synthesis, such as the one used for designing circuits in this work, explicitly report cell-level power dissipations using switching activity factors at the nets. The non-uniformity in cell-level power dissipations allows us to capture the non-uniform power distribution in a circuit layout. To compute the power density profile in mW/um^2, the average cell power dissipation is evenly distributed within the cell's boundary. Ideally, power dissipation or equivalently heat generation needs to be concentrated into the transistor channels to capture device self-heating accurately. However, for the sole purpose of modeling interconnect temperature, capturing uniform average temperature in a

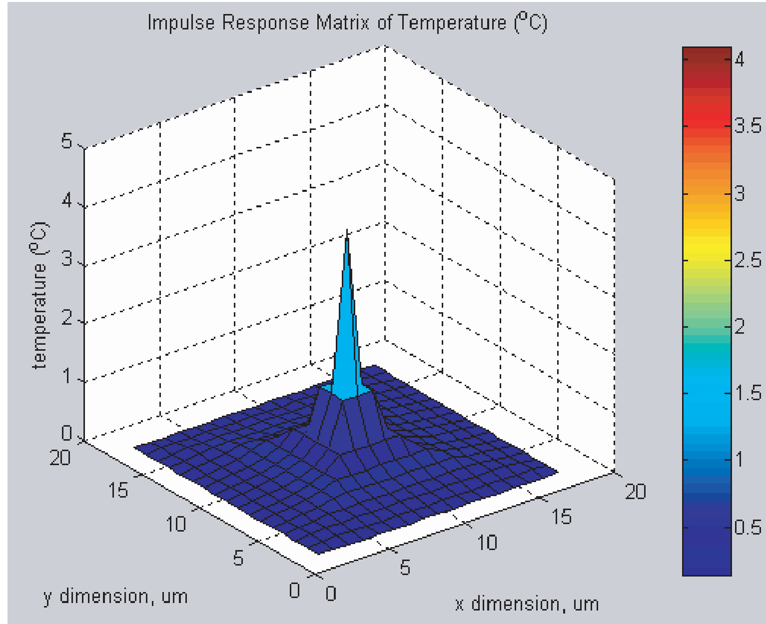

Figure 8-7: *Thermal impulse response at the top surface of a substrate.*

cell is computationally efficient. Cell-level uniform temperature assignment has been traditionally used for fast thermal analysis and temperature driven cell placement [83].

The cell-based temperature estimation technique is implemented in a Matlab program TProfile. The source code for TProfile is included in Appendix C. TProfile accepts the temperature impulse response and power density matrix as its inputs. The program uses a 2048-pt FFT and inverse-FFT for computing the temperature profile matrix. The power density matrix of a single inverter cell ($2um \times 8um$) with a power consumption of $12mW$ is used to compare TProfile's temperature estimates with that from ANSYS simulation. Figure 8-8 and 8-9 show the power density matrix and corresponding temperature profile, respectively, in the single inverter cell in a $30um \times 30um$ substrate. The average cell temperatures from TProfile and ANSYS simulation are 36.3^oC and 39.7^oC, respectively. The difference is less than 10%.

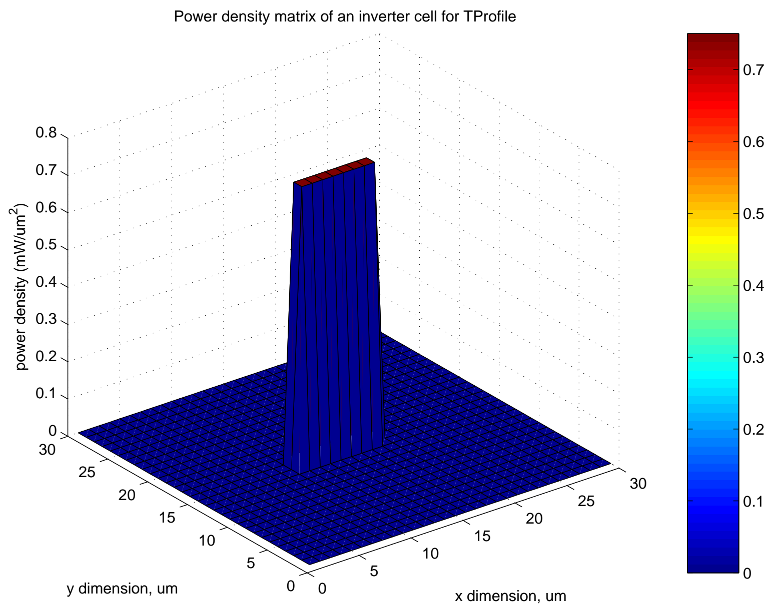

Figure 8-8: *Power density profile of an inverter cell in a* $30um \times 30um$ *substrate.*

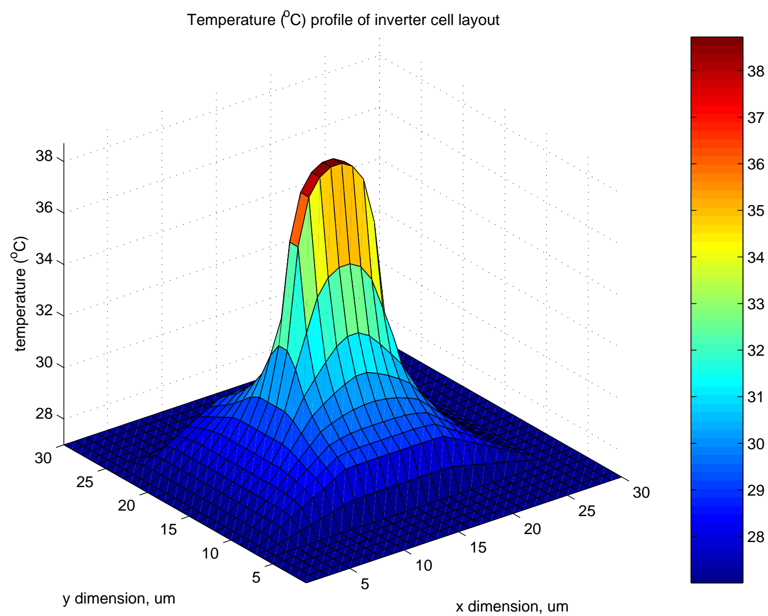

Figure 8-9: *Temperature profile calculated using TProfile of an inverter cell in a* $30um \times 30um$ *substrate.*

8.5 Interconnect Joule Heating

Interconnect joule heating causes the temperature rise due to current flow through the metal. As technology scaling progresses, interconnect joule heating effects are becoming significant due to the increase in current density and number of wiring levels [84]. In [76], temperature distribution has been modelled in a straight line interconnect in the first metal level with two end vias connecting it to the substrate (figure 8-10). Assuming that the substrate temperature is T_o, via temperature is also

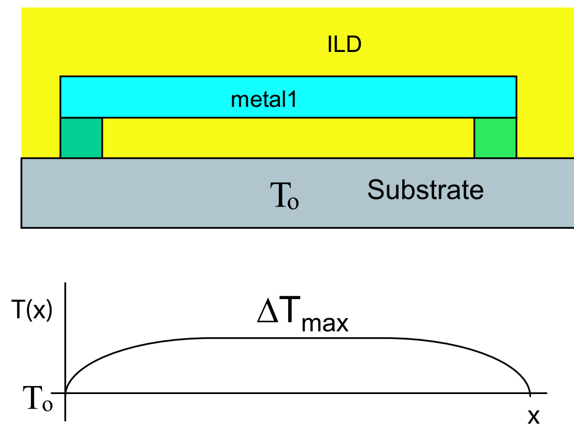

Figure 8-10: *Modeling the temperature profile in a straight line interconnect connected to a substrate at temperature T_o.*

T_o due to higher thermal conductively of the vias. Ignoring the effect of variation in metal resistivity, ρ, with temperature, we can write the diffusion equation as

$$\frac{d^2T}{dx^2} - \frac{T - T_o}{L_H^2} = -\frac{\rho j_{rms}^2}{k_M} \tag{8.18}$$

and

$$L_H = \left[\frac{k_M H t_{ILD}}{k_{ILD}}\left(\frac{1}{s}\right)\right]^{\frac{1}{2}} \tag{8.19}$$

where L_H is the thermal characteristic length, k_M and H are the thermal conductivity and height of the metal line, and k_{ILD} and t_{ILD} are the thermal conductivity and thickness of the ILD material underneath. The heat spreading factor, s, is used to correct the deviation from 1D heat flow between the metal line and the underlying

layer. $s = 1$ ignores any heat spreading effect and accounts for the worst case temperature. It has been shown that within the range of thermal characteristic length, L_H, heat can flow through the via. Beyond L_H, heat flows through the ILD causing the maximum temperature rise, $\triangle T_{\max}$, at the metal line. $\triangle T_{\max}$ has been derived in [76] as following.

$$\triangle T_{\max} = \frac{j^2 \rho L_H^2}{k_M} \tag{8.20}$$

$$= \frac{j^2 \rho H t_{ILD}}{k_{ILD} s} \tag{8.21}$$

$$= \frac{j^2 \rho H t_{ILD}}{k_{ILD}} \tag{8.22}$$

Using equation 8.22, we have computed $\triangle T_{\max}$ in global interconnect lines in different technology generations as illustrated in table 8.3. The technology generation parameters are as reported in the ITRS 2003 edition [2]. Interconnect height, H, and ILD thickness, t_{ILD}, are computed using the minimum global wiring pitch and aspect ratio of Cu lines and vias. $\triangle T_{\max}$ values are computed with two cases of ILD thermal conductivity, k_{ILD}; one with projected k_{ILD} values used in [84] associated with low-k materials, and the other where a fixed k_{ILD} value for SiO_2 is used in all technology generations. The computed $\triangle T_{\max}$ values in the table suggest that interconnect joule heating is not significant in near term technology generations; however, it can be severe in future technology generations particularly with low-k ILD materials.

Multi-level complex shaped interconnects exist in actual circuit layouts. In SysRel, we model the worst case joule heating at a via while applying the default model to compute its lifetime. Figure 8-11 illustrates the method for computing via temperature at different metal levels. The non-uniform substrate temperature is first computed using the cell-based approach described in the previous section. A via connected to the substrate is modelled to be at the same temperature. A higher metal level via is at a temperature above the substrate temperature by multiples of $\triangle T_{\max}$.

Table 8.3: *Maximum temperature increase due to interconnect joule heating in different technology generations.*

Technology Node	$90nm$	$65nm$	$45nm$	$32nm$	$22nm$
j_{max} (MA/cm^2)	0.5	1	3	4.3	5.8
H (um)	0.4305	0.3190	0.2357	0.1680	0.1250
t_{ILD} (um)	0.3895	0.2900	0.2153	0.1540	0.1150
ρ $(\mu\Omega - cm)$	2.2	2.2	2.2	2.2	2.2
k_{ILD} (W/cm^oC) (projected)	$1.4e-2$	$1.02e-2$	$0.54e-2$	$0.19e-2$	$0.12e-2$
$\triangle T_{\max}$ (^oC)	0.0659	0.1995	1.8607	5.5391	8.8655
k_{ILD} (W/cm^oC) (SiO_2)	$1.4e-2$	$1.4e-2$	$1.4e-2$	$1.4e-2$	$1.4e-2$
$\triangle T_{\max}$ (^oC)	0.0659	0.1454	0.7177	0.7517	0.7599

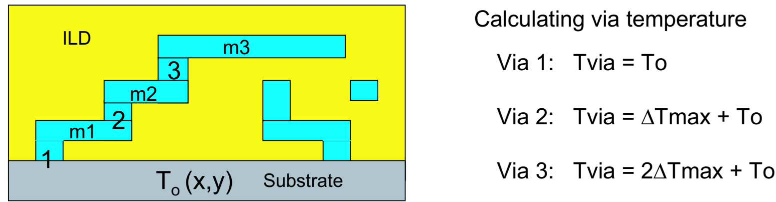

Figure 8-11: *Modeling the temperature rise due to joule heating in multi-level interconnects.*

8.6 Thermo-file Implementation in SysRel

SysRel has a menu bar, named Thermo-file, that adds thermal analysis capabilities and a selection tab to disable or enable the thermal-aware analysis using the estimated temperature profile. Figure 8-12 shows the menu options available in Thermo-file. Before invoking Thermo-file menu operations, a user needs to open a flattened or hierarchical layout and read the *sysrel.emparam* file using the "File" menu items in SysRel. The "Read Hierarchical Magic file..." menu item invokes a file chooser dialog box to allow a user to select the corresponding hierarchical Magic layout file. A Thermo-file layout window appears with a cell-level view of the layout. The "Read Power Distribution (.cpower) File..." menu item invokes a file chooser dialog box to select a cell-level power report file with the extension *.cpower*. The *.cpower* file is an ascii text file. Cell-level power dissipation can be computed using either Synopsys

Design Analyzer or Nanosim. The *.cpower* file, for circuit layouts used in this work, is constructed using cell-level power analysis with Synopsys Design Analyzer.

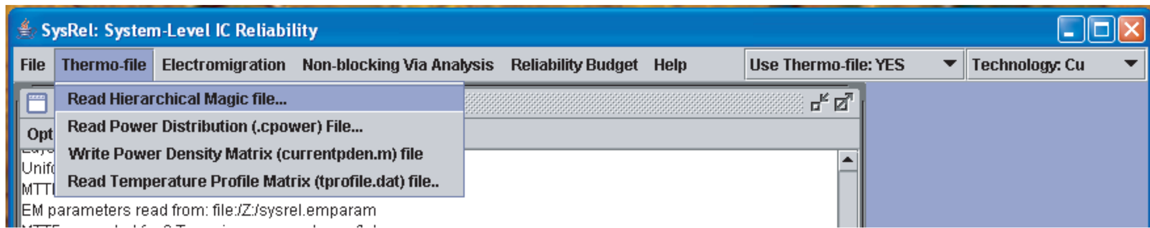

Figure 8-12: *Thermo-file menu options in SysRel for thermal-aware reliability analysis.*

The "Write Power Density Matrix (currentpden.m) File" menu item creates a power density map for the cell-based layout using the methodology discussed in section 8.4. The *currentpden.m* file is written in user's current directory. This file serves as an input to TProfile while calculating the layout-level temperature profile using the temperature impulse method. TProfile writes an output file, *tprofile.dat*, that can be ported to SysRel using the "Read Temperature Profile Matrix (tprofile.dat) file.." menu item. When the temperature profile is read, the Thermo-file window updates the cell-based layout with a color plot of its temperature. The selection tab "Use Thermo-file" in SysRel consequently becomes active. A user can either select to incorporate temperature profile and interconnect joule heating into electromigration lifetime calculations with "Use Thermo-file: YES" or just use a fixed temperature as defined in *sysrel.emparam* file with "Use Thermo-file: NO" option. The reliability analysis follow continues using the "Electromigration" menu.

8.7 Thermal-Aware Reliability Simulation with 32-bit Comparator

The 32-bit comparator circuit described in section 5.6 is used for thermal-aware reliability analysis in SysRel. The power density matrix is generated from SysRel and fed into TProfile to compute the layout-level non-uniform temperature profile as shown

in figure 8-13. With an average power dissipation of $26.6mW$, the mortal interconnect temperature is only 27.078^oC. The lifetimes of mortal trees are significantly higher due to such a low temperature rise (table 8.4). Full-chip probability of no failure within a target lifetime of 10 years is 1. The thermal-aware reliability simulation results suggest that full-chip reliability can be pessimistic using a worst case interconnect temperature leading to unnecessarily conservative designs.

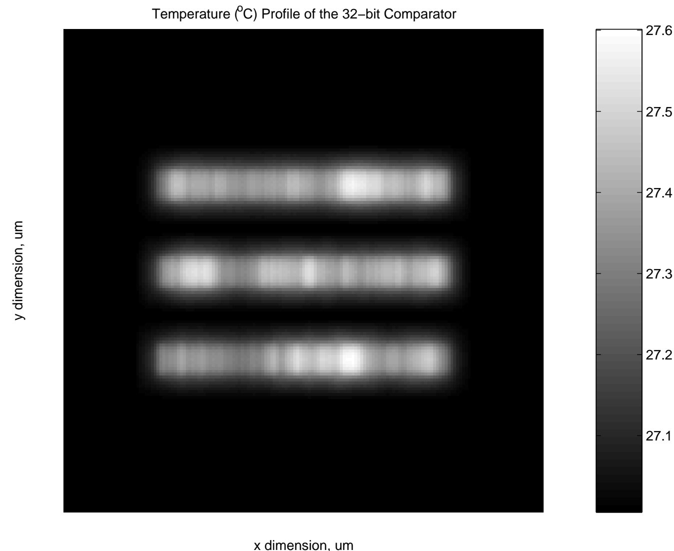

Figure 8-13: *Layout-level non-uniform temperature profile in the* 32-*bit comparator circuit layout.*

8.8 Summary

Interconnect lifetime due to electromigration is exponentially dependent on the inverse temperature. Using a worst case temperature for reliability analysis, and then setting a reliability goal may lead to excessive conservatism in IC design. Therefore, accurate thermal analysis is a prerequisite to circuit-level reliability analysis and reliability-aware IC design. We have conducted ANSYS thermal simulations with a bulk CMOS device with metal routing and diffusion contacts. The junction temperature due to self-heating is in close agreement with reports in other literature. We have also conducted transient thermal simulations with the device. It is possible to accurately

Table 8.4: *Thermal-aware reliability simulation results with Cu metallization in the 32-bit comparator circuit layout.*

Step 1	Layout Extraction (total # of interconnect trees=1143)	
Step 2	Via-based ($j_{\max}L$) filter	
	Number of immortal trees identified	1023
Step 3	Via-based ($j_{local}L$) filter	
	Number of immortal trees identified	112
Step 4	Default Model (with $\sigma_{nucl} = 40MPa$) on 8 mortal trees	
	t_{50} of 4 mortal tees in metal1 (via-above)	17484 years
	t_{50} of 4 mortal tees in metal2 (via-below)	103467 years
Step 5	Full chip stochastic analysis ($\sigma = 0.81$, lognormal)	
	Target chip lifetime	10 years
	Probability of no failure	1
	max FIT	$8e-15$
	t_{50} for full chip	$\gg$ 10 years

model the transient temperature in digital circuits using the temperature rise due to DC/static power. When the circuit is operating at a high enough frequency such that there is minimal cooling between energy dissipation periods in a device, the device temperature reaches a steady-state value proportional to the average power.

Based on our observations from ANSYS thermal simulations, temperature profile due to a power source at the top surface of a substrate is axisymmetric and linearly dependent on power. Treating temperature as a linear and spatially independent variable, we define an impulse response as the temperature rise at the top surface of a substrate (transistor plane) due to $1mW$ of power dissipation in a unit area ($um \times um$). Thus it is possible to compute the temperature distribution at the top surface of the substrate by convolving the impulse response with the power density profile of a layout. A Matlab program TProfile has been developed to compute temperature profile using the proposed method in frequency domain. TProfile estimates the average temperature of a cell within 10% of that from ANSYS simulation. Using the layout-level temperature profile as a boundary condition, SysRel computes the interconnect temperature using a conservative joule heating model. The 32-bit comparator circuit has been reanalyzed with thermal capabilities. While reliability simulation with the worst case interconnect temperature of 105^oC predicts full-chip

probability of no failure to be 0.524 for Cu metallization technology, thermal-aware reliability simulation indicates no failure within the chip's target lifetime.

Chapter 9

Reliability Analysis with Arithmetic and Logic Unit

SysRel reliability simulations discussed in earlier chapters are of rather small adder and comparator circuits. To demonstrate SysRel's capability and to investigate reliability behavior in large circuit layouts, a 64-bit Arithmetic and Logic Unit (ALU) has been designed and analyzed for full-chip reliability. The ALU circuit has been used to investigate the relative reliability with Al and Cu metallization technologies and to predict future reliability direction with low-k Cu interconnect technology. This chapter describes the design and analysis work with the 64-bit ALU.

9.1 Arithmetic and Logic Unit Design and Synthesis

An ALU is the computational core of a processor. A typical ALU performs basic arithmetic functions, such as addition, multiplication, and comparison, as well as basic logic operations, such as AND, OR, and EXOR. The data in an ALU is arranged as a collection of bits. For instance, a 32-bit ALU operates on 32 bits wide inputs, and produces a 32 bits wide output. 32-bit datapath is widely used in modern day processors, although 64-bit datapath is available for very high performance applications.

A 64-bit ALU has been designed and synthesized using a behavioral-level description to synthesis flow. The input instruction is 132-bit wide where the four most significant bits are the opcode[1] followed by two 64-bit operands. The input operands, A and B, and output, C, are signed integer numbers. The behavioral-level description of the ALU is written in VHDL modelled after Tiny64, a simple configurable microprocessor [85]. Table 9.1 lists the opcodes and their corresponding operations.

Table 9.1: *Operations in the* 64*-bit Arithmetic and Logic Unit.*

Opcode	Instruction	Description
0000	mov	Move B to C.
0001	and	C is A and B.
0010	or	C is A or B.
0011	xor	C is A xor B.
0100	add	C is $A + B$.
0101	sub	C is $A - B$.
0110	ror	Rotate right A by 1 bit and store in C.
0111	lsr	Right shift A by 1 bit, pad with 0, and store in C.
1000	lsra	Right shift A by 1 bit, keep sign unchanged, and store in C.
1001	swap	Swap most and least significant 32 bits in A and store in C.
1010	swapb	Swap most and least significant 32 bits in B and store in C.
1011	inc	C is $A + 1$.
1100	dec	C is $A - 1$.
1101	rorb	Rotate right A by 8 bits and store in C.
1111	mul	C is $A(31 - 0) \times B(31 - 0)$.

Using the VHDL description of the ALU, a Verilog gate-level netlist is synthesized in Synopsys Design Analyzer with the TSMC 0.18um cell library. The Verilog netlist is then placed and routed to create the layout in Silicon Ensemble. Silicon Ensemble outputs layout in gds2 format which is converted to a Magic format using Magic layout editor. The 64-bit ALU layout's dimension is 1035um $\times$ 1036um (figure 9-1). Up to five metal levels are used for interconnect routing. The layout has 4193 cells (including filler cells). Table 9.2 shows the number of cells of various types in the layout.

[1] An opcode is a sequence of bits to indicate designated operations in ALU.

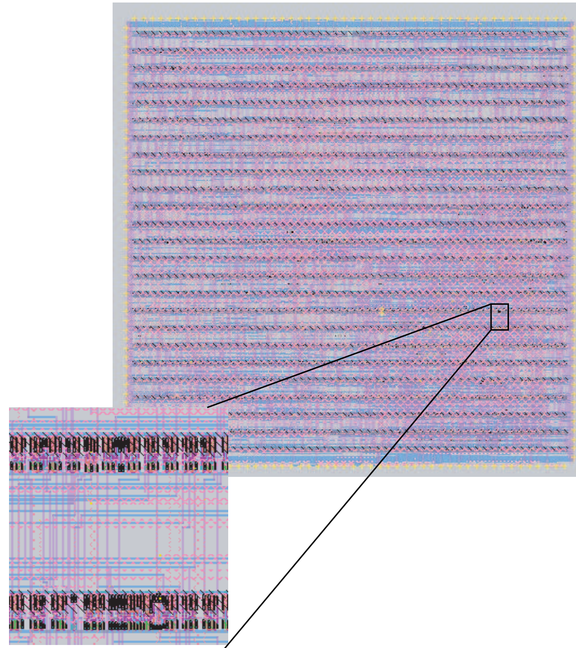

Figure 9-1: *The 64-bit Arithmetic and Logic Unit layout in Magic.*

Table 9.2: *The number of cells of various types in the 64-bit ALU circuit.*

Cell Name	Number	Cell type	Number
XOR2X1	147	XNOR2X1	4
OAI21X1	32	NAND3X1	2
AOI22X1	136	AND2X1	67
INVX1	447	OR2X1	3
AOI21X1	14	BUFX2	75
NAND2X1	110	FAX1	994
NOR2X1	1133	MUX2X1	832
INVX2	196	AND2X2	1

9.2 Reliability Simulation with Different Metallization Technologies

The 64-bit ALU circuit is analyzed in SysRel using the cell-based hierarchical layout from Magic. Total power dissipation of $2.129W$ is estimated using Design Analyzer during synthesis. The maximum current density in the power delivery lines is approximately $0.46MA/cm^2$. The worst case interconnect temperature of 105^oC and full-chip target lifetime of 20 years are used for following simulations with the 64-bit ALU circuit. The number of tiles representing different mask layers in the layout is 807124. Therefore, layout drawing is switched off to minimize simulation time. SysRel reports 89176 interconnect trees in the layout initially outlined in red (figure 9-2(a)). In Cu/SiO_2 based interconnect technology, 81619 trees are first identified to be 'immortal' using the $(j_{max}L)$ filter (figure 9-2(b)). Another 6415 trees are filtered out using the $(j_{local}L)$ filter. As illustrated in figure 9-2(c), only 1142 trees are reported to be mortal in Cu/SiO_2 based interconnect technology.

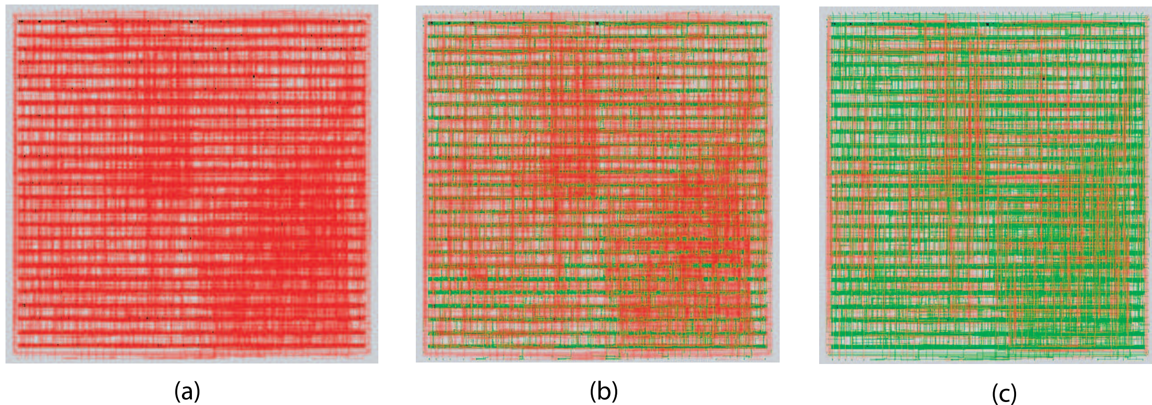

Figure 9-2: *Immortality in interconnect trees in the* 64*-bit ALU circuit with Cu metallization. (a)* 89176 *trees are extracted from the layout. (b)* 81619 *trees are identified as immortal and outlined in green after* $(j_{max}L)$ *filter. (c)* 6415 *additional trees are identified as immortal after* $(j_{local}L)$ *filter.* 1142 *mortal trees are outlined in red.*

The 64-bit ALU circuit has been analyzed for full-chip reliability with various interconnect technologies. The material parameters used here for diffusivity and lifetime calculations in Al bamboo, Al polygranular, and Cu/SiO_2 interconnects are

as reported in tables 4.3, 4.4, 4.5, and 4.6 in Chapter 4. Table 9.3 lists the intermediate results from immortal tree filtration as well as full-chip reliability in various metrics. Due to higher (jL) product thresholds in Al metallization, only 179 trees are identified as mortal. Consistent with our previous analysis, full-chip reliability is the best with Al bamboo type interconnects and worst with Al polygranular type interconnects. On the other hand, lower (jL) product thresholds in Cu/SiO_2 based interconnects identify 1142 trees to be immortal and predict full-chip reliability to be worse than that with Al bamboo type interconnects (row 3 in table 9.3).

Table 9.3: *Full-chip reliability analysis with the* 64*-bit ALU in different metallization technologies. Here* $(jL)_{M1}$ *and* $(jL)_{M2}$ *correspond to* $(jL)_{via-above}$ *and* $(jL)_{via-below}$*, respectively.*

	# of trees = 89176	Critical parameters	# of imm. trees ($j_{\max}L$)	# of imm. trees ($j_{local}L$)	# of mortal trees	Full-chip reliability		
						Prob. of no fail.	Max FIT	t_{50} (years)
1	Al bam	$(jL)_{M1}$ = $(jL)_{M2}$ = $4000A/cm$	84724	4273	179	0.994	153	84.1
2	Al poly	$\sigma_{nuc} = 500MPa$, $B = 50GPa$	84724	4273	179	$8.68E-16$	$272k$	1.07
3	Cu/SiO_2	$(jL)_{M1}$ = $1500A/cm$, $(jL)_{M2}$ = $3700A/cm$, σ_{nuc} = $40MPa$, $B = 28GPa$	81619	6415	1142	0.048	25623	8.51
4	Cu $(jL)_1$	$(jL)_{M1}$ = $(jL)_{M2}$ = $1500A/cm$, σ_{nuc} = $40MPa$, $B = 28GPa$	78381	9653	1142	0.048	25623	8.51
5	Cu $(jL)_2$	$(jL)_{M1}$ = $(jL)_{M2}$ = $700A/cm$, σ_{nuc} = $40MPa$, $B = 28GPa$	67941	18496	2739	0.044	27809	8.50
6	Cu/low-k	$(jL)_{M1}$ = $(jL)_{M2}$ = $375A/cm$, σ_{nuc} = $12.5MPa$, $B = 10GPa$	55134	28710	5332	$3.42E-25$	$839k$	2.33

To decrease interconnect delay, a wide range of materials with a low dielectric constant, κ, generally referred to as low-k dielectrics, are being investigated as replacements for SiO_2. Lower (jL) product thresholds are observed for such ILD materials [86]. Therefore, we have conducted (jL) product sensitivity analysis on the full-chip reliability of the 64-bit ALU using lower (jL) product thresholds as indicated

in rows 4 and 5 in table 9.3. The number of mortal trees increases with lower (jL) product thresholds which in turn degrades the full-chip reliability.

Hau-Riege *et al.* reported electromigration experimental work on straight line via-to-via Cu interconnects with a particular low-k dielectric material [87]. They experimentally reported the upper limit of the (jL) product threshold to be $375A/cm$ and the corresponding critical stress for void nucleation $\sigma_{crit_nuc} \leq 12.5MPa$ when initial stress in the line is zero. In addition, the decrease in dielectric constant corresponds to a decrease in the elastic moduli of the dielectric material which in turn reduces the effective bulk modulus, B, of the metal-dielectric system. $B = 10GPa$ is reported for the Cu/low-k interconnects in [87]. Using the Cu/low-k interconnect system in the 64-bit ALU layout, we have analyzed the impact on full-chip reliability as illustrated in row 6 in table 9.3. The full-chip reliability is the worst among all our simulation results. Following are the two contributing factors.

- The (jL) product threshold of $375MPa$ leads to 5332 mortal trees in the 64-bit ALU layout with the Cu/low-k interconnects. The number of mortal trees is approximately 30 times and 5 times more than that with Al and Cu/SiO_2 interconnects, respectively.

- According to equation 2.5, time to void nucleation, $t_{nucl} \propto (sigma^2_{crit_nuc}/B)$. A comparatively lower $(sigma^2_{crit_nuc}/B)$ factor, as computed using the experimentally reported data, leads to shorter time to void nucleation in mortal trees in Cu/low-k interconnects. Table 9.4 lists the t_{nucl} and t_{growth} values using equations 2.5 and 2.7 of the default model for different interconnect technologies. In void nucleation limited failures, for example, a small void blocking electron flow in Cu via-above type interconnect trees, the lifetime of a mortal tree in Cu/low-k is affected most.

During SysRel simulations, it is possible to vary any material parameters and investigate the impact on full-chip reliability. The reverse is also possible and particularly useful in predicting material and process parameters required to achieve desired full-chip reliability. We have conducted similar studies with the 64-bit ALU circuit

Table 9.4: *Times to void nucleation and void growth in various interconnect technologies. Here* $j = 0.5MA/cm^2$ *and* $T = 105^oC$.

Interconnect technology	t_{nucl} (years)	t_{growth} (years)
Al bamboo	164.44	58.99
Al polygranular ($w = 4um$)	2.15	0.77
Cu/SiO_2	16.19	121.39
Cu/low-k	4.42	121.39

to investigate the required Cu atomic diffusivity, a parameter that can be controlled during process development. The full-chip reliability goal is set to the best case probability of 0.994 of no failure within the target lifetime achieved with Al bamboo type interconnects. The atomic diffusivity of Cu is varied during simulations using the prescaler, D_o, in equation 4.4. The lower the atomic diffusivity, the better the lifetime. Initially the diffusivity ratio D_{Al_bamboo}/D_{Cu} is 0.35. As illustrated in table 9.5, Cu diffusivity needs to be reduced by approximately 12 times to achieve the equivalent reliability with Cu/SiO_2 based interconnects. With Cu/low-k interconnects, the required improvement in atomic diffusivity is even higher. Approximately 44 times decrease in Cu atomic diffusivity is required with Cu/low-k interconnects.

Table 9.5: *Cu atomic diffusivity requirements for equivalent circuit-level reliability in the 64-bit ALU with Al and Cu metallizations.*

# of trees = 89176	Critical parameters	# of mortal trees	Prob. of no failure	Desired prob. of no fail.	Required diff. ratio D_{Al_bam}/D_{Cu}	Required decrease in D_{Cu}
Al poly	$(jL)_{M1}$ = $(jL)_{M2}$ = $4000A/cm$	179	0.994	0.994	0.3495	
Cu/SiO_2	$(jL)_{M1}$ = $1500A/cm$, $(jL)_{M2}$ = $3700A/cm$, $\sigma_{nuc} = 40MPa$, $B = 28GPa$	1142	0.048	0.995	4.3214	12.36×
Cu/low-k	$(jL)_{M1}$ = $(jL)_{M2}$ = $375A/cm$, σ_{nuc} = $12.5MPa$, $B = 10GPa$	5332	$3.42E-25$	0.995	15.41	44.10×

Electromigration tests are usually done with simple interconnect structures. Test-level reliability can be engineered using different fabrication processes and materials,

for example using a different capping layer other than Si_3N_4 on top of Cu [88], that impact critical parameters such as atomic diffusivity and effective modulus. Data in table 9.5 conveys a profound message on the required test-level reliability target for Cu metallization technology. Significantly improved test-level reliability in Cu is required to achieve equivalent circuit-level reliability with Al bamboo type interconnects. Moreover, the required improvement will increase as low-k/low-modulus dielectrics are introduced.

9.3 Thermal-Aware Reliability Analysis with the 64-bit ALU

Cell-level power dissipations in the 64-bit ALU are calculated in Design Analyzer during gate-level netlist synthesis. The default input toggle rate in Design Analyzer is 0.5 in $1ns$ which corresponds to an operating frequency of $250MHz$. Total power dissipation is reported to be $2.129W$ when $Vdd = 5V$. Using the cell-level power dissipations, the layout-level nonuniform power density matrix of the 64-bit ALU is first computed in SysRel. Using the power density matrix in conjunction with a temperature impulse response, the nonuniform temperature profile is calculated in TProfile. Figure 9-3 illustrates the temperature profile in the 64-bit ALU in bulk CMOS technology. There are few hot spots where the temperature rise is as high as 4.5^oC above the ambient.

The average dynamic power consumed in a digital circuit operating at a frequency f is expressed by the well known relation

$$P = \alpha_{0\rightarrow 1} \times C_{load} \times Vdd^2 \times f \tag{9.1}$$

where C_{load} is the capacitive load, Vdd is the operating voltage, and $\alpha_{0\rightarrow 1}$ the probability that a power consuming transition occurs at the output [89]. Assuming that the leakage power component does not contribute to significant temperature rise and the dynamic power consumption is dominant, we can estimate power density matrices in

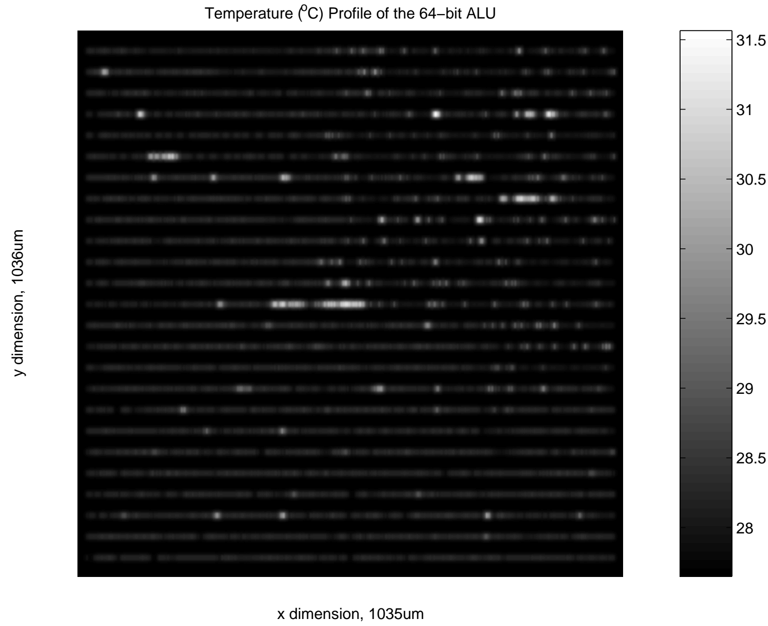

Figure 9-3: *Nonuniform temperature profile of the* 64*-bit ALU circuit in bulk CMOS technology at* $Vdd = 5V$ *and operating frequency*$= 250MHz$.

the 64-bit ALU at higher operating frequencies using the linear relation depicted in equation 9.1. The power density matrices can then be used to compute layout-level temperature profiles at those frequencies. It is important to note that the ALU circuit may need to be pipelined to operate at higher frequencies and associated circuitry will add to the total power consumption.

Thermal-aware full-chip reliability simulation of the 64-bit ALU circuit has been carried out in SysRel using Cu/SiO_2 based interconnects in the layout. The simulations are at various circuit operating frequencies in both bulk CMOS and silicon-on-insulator (SOI) technologies. The temperature impulse response for bulk CMOS technology has been calculated from ANSYS simulation results as discussed in Chapter 8. The thermal resistance of a bulk CMOS device is estimated to be $5.76^oC/mW/um$. Using the thermal resistance of an SOI device of $50^oC/mW/um$ as reported in [78, 90], we would expect the junction temperature rise in SOI technology to be approximately nine times of that in bulk CMOS technology. Therefore, the temperature impulse response is scaled accordingly for estimating temperature profile of the 64-bit ALU in

SOI technology. In addition, circuits in SOI technology can operate at a faster frequency than those in comparable bulk CMOS technology due to lower capacitance in SOI devices [91]. As reported in [91], a 64-bit PowerPC in SOI technology has a maximum operating frequency, f_{MAX}, that is approximately 1.24 times the f_{MAX} in bulk CMOS technology. Therefore, we scale the power dissipation in SOI technology by a factor of 1/1.24 while estimating the power dissipation of the 64-bit ALU circuit operating at the same frequency as in bulk CMOS technology. Using the estimated power dissipation and scaled temperature impulse response for SOI technology, we compute the temperature profile of the 64-bit ALU circuit for thermal-aware reliability analysis.

Tables 9.6 and 9.7 list the thermal-aware full-chip reliability results in bulk CMOS and SOI technology, respectively, with Cu/SiO_2 interconnects in the 64-bit ALU. $T_{\max}$ is the maximum temperature at a hot-spot on the substrate. T_{avg} is the average temperature at the top surface of the substrate. The $T_{\max}$ and T_{avg} values are stated here for record. The nonuniform layout-level temperature profile computed for each operating frequency is used during full-chip reliability analysis with SysRel. As indicated

Table 9.6: *Thermal-aware full-chip reliability report for the* 64*-bit ALU with* Cu/SiO_2 *metallization in bulk CMOS technology. Here the target lifetime of the chip is* 20 *years.*

Operating Frequency	Bulk CMOS Technology			
	$T_{\max}$ (^oC)	T_{avg} (^oC)	Full-chip prob. of no failure	Max FIT
$250MHz$	31.56	27.84	≈ 1	$1.01E-10$
$1GHz$	45.26	30.36	≈ 1	$7.56E-10$
$2GHz$	63.51	33.72	0.9999	$9.6E-9$
$3GHz$	81.77	37.07	0.9999	$1.41E-7$
$4GHz$	100.0	40.43	0.9999	$2.95E-4$

in table 9.6, the circuit is highly reliable at low operating frequencies in bulk CMOS technologies. We notice a temperature rise significant enough to degrade reliability as indicated by a lower Max FIT only at frequencies such as 3 and 4 GHz. On the other hand, an increased temperature rise in SOI technology significantly affects full-chip reliability at frequencies as low as $1GHz$ (table 9.7). The maximum hot-spot

temperature reaches higher than the ITRS reported T_{max} of 105^oC indicating that active hot-spot management is required in SOI technology.

Table 9.7: *Thermal-aware full-chip reliability report for the* 64*-bit ALU with* Cu/SiO_2 *metallization in SOI technology. Here the target lifetime of the circuit is* 20 *years.*

Operating Frequency	SOI Technology			
	$T_{\max}$ (^oC)	T_{avg} (^oC)	Full-chip prob. of no failure	Max FIT
$250MHz$	55.94	28.91	≈ 1	$8.88E-11$
$1GHz$	142.76	34.63	0.997	68.04
$2GHz$	258.51	42.25	$2.81E-27$	$395k$

Simulation results in tables 9.6 and 9.7 reiterate the importance of thermal-aware reliability analysis. Electromigration lifetime is exponentially dependent on the inverse of temperature. At low operating frequencies when power dissipation is not high enough, predicted full-chip reliability in the 64-bit ALU is very high. As the temperature increases with higher performance and power dissipation in the circuit, only then do we observe some degradation in reliability. The reliability degradation is more pronounced in SOI technology where the temperature rise is higher due to an oxide layer underneath the devices.

9.4 Summary

A 64-bit Arithmetic and Logic Unit (ALU) has been designed and analyzed for full-chip reliability to demonstrate SysRel's capability as well as to investigate reliability behavior in large circuit layouts. The layout is $1035um$ by $1036um$ in dimension and consists of 4193 cells. The cell-based layout with various Al and Cu interconnect technologies has been analyzed in SysRel. The number of trees is 89176. The immortality condition filters indicate that a significant number of trees are indeed immortal. Moreover, the number of mortal trees is sensitive to the (jL) product threshold. The higher (jL) product threshold in Al metallization identifies only 179 trees to be mortal. On the other hand, the lower (jL) product thresholds in Cu/SiO_2

based interconnects lead to 1142 mortal trees. Consistent with our earlier findings, the best full-chip reliability is achieved with Al bamboo type interconnects.

The 64-bit ALU circuit has been used to explore future reliability direction with Cu/low-k based interconnects. Using experimentally reported parameters for a particular low-k material, we find that the lowest (jL) product threshold in the Cu/low-k based interconnects leads to the highest number of mortal trees in the layout. In addition, lower lifetimes in the mortal trees generate the worst full-chip reliability metrics among other types of Al and Cu interconnects. SysRel has been used to investigate the required improvement in Cu atomic diffusivity such that full-chip reliability can be matched with the best case with Al bamboo type interconnects. Test-level reliability can be engineered by changing parameters, such as Cu atomic diffusivity, in the interconnects. Our simulation results indicate that significantly improved test-level reliability in Cu is required to achieve equivalent circuit-level reliability with Al bamboo type interconnects. Moreover, the required improvement will increase as low-k/low-modulus dielectrics are introduced.

Thermal-aware full-chip reliability simulation of the 64-bit ALU circuit has been carried out in SysRel using Cu/SiO_2 based interconnects in the layout. The operating frequency is varied in the circuit in both bulk CMOS and SOI technologies. The electromigration lifetime decreases exponentially with temperature increase. At low operating frequencies when the power dissipation is not high enough to generate significant temperature rise, predicted full-chip reliability in the 64-bit ALU is very high. As the temperature increases with higher performance, only then do we observe notable degradation in reliability. The reliability degradation is more pronounced in SOI technology where the temperature rise is higher due to an oxide layer underneath the devices. The simulation results establish that reliability is indeed a concern in high performance circuits and technology. Moreover, SysRel can demonstrate performance-reliability trade-off in high performance circuits and facilitate reliability-aware circuit design.

Chapter 10

Conclusions and Future Work

In the pursuit of higher performance and integration, Integrated Circuit (IC) technology is heading towards the nanotechnology era. The gate length of a state-of-the-art active device, such as a metal oxide silicon field-effect transistor (MOSFET), is in the range of 45 to 50 nm allowing millions of such devices to be fabricated in a single chip. In addition to achieving higher density, smaller devices increase speed due to higher drive current during the "on" state. While more and more devices are desirable for integrating more functionality in a single chip, interconnecting the devices using metal wires takes up an even greater percentage of space. Interconnect delay related to its resistance (R), capacitance (C), and in some cases inductance (L) has become dominant over gate delay. Cu has been replacing Al as the material of choice for interconnects due to its lower electrical resistance and expected improved resistance to electromigration-induced failures. However, with the present fabrication technology and interconnect architecture scheme, the reliability of Cu interconnects is not as high as it could be.

Electromigration experimental work often focuses on via-terminated straight line interconnects (figure 10-1(a)). However, in circuit layouts, straight line segments are linked through junctions to other segments in the same layer or different layers of metallization. Such junctions are often complex involving multiple linkage to segments with different dimensions and current densities. It has been established that an interconnect tree, a continuously connected piece of metal within one layer of metallization

(figure 10-1(b)), is the appropriate fundamental reliability unit for circuit-level reliability assessments for metallization schemes with fully blocking boundaries at the vias. Metal atoms within an interconnect tree can diffuse freely among the segments, and the stress evolutions and atomic fluxes in the segments are coupled. However, most circuit-level reliability assessment tools are still based on breaking up interconnect trees into individual segments and assessing the reliability of each segment separately, using the results from straight via-to-via test lines [33, 34]. This method is generally inaccurate.

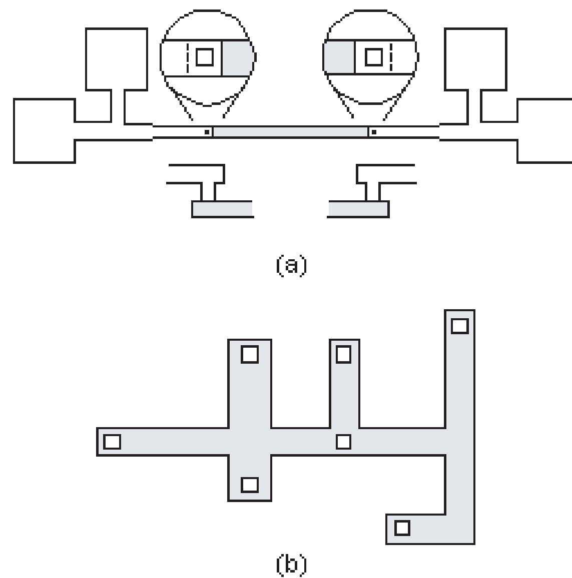

Figure 10-1: *(a) A simple via-terminated segment used in tests, (b) an example of a more complex interconnect tree on a single level of metallization.*

Another shortcoming in most existing tools is the treatment of all interconnect lines as being equally prone to electromigration failure. Via-terminated interconnect lines can be immune to electromigration-induced failure when operated below a critical product of current density and line length $(jL)_{crit}$. Rigid vias block electromigration and prevent the build-up of high enough mechanical stresses causing failure. Consequently, many interconnects in circuits with either Cu or Al metallization are 'immortal'. Immortality conditions need to be accounted in order for circuit-level reliability assessments to accurately predict full-chip reliability as well as to manage

workload from numerous interconnect trees in circuit layouts.

The present fabrication technology and interconnect architecture scheme for Cu interconnects lead to distinct electromigration failure characteristics. Cu interconnects are fabricated by the damascene method, in which a trench is first etched into a blanket layer of dielectric material before filling it with Cu by electroplating. To prevent Cu atoms diffusing into the device layer, thin refractory metal layers consisting of Tantalum (Ta) or Tantalum Nitride (TaN) are placed at the sides and bottom of the Cu interconnect lines. Cu-filled vias are used to connect multiple layers of metallization. The Cu lines are capped with a dielectric diffusion barrier, which is usually made of Silicon Nitride (Si_3N_4). It has been widely reported that the Cu/Si_3N_4 interface acts both as the dominant diffusion pathway for atoms and as the likely site for void nucleation. Gan *et al.* experimentally demonstrated that, due to this fact, the lifetimes of the M2 test structures were always higher than those of the M1 test structures with the same length, width, and number of vias at each end [40]. The underlying phenomenon applies to all metal layers and has been generalized using via-above and via-below definitions.

During electromigration in Cu interconnects, a tensile stress develops at the cathode end, where the Ta liner underneath a via usually forms a blocking boundary to diffusing Cu atoms. In via-below structures, voids preferentially nucleate at the Cu/Si_3N_4 interface due to the low critical stress (41 MPa or less) required for that interface [43]. An open-circuit failure would occur only when the void grows to span the whole thickness of the metal line, resulting in a very large void volume. On the other hand, in via-above structures the maximum tensile stress develops at the Cu/Si_3N_4 interface near the cathode via. Therefore, an open-circuit failure would occur if a small void forms below the via, such that the pathway for electron flow is blocked. This asymmetry in the void volume required for failure not only accounts for the asymmetry in lifetime but also contributes to the different $(jL)_{crit}$ products in immortality conditions for via-above and via-below type of interconnects.

10.1 Summary and Implications of Results

In this work, we have developed a new reliability CAD tool, SysRel, that addresses the shortcomings of existing tools and incorporates the circuit-level reliability analysis with Cu metallization technology. SysRel applies an interconnect tree based hierarchical reliability analysis flow with either Al or Cu metallization technology in both conventional and 3D circuit layouts. Along the tool development work, we have proposed and implemented various methodologies, such as full-chip reliability analysis, cell-based reliability analysis, layout-level temperature profiling, and reliability analysis with non-blocking via analysis in Cu metallization. Information on SysRel software release is available in Appendix D.

10.1.1 SysRel: Interactive Reliability CAD Tool

SysRel is an operating system independent stand-alone application written in Java 2. It treats circuit layout files in magic format created with Magic and 3D-Magic. It is an interactive tool initiating the steps in reliability simulation after receiving user inputs via menu bars. SysRel has its own desktop which hosts the output window, layout window, and reliability statistics reports after simulation (figure 10-2). A user can seamlessly migrate between Al and Cu metallization technologies in a circuit layout using the technology selection tab. Other significant features in SysRel include cell-based and thermal-aware reliability analysis.

SysRel utilizes the hierarchical flow for reliability analysis using interconnect trees extracted from circuit layouts as the fundamental reliability unit (FRU). The FRUs are filtered out for immortality in two stages, first using a current density of $j_{\max}$ from the ITRS, and then using j_{local} from circuit simulation. Via-type based (jL) product thresholds are applied in the filtering stages for Cu technology. However, no such via classification is required for Al technology as a single (jL) product threshold is sufficient. After the filtering stages, a default model is applied to the mortal FRUs to compute the expected lifetimes due to electromigration failure. Lifetimes are combined using a joint stochastic process with a series combination of all FRUs. Given a

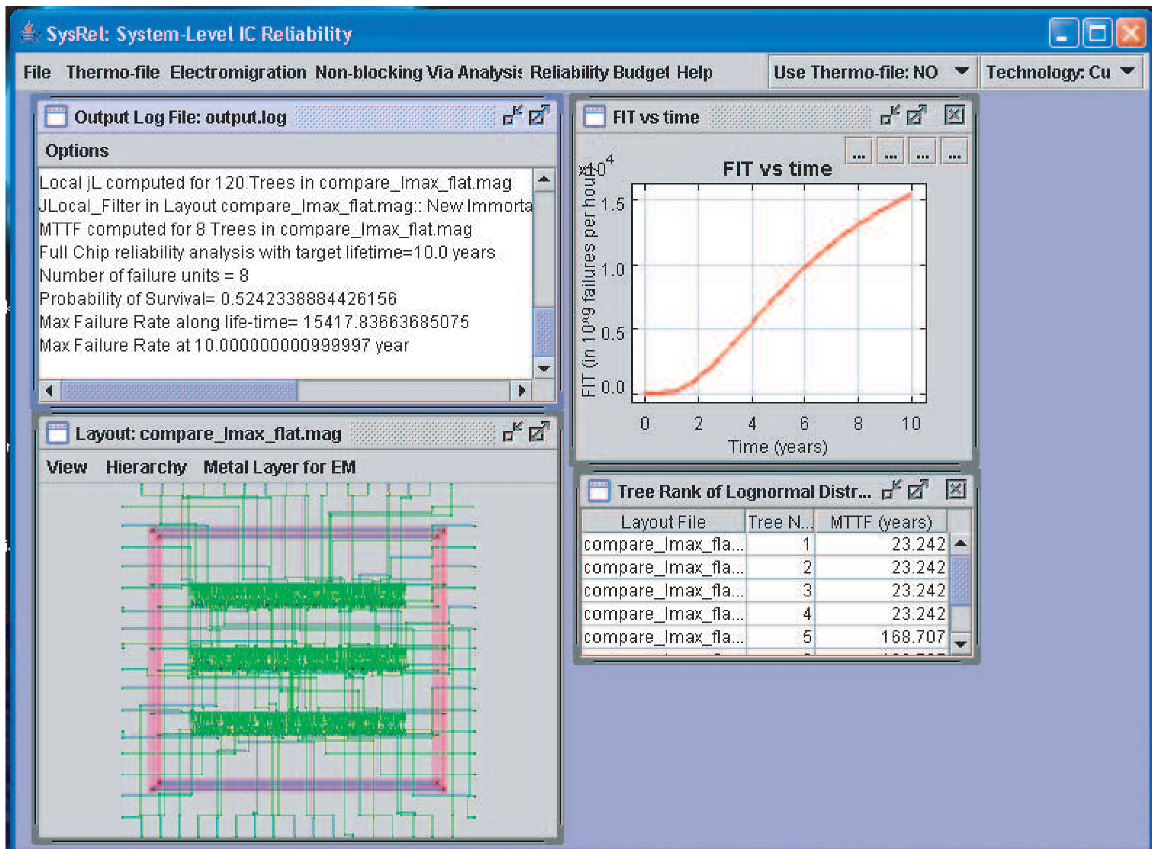

Figure 10-2: *Screen shot of SysRel desktop pane. At the top right corner are selection tabs to enable thermal analysis and to select either Cu or Al metallization. Inside the desktop pane, there are Output Window, Layout Window, Plot Window, and Tree Rank Table.*

target lifetime for the chip, the full-chip reliability metrics for output are: probability of survival, FIT (failure unit) along lifetime, maximum FIT, and time to cumulative % failure. A 32-bit comparator circuit layout in 2D and 3D technology is analyzed in SysRel to demonstrate its basic functionality and significance. SysRel identifies electromigration critical (mortal) trees and allows a designer to focus on those trees while ignoring numerous other immortal trees in the layout. Thus, a "what-if" analysis is possible on selected mortal trees to observe the impact on full-chip reliability. The hierarchical reliability flow in SysRel is crucial not only for computational efficiency, but also for accuracy in full-chip reliability analysis. Simulation results with a 64-bit ALU circuit demonstrate that 99.8% and 98.7% of the interconnect trees are immortal in Al and Cu metallization, respectively. Therefore, the lack of filtering algorithms in the flow would produce inaccurate full-chip reliability results.

Large circuits are often designed in a hierarchical flow where individual blocks or cells are first designed and characterized. We have developed and implemented in SysRel the concept of cell-level reliability characterization and computationally efficient full-chip reliability analysis. An IIT standard cell library has been characterized using SysRel. SysRel reads and retains the hierarchy levels in a hierarchical Magic layout. Reliability simulation results with the hierarchical Magic layouts of a 4-bit adder and 32-bit comparator demonstrates the significance of cell-based reliability analysis in SysRel. Statistics from the hierarchical layouts suggest that only a few cells are used multiple times and a large percentage of extracted interconnect trees are internal to cells. Therefore, cell-based reliability analysis, where cells are pre-characterized and hidden in a hierarchical layout, improves computational efficiency. The maximum improvement in times required for layout drawing and interconnect tree extraction of 60% and 45%, respectively, have been demonstrated with the 4-bit adder and 32-bit comparator layouts. The cell-based reliability assessment approach has allowed the simulation with the fairly large and complex 64-bit Arithmetic and Logic circuit.

10.1.2 Thermal-Aware Reliability Analysis in SysRel

Interconnect lifetime due to electromigration is exponentially dependent on the inverse temperature. Therefore, accurate thermal analysis is a prerequisite to circuit-level reliability analysis and reliability-aware IC design. We have conducted both steady state and transient thermal simulations with bulk CMOS transistors using ANSYS. It is possible to accurately model the transient temperature in digital circuits using the temperature rise due to DC/static power. When the circuit is operating at a high enough frequency such that there is minimal cooling between energy dissipation periods in a device, the device temperature reaches a steady state value proportional to the average power.

Based on our observations from ANSYS thermal simulations, the temperature profile due to a power source at the top surface of a substrate is axisymmetric and linearly dependent on power. Treating temperature as a linear and spatially independent variable, we define an impulse response as the temperature rise at the top surface of the substrate (transistor plane) due to $1mW$ of power dissipation in a unit area ($1um \times 1um$). Thus it is possible to compute the temperature distribution at the top surface of the substrate by convolving the impulse response with the power density profile of a layout. A Matlab program TProfile has been developed to compute temperature profile using the proposed method in frequency domain using Fast Fourier Transform. TProfile estimates the average temperature of a cell within 10% of that from ANSYS simulation.

Given a cell-based circuit layout and cell-based average power dissipation, SysRel outputs a power density matrix for TProfile. TProfile computes the non-uniform temperature profile at the top surface of the substrate. Using the layout-level temperature profile as a boundary condition, SysRel computes interconnect temperature using a conservative joule heating model. Thermal-aware reliability simulation with the 32-bit comparator layout suggests that using a worst case temperature for reliability analysis, and then setting a reliability goal may lead to excessive conservatism in IC design. While reliability simulation with the worst case interconnect temper-

ature of 105^oC predicts low full-chip reliability, thermal-aware reliability simulation indicates no failure within the chip's target lifetime due to low temperature rise.

Using the thermal-aware capability in SysRel, we have investigated the thermal effect from high performance operation and technology, such as silicon-on-insulator (SOI). The operating frequency is varied in the 64-bit ALU circuit in both bulk CMOS and SOI technologies. The electromigration lifetime decreases exponentially due an increase in temperature. At low operating frequencies when the power dissipation is not high enough, predicted full-chip reliability in the 64-bit ALU is very high. As temperature increases with higher performance, only then do we observe some degradation in reliability. The reliability degradation is more pronounced in SOI technology where the temperature rise is higher due to an oxide layer underneath the devices. The simulation results establish that reliability is indeed a concern in high performance circuits and technology. Moreover, SysRel can demonstrate performance-reliability trade-off in high performance circuits and facilitate reliability-aware circuit design.

10.1.3 Reliability Comparison of Copper and Aluminum Metallizations

Under similar test conditions, electromigration reliability of Al and Cu metallization interconnect trees demonstrate significant differences because of the differences in interconnect architectural schemes. Using SysRel, we have been able to compare for the first time circuit-level reliability with the two metallizations. SysRel differentiates between Cu and Al technologies during immortality filtering and lifetime estimation. Unlike Al technology, a (jL) product filtering algorithm with a classification of separate via-above and via-below treatments is required in Cu interconnect trees. Different diffusivity mechanisms lead to the differences in electromigration lifetimes. Al lines wider than the grain size $(d = 0.5um)$ have polygranular microstructure with the highest atomic diffusivity. On the other hand, narrow Al lines with bamboo type microstructure have lower diffusivity comparable to that of Cu metallization.

Using the best estimates of material parameters, electromigration lifetimes of a via-to-via straight line interconnect are compared for four configurations: Cu via-above, Cu via-below, Al bamboo, and Al polygranular. In typical operating condition ($j \leq 0.5MA/cm^2$ and $T \leq 105^oC$), an Al bamboo type line has the longest time-to-failure followed by Cu via-below, Cu via-above, and Al polygranular type lines. Circuit-level reliability simulations with the 32-bit comparator and 64-bit ALU circuits affirm that the best full-chip reliability is indeed achieved with Al bamboo type interconnects. The lower (jL) product thresholds in Cu/SiO_2 based interconnects lead to an increased number of mortal trees in circuit layouts. Cu/low-k based interconnects are reported to have even lower (jL) product thresholds. As predicted using the default model as well as reported experimentally, mortal trees have shorter lifetimes than Cu/SiO_2 based interconnects. Both effects contribute to the worst reliability prediction in the 64-bit ALU circuit with Cu/low-k interconnects.

Electromigration tests are done with simple interconnect structures. Test-level reliability can be engineered using different fabrication processes and materials that impact critical parameters, such as atomic diffusivity and effective modulus. Circuit-level reliability comparisons using various circuits and circuit elements in Al and Cu metallizations convey a profound message on the required test-level reliability target for Cu. Significantly improved test-level reliability in Cu is required to achieve the equivalent circuit-level reliability with Al bamboo type interconnects. Moreover, the required improvement will increase as low-k/low-modulus dielectrics are introduced.

10.1.4 Non-blocking Via Analysis with Copper Metallization

An interconnect tree is the fundamental reliability unit for circuit-level reliability assessments for metallization schemes with fully-blocking boundaries at the vias. Via-terminated interconnect lines in Al metallization can be immune to electromigration-induced failure when operated below a critical product of the current density and line length. Tungsten-filled vias block electromigration and prevent the build-up of the mechanical stress required to cause failure. In Cu metallization, refractory liners at vias generally block electromigration. However, some experimental studies suggest

that this is not always the case, and as liner thicknesses are decreased, fully-blocking liners at the vias become less certain due to liner ruptures. When *Cu*-filled vias are not fully blocking, an interconnect tree is connected to other metallization layers. While liner ruptures can lead to increased lifetimes in test structures, the impact of tree linking on circuit-level reliability is not clear. Therefore, we incorporated in SysRel a capability for making reliability analysis with non-blocking vias. Users can stochastically or deterministically assign non-blocking vias in a circuit layout. Multiple trees, when linked by non-blocking vias, are merged to create a single tree, which is then treated as the fundamental reliability unit.

The presence of non-blocking vias leads to either an increase or decrease in the number of mortal trees. Trees that would be immortal with blocking vias can be linked to form mortal trees. However, two or more mortal trees can be linked to form a single mortal tree with a longer effective length. The latter effect results in an overall reduction in the number of mortal trees, as illustrated with the 32-bit comparator circuit layout. If the default model for estimating the lifetime of a mortal tree assumes semi-infinite segment lengths, there is a predicted reliability improvement with non-blocking boundaries. However, if the impact of line length is taken into account (e.g., the lifetime to be inversely proportional to line length), the full-chip reliability would degrade. The impact of non-blocking vias depends on how reliability depends on length, and how non-blocking vias affect the line length dependence of reliability. SysRel simulation results demonstrate the importance of development and use of long test structures with non-blocking vias for accurate circuit-level reliability assessments with Cu metallization.

10.2 Potential Applications of SysRel

SysRel is a circuit-level reliability assessment tool. Therefore, circuit designers can use SysRel to get interactive feedback on full-chip reliability during design time. SysRel identifies mortal trees in a circuit layout. Circuit designers can then perform "what-if" analysis with the mortal trees to investigate different methods for potential reliability

improvement. The chief consequences of this research are a design-in reliability tool and associated methodologies. If circuit layouts are not compatible for analysis with SysRel due to layout format, one can still implement the methodologies from SysRel in other circuit-level analysis tools.

SysRel is a valuable tool for reliability and process engineers. One can conduct sensitivity analysis of material parameters, such as atomic diffusivity, critical stress, and bulk modulus, on full-chip reliability. This would help identify critical process parameters that impact full-chip reliability and allow process engineers to focus on those for optimum reliability improvement. In addition, SysRel is an ideal tool for predictive assessment of future technologies currently in development. It is possible to incorporate preliminary models and material parameters for new interconnect technologies and investigate the impact on circuit-level reliability.

10.3 Future Directions

The Reliability CAD tool, SysRel, and associated research work provide the framework of future research in various directions.

10.3.1 Optimization of SysRel

SysRel is written in Java 2. Java programs have been reported to be slow relative to other languages, such as C++. However, with the advent of just-in-time (JIT) compilers and several source code optimization techniques, the performance of Java programs can be significantly improved [92]. As SysRel has been developed as a pilot RCAD tool, only a few critical operations, such as Magic layout out parsing, are optimized for speed and memory usage. Further optimization is possible and would be essential for analysis with large fully flattened circuit layouts. In fact, non-blocking via analysis with the 64-bit ALU circuit was not feasible with the current version of SysRel due to long simulation time. Performance bottlenecks in SysRel need to be identified using a Java profiler tool, for example JProfiler [93]. Data-structures and algorithms can be optimized for fast interconnect tree extraction and layout drawing.

Several pointers for optimizing Java codes for speed and memory usage are available at [94, 95].

Another venue for improvement in SysRel is the development of sophisticated methodologies for local current density calculation in interconnect trees. While current density calculation in power delivery lines (unidirectional current flow) are relatively straight forward, segment-based current density estimation in signal networks is an active area of research [57]. Signal networks carry a combination of bidirectional and unidirectional currents in different segments. Detail current flow information including current directions in different segments of an interconnect tree is not available from any power analysis tool, and more importantly, simple algorithms are often not computationally feasible for application in a large circuit layout.

10.3.2 Enhanced Electromigration Lifetime Model

SysRel uses a default model for estimating electromigration lifetime at the vias originally developed for Al metallization, and then extended for Cu metallization technology [18, 41]. When compared to experimental lifetime results, the default model gives a conservative lifetime prediction for an interconnect tree. Moreover, the model assumes semi-infinite segment lengths while calculating the lifetime of a via connecting those segments. As indicated in this work, the line length effect on electromigration lifetime is a significant factor for accurate assessment of the impact of non-blocking vias in Cu metallization technology. Therefore, the line length dependence needs to be investigated for incorporation into the default model.

The default model for computing the lifetime of a via is appropriate for a single via at junctions as shown in figure 10-1(b). While such single via junctions are more common in signal lines, power delivery and clock lines are often wider with multiple vias at a junction. As illustrated in figure 10-3, multiple vias are used to reduce the current density through each via and to reduce the electrical resistance of the junction. Current implementation in SysRel lumps all vias into one and uses the current density computed for a single via. This is a conservative approach that does not take into account the effect of parallel paths in the junction. Experimental work

is required to characterize the impact of multiple vias on electromigration lifetime. The default model needs to be upgraded accordingly.

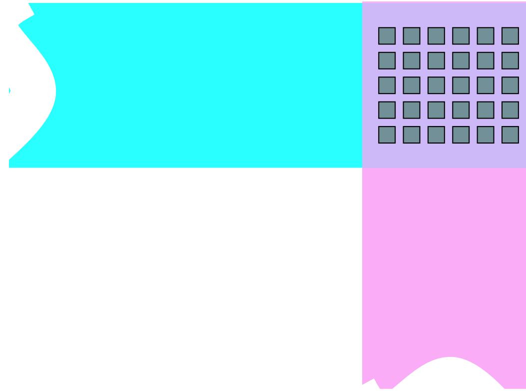

Figure 10-3: *Multiple vias at a junction in wide metal interconnects.*

The electromigration failure mechanisms and characteristics may change in future with the advent of low-k dielectric materials in Cu metallization technology. It has been reported that the electromigration lifetime of Cu/low-k interconnects is shorter than Cu/SiO_2 and that extrusions are more frequently seen at the anode end of the line during failure analysis [96]. The electromigration lifetime needs to be remodelled to account for new findings in future interconnect technologies. New models can be easily incorporated into SysRel to assess circuit-level reliability.

10.3.3 Thermal Analysis and Management in 3D IC Technology

SysRel currently has the capability for thermal-aware reliability analysis with conventional or 2D circuits. Thermal modeling and analysis that incorporates new thermal management strategies in 3D ICs is a promising area of research. SysRel provides the framework for layout-level thermal profiling in 3D circuit layouts created using 3D-Magic. A new 3D integration scheme with face-to-face bonds for high density inter-wafer connections and back-to-back bonds with embedded micro-channels (*thermal*

connects) at the Si interface has been proposed in section 1.4.3. Future research work with the 3D integration scheme can include fabrication and experiments with bonded wafers to quantify bonding strength in the presence of micro-channels. In addition, new thermal modeling techniques need to be developed that generate layout-level temperature profiles in 3D. The layout-level temperature profile in 3D circuits needs to incorporate nonuniform power dissipation in both horizontal and vertical planes. One can also develop methodologies for the optimal placement of micro-channels in the presence of nonuniform power dissipation. It would be interesting to incorporate thermal analysis capabilities for 3D ICs into SysRel and investigate the reliability impact from the proposed thermal management technique.

10.3.4 Impact of Process Variations

Due to process variations in advanced technologies, interconnect parameters, such as ILD thickness, metal height, and metal resistivity, are stochastic in nature [97]. Consequently, CAD simulation results may vary from Si implementations if the process variation effects are not taken into account. According to [98], the most common examples of deviations related to interconnect technology include reliability effects such as signal electromigration, power electromigration and IR drop, crosstalk glitch, and crosstalk delay. If the uncertainly is not taken into account, circuit design may lead to be either less reliable than expected or excessively conservative trading off maximum attainable performance. SysRel, being a layout-level analysis tool, has the framework to incorporate the effects of process variations. Interconnect pattern density, a contributing factor in interconnect parameter variations, is readily available from circuit layouts in SysRel [98]. Instead of using fixed values for ILD thickness, metal height, and metal resistivity, one can extend SysRel to include associated distributions for stochastic analysis. Using the extended SysRel, it will be interesting to quantify circuit-level reliability impact in the presence of process variations.

Appendix A

SysRel Source Code Overview

This appendix contains an overview of SysRel's source code in Java. The source code is available from Prof. Donald E. Troxel or Prof. Carl V. Thompson at MIT [99]. The code is written in Java 2. The Java classes are categorized into functional groups. The major functionalities and features of the classes, defined in the java file, *classname.java*, are briefly described under each category.

A.1 Main Application Classes

The main application classes are responsible for starting the SysRel application when it is invoked from the command line or by double clicking the jar file. These classes also define global variables for use in the rest of the application.

SysRel.java SysRel.java is the topmost class of SysRel. The primary function of SysRel is to invoke the graphical user interface. Menu bars and corresponding action listeners are defined in SysRel.java. SysRel uses multiple document interface (MDI) functionality to keep the internal modules, such as layout window, thermo-file window, and output window, centrally contained within the master application's window.

Globals.java This class contains the declaration for global constants, runtime variables, and debugging flags for use in the entire SysRel application.

A.2 Graphical Interface Classes

The graphical interface classes are developed using the Java Swing package available in Java Development Kit (JDK) version 1.3 [63]. Swing supports everything from creating menus, dialog boxes, and buttons in a Graphical User Interface (GUI) application.

MyLayoutFrame.java The MyLayoutFrame class extends or inherits from JInternalFrame class in Swing. This class is responsible for creating the layout window inside the SysRel desktop. MyLayoutFrame defines and supports basic menu operations for viewing a layout, such as zoom in, zoom out, zoom fit, and pan left, right, top, and bottom. Other menu operations include selecting different levels of hierarchy if a cell-based hierarchical Magic layout file is opened in the layout window. The third menu option lists all metal levels used in the layout and allows a user to elect either all levels or any one level for interconnect tree extraction.

LayoutPanel.java The LayoutPanel class inherits from the Swing JPanel class and provides infrastructure to display a layout in a scrollable window. The class has its own Graphics2D[1] object and a paint method to handle all rectangle drawing tasks.

MyThermoFrame.java The MyThermoFrame class extends JInternalFrame. The purpose of this class is to create a thermo-file window in SysRel that displays cell-level view of the layout with its temperature profile. Similar to MyLayoutFrame, MyThermoFrame defines and supports basic menu operations for viewing the layout, such as zoom in, zoom out, zoom fit, and pan left, right, top, and bottom.

ThermalPanel.java The ThermalPanel is equivalent to LayoutPanel class responsible for displaying the cell-based layout and temperature profile in the thermo-file window. The class has its own Graphics2D object and a paint method.

[1]The Graphics2D class is the abstract base class of Java for all graphics contexts that allows an application to draw onto components.

MyOutputFrame.java The MyOutputFrame class extends JInternalFrame and is responsible for creating the output window in SysRel's desktop pane. The class supports a scrollable text area in the output window.

A.3 File Parser Classes

The file parser classes primarily serve two purposes, parsing the Magic technology and layout files. While creating a layout window, MyLayoutFrame instantiates the file parser objects.

TechDB.java The TechDB class implements a recursive parser for reading Magic's technology file. The main constructor for the class is called with a string representation of technology file's Uniform Resource Locator (URL[2]). This allows SysRel to incorporate a technology file from anywhere in a networked system. An InputStream[3] object is created from the URL for reading the ASCII representation of a technology file. Then all the sections in a technology file are parsed using the parser methods, and information on different planes, tiles, and contacts are stored in Hashtable[4] data-structures.

ParseMAGFile.java The ParseMAGFile class implements the recursive parser for reading a Magic layout or a *.mag* file, and works in a similar fashion as TechDB. Given an URL, it opens a *.mag* file, and conditionally calls its parser methods, ParseSection(), ParseRect(), or ParseRLabel(), depending on what ParseLine() interprets from the current line in the file. All parsed information is "cached" internally to reduce the amount of time required for parsing large layouts. Following classes implement the cached representation of layout data.

[2]An URL is a pointer to a "resource" on the World Wide Web. A resource can be something as simple as a file or a directory.

[3]A superclass of all classes representing an input stream of bytes in Java.

[4]Hashtable is a Java data-storage class that maps distinct keys to stored values.

CachedSection.java This class implements the object representation of a "section" in a *.mag* file. A section in a Magic file is defined to start with a "<< foo >>", where "foo" is the name of a mask layer (e.g. polysilicon, metal1, etc.) or some control sections (e.g error_s, checkpaint, end). A CachedSection object contains all the coordinates of rectangular tiles that follow the section header tag.

CachedStmt.java The CashedStmt class implements the cached representation of any statement that is in *.mag* file's sections. It is an abstract type that must be subclassed to CachedRect, CachedRLabel, or CachedUse.

CachedRect.java This class extends the CachedStmt class. Given the parsed coordinates from the "rect ll lr ul ur" statement in a "section" of a *.mag* file, it defines a rectangular tile representing a mask layer.

CachedRlabel.java The CachedRlabel class extends CachedStmt, and implements the object representation for an "rlabel" statement in a section. The "rlabel" statement assigns user defined text labels to mask layers in a layout file.

CachedUse.java The CachedUse class also extends CachedStmt, and implements the object representation for an "use" statement in a *.mag* file. The "use" statement allows users to import instances of another *.mag* file into the layout.

CachedCell.java Finally, the CachedCell class implements the cached representation of the entire *.mag* file. A CachedCell contains multiple instances of CachedSection, CachedStmt, CachedRect, CachedRlabel, and CachedUse to fully define a layout. In a hierarchical layout, where a *.mag* file imports other *.mag* files, the CachedCell object for the parent layout has pointers to other CachedCell objects for the sub-layouts.

A.4 Corner-stitched Data-structure Classes

These classes define corner-stitched data-structure for the internal representation of a layout file. The corner-stitched data-structure is described in section 5.3.

Tile.java The Tile class implements the basic representation of a corner-stitched "tile" object in a layout. A "tile" is a rectangle corresponding to either a mask layer or, in case of a "space tile", absence of it. The Tile class contains "pointers" to eight tiles adjacent to the corner-edges. The neighboring tiles are stored in an array to facilitate faster search algorithms.

TileType.java The TileType class implements different categories of tiles in a particular layout. The different categories are defined according to the input from "types" and "contact" sections in a technology file. For every instance of a Tile class, there is a TileType object to represent its mask layer, such as metal1, poly, and metal2.

Plane.java The Plane class inherits from Tile, and represents a collection of Tile objects, including "space", that exists in any particular plane of a layout. Initially, a plane is a large "space tile" representing the absence of mask layers. As a layout is parsed, solid tiles are added to this large space plane at the proper positions.

PlaneType.java Similar to the TileType class, this class implements different categories of planes in a particular layout. The planes are defined according to the input from the "planes" section in a technology file.

A.5 Layout and Tree Representation Classes

The layout and tree representation classes implement the internal representation of a parsed *.mag* file with the corner-stitched data-structures. These classes also store the interconnect trees for further analysis.

Layout.java The Layout class defines a layout as a stack of multiple Plane objects where an individual Plane object has different types of Tiles to represent the mask layers. The constructor for this class takes CachedCell and TechDB objects as its input parameters for creating such a representation.

ITree.java This class represents an interconnect tree built from corner-stitched layout representation. An interconnect tree is stored as a collection of adjacent tiles that would form a continuous electrical path in one layer of metallization.

ISegment.java The ISegment class implements a Tile that is a part of an ITree. The ISegment class allows differentiation of Tile objects based on whether they form an interconnect tree or not. An ISegment object is specially tagged if it is a contact tile in a tree.

ISurface.java This class implements the surface area between adjacent ISegments in an ITree. The ISurface class can be used for generating tree input data for MIT EMSIM [50].

Path.java The Path class inherits a Vector[5] class and defines a specific path along an ITree object. It consists of a sequence of ISurfaces between starting and ending ISegments. An ITree object can have multiple paths as more than one ending ISegments can exist in an interconnect tree.

CellView.java The CellView class supports the cell representation in a hierarchical cell-based magic layout. In a cell-based layout, each cell has a CellView object with its own Tiles and Planes arrays. Multiple CellView objects are linked to the root Cellview. In case of a flattened layout, only one CellView object is created to represent the root layout.

[5]The Vector class in Java implements a growable array of objects. Like an array, it contains components that can be accessed using an integer index.

A.6 Reliability Computation Class

The reliability computation class is responsible for defining the methods and associated algorithms for reliability analysis. Only this class needs to be extended to add more features and new models in future.

Reliability.java The Reliability class implements the methods that operate on ITree objects and apply different filtering algorithms to isolate mortal interconnect trees. Methods, such as *applyJmaxLfilter()*, *applyRandomBC()*, and *applyExtraction()*, are called when the main application invokes reliability analyses from the "Electromigration" menu. The Reliability class contains values for technology-specific material parameters. Using the values, the method *applyDefaultModel()* computes the lifetime of mortal trees in Cu or Al metallization technology depending on the current technology selection. The Reliability class also supports non-blocking via analysis. The methods *generateNBvias()* assigns non-blocking vias in a flat layout with Cu metallization and *joinTrees()* creates merged trees linked by non-blocking vias.

Appendix B

Cadence to ANSYS Geometric Model Conversion: cds2inp

This appendix contains the perl script, cds2inp, that converts a Cadence circuit layout to ANSYS input file defining a 3D geometric model. Figure B-1 shows an example of the conversion using a transistor layout with diffusion contacts and metal routing. The Cadence circuit layout needs to be saved in an ascii text file. The instructions for using cds2inp are as follows:

1. Draw an outline or ESEXCL layer in the Cadence layout to indicate a boundary for the substrate in ANSYS simulation.

2. Move the origin to the lower left corner of the outline. Use menu operations: Edit→Other →Move Origin.. (to define a simple coordinate system in ANSYS)

3. Select layers, including the outline, for converting to ANSYS geometry. To select all layers in the layout, use menu operations: Edit→Select All.

4. From the Window menu in Cadence, select “Show Selected Set”. This will open a new window with the text representation of selected layers.

5. Save the text in a file, such as file1.out, as text format (default) using the File menu in the opened window. Copy the saved file to cds2inp script's directory location.

6. Run “cds2inp file1.out file2.inp [optional arguments]”. Run “cds2inp” to see a list of all optional arguments. File2.inp is the output file name to be created.

7. cds2inp reads material properties from /materials/*.mat files. A single .mat file defines desired material properties for a give material.

8. After starting ANSYS, read “file2.inp” using the File menu’s “read from file” option. This would create the geometry and define material properties automatically.

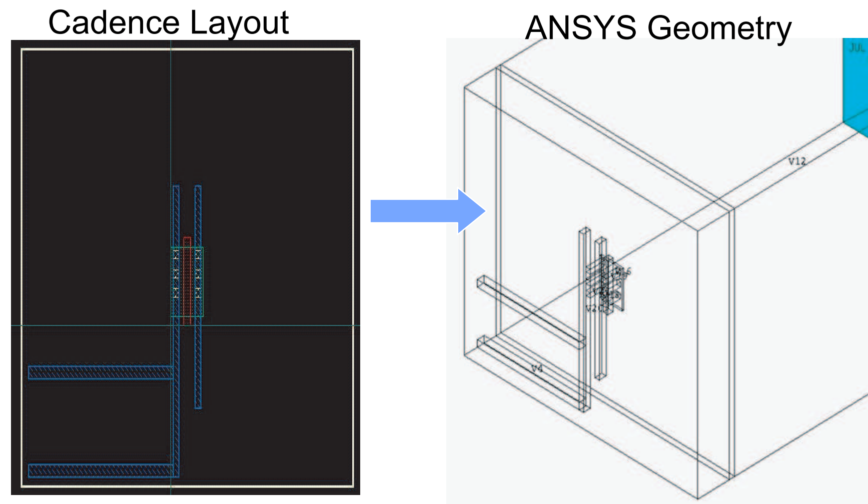

Figure B-1: *Cadence circuit layout to ANSYS geometry conversion demo with a transistor with diffusion contacts and metal routing.*

Perl script for cds2inp

```
use strict;
use warnings;
print "\n\n";
print "Starting the Cadence layout to ANSYS input geometryconverter\n";
print "Author: Syed M. Alam\n";
print "Last Modified @ MIT:November 7, 2003\n";
```

```
print "Usage: cds2inp ascii_layout outputfile [sub1_thickness in um]
[3d] [ff/bb/fb] [sub2_thickness in um]\n";
print "Bulk CMOS Technology\n"; print "----------------------------\n";
my $sub1_h; #default substrate_1 height is 10um
$sub1_h=10;
if ($ARGV[2] && $ARGV[2]>0) {$sub1_h=$ARGV[2];}
               #Add more layers and z-dimensions here in um
my @layers=qw(ESEXCLDIFF CONT POLY1 METAL1 ILD);
     #original layers, working for these layers as well
#my @layers=qw(OUTLINE RX CA PC M1 ILD);

# thickness of layers
my %layer_h= ( ESEXCL => 0,
        DIFF => 0.12,
        CONT => 0.68,
        POLY1 => 0.34,
        METAL1 => 0.50,
        ILD => 2.00);

my %layer_z1= ( ESEXCL => 0,
        DIFF => ($sub1_h-$layer_h{DIFF}),
        CONT => $sub1_h,
        POLY1 => $sub1_h,
        METAL1 => ($sub1_h+$layer_h{CONT}),
        ILD => $sub1_h);

my %layer_z2= ( ESEXCL => $sub1_h,
        DIFF => $sub1_h,
        CONT => ($sub1_h+$layer_h{CONT}),
        POLY1 => ($sub1_h+$layer_h{POLY1}),
```

```
        METAL1 => ($sub1_h+$layer_h{CONT}+$layer_h{METAL1}),
        ILD => ($sub1_h+$layer_h{ILD}) );

if (!$ARGV[0]) {print "Thank you for using cds2inp. See usage for
command line arguments\n\n"; exit;}

my $temp_v = 0; my $inputfile=$ARGV[0]; my $outputfile=$ARGV[1];
my $tdic; my $tdic_bond; my $sub2_h;

# default substrate 2 height is 10um $sub2_h=10; # dafault bonding
scheme is fb $tdic_bond="ff";

if ($ARGV[3] && $ARGV[3] eq "3d") {$tdic=1;} else {$tdic=0;} if
($ARGV[4]) {$tdic_bond=$ARGV[4];} if ($ARGV[5] && $ARGV[5]>0)
{$sub2_h=$ARGV[5];}


# z-dimensions for face-to-face bonding
my %layer_ff_z1= ( ESEXCL => ($layer_z2{ILD}+$layer_h{ILD}),
        DIFF => ($layer_z2{ILD}+$layer_h{ILD}),
        CONT => ($layer_z2{ILD}+$layer_h{ILD}-$layer_h{CONT}),
        POLY1 => ($layer_z2{ILD}+$layer_h{ILD}-$layer_h{POLY1}),
        METAL1 => ($layer_z2{ILD}+$layer_h{ILD}
                   -$layer_h{CONT}-$layer_h{METAL1}),
        ILD => $layer_z2{ILD} );
my %layer_ff_z2= ( ESEXCL => ($layer_z2{ILD}+$layer_h{ILD}+$sub2_h),
        DIFF => ($layer_z2{ILD}+$layer_h{ILD}+$layer_h{DIFF}),
        CONT => ($layer_z2{ILD}+$layer_h{ILD}),
        POLY1 => ($layer_z2{ILD}+$layer_h{ILD}),
        METAL1 => ($layer_z2{ILD}+$layer_h{ILD}-$layer_h{CONT}),
```

```
        ILD => ($layer_z2{ILD}+$layer_h{ILD})  );

# z-dimensions for face-to-back bonding
my %layer_fb_z1= ( ESEXCL => $layer_z2{ILD},
        DIFF => ($layer_z2{ILD}+$sub2_h-$layer_h{DIFF}),
        CONT => ($layer_z2{ILD}+$sub2_h),
        POLY1 => ($layer_z2{ILD}+$sub2_h),
        METAL1 => ($layer_z2{ILD}+$sub2_h+$layer_h{CONT}),
        ILD => ($layer_z2{ILD}+$sub2_h) );
my %layer_fb_z2= ( ESEXCL => ($layer_z2{ILD}+$sub2_h),
        DIFF => ($layer_z2{ILD}+$sub2_h),
        CONT =>($layer_z2{ILD}+$sub2_h+$layer_h{CONT}) ,
        POLY1 => ($layer_z2{ILD}+$sub2_h+$layer_h{POLY1}),
        METAL1 =>($layer_z2{ILD}+$sub2_h+
                  $layer_h{CONT}+$layer_h{METAL1}),
        ILD => ($layer_z2{ILD}+$sub2_h+$layer_h{ILD}) );

# z-dimensions for back-to-back bonding
my %layer_bb_z1= ( ESEXCL => -$sub2_h,
        DIFF => -$sub2_h,
        CONT => (-$sub2_h-$layer_h{CONT}),
        POLY1 => (-$sub2_h-$layer_h{POLY1}),
        METAL1 => (-$sub2_h-$layer_h{CONT}-$layer_h{METAL1}),
        ILD => (-$sub2_h-$layer_h{ILD}) );
my %layer_bb_z2= ( ESEXCL => 0,
        DIFF => (-$sub2_h+$layer_h{DIFF}),
        CONT => -$sub2_h,
        POLY1 => -$sub2_h,
        METAL1 => (-$sub2_h-$layer_h{CONT}),
        ILD => -$sub2_h );
```

```
print "Openning File: $inputfile\n"; open RFILE, "< $inputfile" or
die "Can't open input file:\n$!"; print "Openning File to Write:
$outputfile\n"; open WFILE, "> $outputfile" or die "Can't open
output file:\n$!"; writeheader(); while (my $line=<RFILE>) {
     $line=~ s/:/ /;
     $line=~ s/:/ /;
     my @larray = split(" ",$line);
     if ($larray[0] && $larray[0] eq "rect") {
    my $wline = block(@larray);
    if ($wline) {
        print "Write: $wline";
        print WFILE "$wline"; $temp_v=1;}
     }
} writefooter(); close WFILE; close RFILE; if($temp_v==0) {
   print "WARNING: No geometric data in $outputfile\n";
   print "WARNING: Removing $outputfile\n";
   #rm $outputfile;
} else
   { print "$outputfile written properly for ANSYS\n" }


sub block {
     my @larray=@_;
     my $solid = $larray[1];
     if ($solid eq "PC") {$solid="POLY1";}
     if ($solid eq "M1") {$solid="METAL1";}
     if ($solid eq "RX") {$solid="DIFF";}
```

```
if ($solid eq "OUTLINE") {$solid="ESEXCL";}
if ($solid eq "CA") {$solid="CONT";}
if (!grep(/$solid/, @layers)){ return;}
print "Read: @larray\n";
my $x1=$larray[4];
$x1=~ s/://;
my $x2=$larray[6];
$x2=~ s/://;
my $y1=$larray[5];
my $y2=$larray[7];
my $z1=$layer_z1{$solid};
my $z2=$layer_z2{$solid};

if ($solid eq "ESEXCL") {
if ($tdic) {
    if ($tdic_bond eq "ff")
    { return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,$x1,$x2,$y1,$y2,
              $layer_z1{ILD},$layer_z2{ILD}\nblock,$x1,$x2,$y1,$y2,
              $layer_ff_z1{ESEXCL},
              $layer_ff_z2{ESEXCL}\nblock,$x1,$x2,$y1,$y2,
              $layer_ff_z1{ILD},$layer_ff_z2{ILD}\n"; }
    if ($tdic_bond eq "fb")
    { return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,$x1,$x2,$y1,$y2,
    $layer_z1{ILD},
    $layer_z2{ILD}\nblock,$x1,$x2,$y1,$y2,$layer_fb_z1{ESEXCL},
    $layer_fb_z2{ESEXCL}\nblock,$x1,$x2,$y1,$y2,
    $layer_fb_z1{ILD},$layer_fb_z2{ILD}\n"; }
    if ($tdic_bond eq "bb")
    {return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,$x1,$x2,$y1,$y2,
    $layer_z1{ILD},$layer_z2{ILD}\nblock,$x1,$x2,$y1,$y2,
```

```
        $layer_bb_z1{ESEXCL},
        $layer_bb_z2{ESEXCL}\nblock,$x1,$x2,$y1,$y2,
        $layer_bb_z1{ILD},$layer_bb_z2{ILD}\n"; }
    }
    else
        { return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,$x1,$x2,$y1,$y2,
        $layer_z1{ILD},$layer_z2{ILD}\n";}
    }

    if ($tdic) {  #if tdic then have 2 block statements for each solid
     if($tdic_bond eq "ff")
          {return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,
          $x1,$x2,$y1,$y2,$layer_ff_z1{$solid},$layer_ff_z2{$solid}\n"; }
     if($tdic_bond eq "fb")
          { return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,$x1,$x2,$y1,$y2,
          $layer_fb_z1{$solid},$layer_fb_z2{$solid}\n"; }
     if($tdic_bond eq "bb")
          {  return "block,$x1,$x2,$y1,$y2,$z1,$z2\nblock,$x1,$x2,$y1,$y2,
          $layer_bb_z1{$solid},$layer_bb_z2{$solid}\n"; }
     }

    #not a tdic, just return one block statement
    return "block,$x1,$x2,$y1,$y2,$z1,$z2\n";
}

sub writeheader {
    print WFILE "/TITLE,Thermal Analysis File\n";
    print WFILE "KEYW,PR_SET,1\n";
    print WFILE "KEYW,PR_STRUC,0\n";
    print WFILE "KEYW,PR_THERM,1\n";
```

```
print WFILE "KEYW,PR_FLUID,0\n";
print WFILE "KEYW,PR_ELMAG,0\n";

print WFILE "KEYW,MAGNOD,0\n";
print WFILE "KEYW,MAGEDG,0\n";
print WFILE "KEYW,MAGHFE,0\n";
print WFILE "KEYW,MAGELC,0\n";
print WFILE "KEYW,PR_MULTI,0\n";
print WFILE "KEYW,PR_CFD,0\n";
print WFILE "/GO\n";
print WFILE "!*\n";
print WFILE "/COM,\n";
print WFILE "/COM,Preferences for GUI filtering set to display:\n";
print WFILE "/COM,  Thermal\n";
print WFILE "!*\n";
print WFILE "/PREP7\n";
print WFILE "!*\n";
print WFILE "ET,1,SOLID90\n";
print WFILE "!*\n";
print WFILE "mat,1\n";
print WFILE "mpread,materials/SiO2.mat,,,\n";
print WFILE "mat,2\n";
print WFILE "mpread,materials/Cu.mat,,,\n";
print WFILE "mat,3\n";
print WFILE "mpread,materials/bulkSi.mat,,,\n";
print WFILE "mat,4\n";
print WFILE "mpread,materials/channelSi.mat,,,\n";
print WFILE "mat,5\n";
print WFILE "mpread,materials/poly.mat,,,\n";
print WFILE "mat,6\n";
```

```
    print WFILE "mpread,materials/W.mat,,,\n";
}

sub writefooter {
    print WFILE "/VIEW, 1 ,1,1,1\n";
    print WFILE "/ANG, 1\n";
    print WFILE "/REP,FAST\n";
    print WFILE "/GRAPHICS,POWER\n";
    print WFILE "!*\n";
    print WFILE "/TYPE,1,8\n";
    print WFILE "/CPLANE,1\n";
    print WFILE "/SHADE,1,1\n";
    print WFILE "/REPLOT\n";
    print WFILE "VOVLAP,ALL\n";
    print WFILE "NUMM,KP\n";
    print WFILE "!*\n";
    print WFILE "/PNUM,KP,0\n";
    print WFILE "/PNUM,LINE,0\n";
    print WFILE "/PNUM,AREA,0\n";
    print WFILE "/PNUM,VOLU,1\n";
    print WFILE "/PNUM,NODE,0\n";
    print WFILE "/PNUM,TABN,0\n";
    print WFILE "/PNUM,SVAL,0\n";
    print WFILE "/NUMBER,0\n";
    print WFILE "!*\n";
    print WFILE "/PNUM,ELEM,0\n";
    print WFILE "/REPLOT\n";
    print WFILE "VPLOT\n";
}
```

Appendix C

Matlab Source Code for TProfile

```
0001 function TProfile(xd, impulse, size_I, PD)
0002 % xd: length axis in um for impulse response
0003 %     required step (delta_xd) = 0.02um
0004 % impulse: temperature (oC) from impulse response as a function of xd
0005 % size_I: discrete radius (unit 1um) of the impulse response for use in
0006 %         analysis
0007 % PD: power density matrix, unit: mW/um^2
0008 % Outputs:
0009 %         - Calculates and plots the temperature profile of the layout
0010 %         - Plots the impulse response matrix
0011 %         - Saves the temperature profile in ascii file: tprofile.dat
0012
0013 % Author: Syed M. Alam
0014 % Last updated: April 17, 2004
0015
0016 %transpose PD for same viewing axis as SysRel
0017 PD=PD';
0018 %plot PD;
0019 figure;
0020 imshow(PD,[min(PD(:)) max(PD(:))]);colorbar;
```

```
0021 title('Power Density (mW/um^2) Profile of the Layout')
0022 xlabel('x dimension, um')
0023 ylabel('y dimension, um')
0024
0025 % create impulse response matrix
0026 I=zeros(2*size_I+1);
0027 center = size_I+1;
0028 Id=[];
0029 % read temperature at 1um offsets upto size_I
0030 for w=0:size_I
0031     index = find(xd == w);
0032     Id = [Id impulse(index)];
0033 end
0034 % fill in the I matrix using axi-symmetric property
0035 for w=0:size_I
0036     for y=-w:w
0037         I(center-w,center+y)=Id(w+1);
0038         I(center+y,center+w)=Id(w+1);
0039         I(center+w,center+y)=Id(w+1);
0040         I(center+y,center-w)=Id(w+1);
0041     end
0042 end
0043
0044 % get row and column length of PD
0045 [nrow ncol]=size(PD);
0046
0047 % number of elements after convolution in row
0048 convlen_row = nrow+2*size_I+1-1;
0049 convlen_col = ncol+2*size_I+1-1;
0050
```

```
0051 % use pt-point FFT, pt need to be power of 2 for FFT speed
0052 pt=2048;
0053 if(max(convlen_row,convlen_col)>pt)
0054 disp('ERROR: PD or I matrix size is greater than FFT points');
0055 return;
0056 end
0057
0058 C=ifft2(fft2(PD,pt,pt).*fft2(I,pt,pt));
0059 C=C(1:convlen_row,1:convlen_col);
0060 % truncate extra points from 4 sides
0061 T=C(size_I+1:convlen_row-size_I,size_I+1:convlen_col-size_I);
0062 T=real(T);
0063
0064 %conventional heat sink technology
0065 Rth=0.26; %oC/W size 12cmX26cmX10cm with air vol. 0.066m^3/s
0066 Tamb = 27; %oC
0067 totalP=sum(sum(PD));   %mW
0068 Tsub=Tamb+totalP*10^-3*Rth;
0069 T=Tsub + T;
0070
0071
0072 figure;
0073 imshow(T,[min(T(:)) max(T(:))]);colorbar;
0074 title('Temperature (^oC) Profile of the Layout')
0075 xlabel('x dimension, um')
0076 ylabel('y dimension, um')
0077
0078 figure;
0079 surf(T);colormap(jet);colorbar;
0080 title('Temperature (^oC) Profile of the Layout')
```

```
0081 xlabel('x dimension, um')
0082 ylabel('y dimension, um')
0083 zlabel('temperature (^oC)')
0084
0085 figure;
0086 surf(I);colormap(jet);colorbar
0087 title('Impulse Response Matrix of Temperature (^oC)')
0088 xlabel('x dimension, um')
0089 ylabel('y dimension, um')
0090 zlabel('temperature (^oC)')
0091
0092 %need to transpose T as power density matrix is transpose
0093 T=T';
0094
0095 save tprofile.dat T -ascii;
0096 %fid=fopen('tprofile.dat','w');
0097 %fprintf(fid,'%3.4f  ',T);
0098 %fclose(fid);
0099 disp('tprofile.dat file is written with temperature profile matrix');
```

Appendix D

SysRel Release

D.1 SysRel v1.1

An internal release of SysRel v1.1 was made at MIT on March 6, 2004. The release consists of an installation CD and User Guide. Except for the thermal-aware analysis, SysRel v1.1 has all other capabilities as described in this thesis.

D.2 SysRel v2.0

SysRel v2.0 was released on August 20, 2004. A new version of the User Guide is available in SysRel v2.0 installation CD. SysRel v2.0 has all the capabilities described in the thesis. Users will be able to reproduce many simulation results and create more using the tutorials in User Guide v2.0.

SysRel release can be obtained by contacting [99].

Bibliography

[1] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat, G. Glass, T. Hoffmann, K. Johnson, C. Kenyon, J. Klaus, B. McIntyre, K. Mistry, A. Murthy, J. Sandford, M. Silberstein, S. Sivakumar, P. Smith, K. Zawadzki, S. Thompson, and M. Bohr. A 90nm High Volume Manufacturing Logic Technology Featuring Novel 45nm Gate Length Strained Silicon CMOS Transistors. In *Proceedings of the IEEE International Electron Devices Meeting*, pages 11.6.1–11.6.3, December 2003.

[2] International Technology Roadmap for Semiconductors. http://public.itrs.net. Interconnects.

[3] Focus Center Research Program on Interconnects. Microelectronics Advanced Research Corporation (MARCO). http://fcrp.src.org Founded in 1998.

[4] M. T. Bohr. Interconnect scaling-the real limiter to high performance ULSI. In *proceedings of IEEE International Electron Devices Meeting*, pages 241–244, December 1995.

[5] P. C. Wang, G. S. Cargill III, I. C. Noyan, and C. K. Hu. *Applied Physics Letter*, 72:1296, 1998.

[6] M. A. Korhonen, P. Boergesen, K. N. Tu, and C. Y. Li. *Journal of Applied Physics*, 73:3790, 1993.

[7] C. H. Hau-Riege. PhD Dissertation, Massachusetts Institute of Technology, Department of Materials Science and Engineering, June 2000.

[8] Y. C. Joo. PhD Dissertation, Massachusetts Institute of Technology, Department of Materials Science and Engineering, June 1995.

[9] V. K. Andleigh. *Modeling of and Experiments Characterizing Electromigration-Induced Failures in Interconnects*. PhD Dissertation, Massachusetts Institute of Technology, Department of Materials Science and Engineering, June 2001.

[10] G. Chee Lip. *Reliability Assessment Methodologies for Copper-Based Interconnects in Integrated Circuits*. PhD Dissertation, National University of Singapore, Singapore-MIT Alliance, Advanced Materials for Micro and Nano Systems, May 2003.

[11] R. G. Filippi, G. A. Biery, and R. A. Wachnik. *Journal of Applied Physics*, 78:3756, 1995.

[12] J. R. Black. In *Proceedings of the International Reliability Physics Symposium*, page 148, 1967.

[13] R. Kircheim and U. Kaeber. *Journal of Applied Physics*, 70:172, 1991.

[14] M. Shatzkes and J. R. Lloyd. *Journal of Applied Physics*, 59:3890, 1986.

[15] J. J. Clement. *Journal of Applied Physics*, 82:5991, 1997.

[16] C. K. Hu, K. P. Rodbell, T. D. Sullivan, K. Y. Lee, and D. P. Bouldin. *IBM Journal of Research and Development*, 39:465, 1995.

[17] C. K. Hu. Electromigration Failure Mechanisms in Bamboo-grained Al(Cu) Interconnections. *Thin Solid Films*, 260:124–134, 1995.

[18] S. P. Hau-Riege. *New Methodologies for Interconnect Reliability Assessments of Integrated Circuits*. PhD Dissertation, Massachusetts Institute of Technology, Department of Materials Science and Engineering, April 2000.

[19] K. W. Guarini, A. W. Topol, and M. Ieong et al. Electrical Integrity of State-of-the-Art 0.13 um SOI CMOS Devices and Circuits Transferred for Three-

Dimensional (3D) Integrated Circuit (IC) Fabrication. In *Proceedings of the IEEE International Electron Devices Meeting*, pages 943–945, December 2002.

[20] A. Fan, A. Rahman, and R. Reif. Copper Wafer Bonding. *Electrochemical and Solid State Letters*, 2:534–536, 1999.

[21] S. J. Souri and K. C. Saraswat. Interconnect Performance Modeling for 3D Integrated Circuits with Multiple Si Layers. In *Proceedings of the IITC*, pages 24–26, 1999.

[22] G. W. Neudeck. Three-Dimensional CMOS Integration. *IEEE Circuits and Devices Magazine*, 6:32–38, 1990.

[23] K. F. Lee, J. F. Gibbons, K. C. Saraswat, and T. I. Kamins. Thin Film MOSFET Fabricated in Laser-Annealed Polycrystalline Silicon. *Journal of Applied Physics Letters*, 35:173–175, 1979.

[24] A. Rahman, A. Fan, J. Chung, and R. Reif. Wire-length Distribution of Three-Dimensional Integrated Circuits. In *Proceedings of the IITC*, pages 233–235, 1999.

[25] A. Chandrakasan. The Collaborative Node. Interconnect Focus Center Program Review, August 15 2000.

[26] S. M. Alam, D. E. Troxel, and C. V. Thompson. A Comprehensive Layout Methodology and Layout-specific Circuit Analyses for Three-dimensional Integrated Circuits. In *Proceedings of the ISQED*, pages 246–251, March 2002.

[27] S. M. Alam, D. E. Troxel, and C. V. Thompson. Layout-specific Circuit Evaluation in Three-dimensional Integrated Circuits. *Analog Integrated Circuits and Signal Processing*, 35:199–206, 2003.

[28] N. Checka and C. Lau. Performance Comparison of 2D and 3D FPGAs. Design and Analysis of Digital Integrated Circuits, Professor Anantha Chandrakasan, Fall 2000.

[29] S. Das. *Design Automation and Analysis of Three-Dimensional Integrated Circuits*. PhD Dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, June 2004.

[30] A. Rahman and R. Reif. Thermal analysis of three-dimensional (3-D) integrated circuits (ICs). In *Proceedings of the IITC*, pages 157–159, 2001.

[31] S. M. Alam, D. E. Troxel, and C. V. Thompson. A Bonding Scheme with Heat Extraction and Dense Inter-wafer connection for Three-dimensional (3D) Integrated Circuit Technology. MIT Technology Licensing Office Application, Case no. 10699, April 2004.

[32] R. Krishnan, R. Tadepalli, and C. V. Thompson. Development of Electrical-mechanical Reliability test structures for characterization of bonded copper interconnects for 3-D IC applications - An Experimental Reliability/Yield Assessment method. MARCO Interconnect Focus Center Program.

[33] C. Hu. (Berkeley Reliability Simulator) BERT: an IC reliability simulator. *Microelectronics Journal*, 23:97–102, 1992.

[34] C. Teng, Y. Cheng, E. Rosenbaum, and S. Kang. iTEM: A Temperature-Dependent Electromigration Reliability Diagnosis Tool. *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, 16:882–893, 1997.

[35] Y. Chery, S. Hau-Riege, S. M. Alam, D. E. Troxel, and C. V. Thompson. ERNI: A Tool For Technology-Generic Circuit-Level Reliability Projections. Interconnect Focus Center Annual Review, December 14-15 1999.

[36] M. Perez. *Java Remote Microscope for Collaborative Inspection of Integrated Circuits*. Bachelor of Science and Master of Engineering Thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 1997.

[37] J. K. Ousterhout, G. T. Hamachi, R. N. Mayo, W. S. Scott, and G. S. Taylor. The Magic VLSI Layout System. *IEEE Design and Test*, pages 19–30, 1985.

[38] S. M. Alam, C. L. Gan, D. E. Troxel, and C. V. Thompson. Circuit level reliability analysis of Cu interconnects. In *Proceedings of the ISQED*, pages 238–243, March 2004.

[39] C. L. Gan, F. Wei, C. V. Thompson, K. L. Pey, W. K. Choi, S. P. Hau-Riege, and B. Yu. Contrasting Failure Characteristics of Different Levels of Cu Dual-Damascene Metallizations. In *Proceedings of the 9th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, page 124, 2002.

[40] C. L. Gan, C. V. Thompson, K. L. Pey, W. K. Choi, H. L. Tay, B. Yu, and M. K. Radhakrishnan. Effect of Current Direction on the Lifetime of Different Levels of Cu Dual-Damascene Metallization. *Applied Physics Letter*, 79:4592, 2001.

[41] C. L. Gan, C. V. Thompson, K. L. Pey, and W. K. Choi. Experimental characterization and modeling of the reliability of three-terminal dual-damascene Cu interconnect trees. *Journal of Applied Physics*, 94:1222, 2003.

[42] C. L. Gan, C. V. Thompson, K. L. Pey, and W. K. Choi. The Reliability of Wide-to-Narrow Cu interconnects. *Journal of Applied Physics*. In submission.

[43] S. P. Hau-Riege. *Journal of Applied Physics*, 91:2014, 2002.

[44] E. T. Ogawa, A. J. Bierwag, K. D. Lee, H. Matsuhashi, P. R. Justison, A. N. Ramamurthi, and P. S. Ho. *Applied Physics Letter*, 78:2652, 2001.

[45] K. D. Lee, E. T. Ogawa, H. Matsuhashi, P. R. Justison, K. S. Ko, , and P. S. Ho. *Applied Physics Letter*, 79:3236, 2001.

[46] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*, chapter 23.3, pages 477–484. MIT Press, 1997.

[47] C. S. Hau-Riege, A. P. Marathe, and V. Pham. The Effect of Line length on the Electromigration Reliability of Cu interconnects. In *Proceedings of the Advanced Metallization Conference*, page 169, 2002.

[48] J. A. Maiz. Characterization of electromigration under bidirectional (BC) and pulsed unidirectional (PDC) currents. In *Proceedings of the 27th International Reliability Physics Symposium*, page 220, 1989.

[49] J. J. Clement. Reliability analysis for encapsulated interconnect lines under dc and pulsed dc current using a continuum electromigration transport model. *Journal of Applied Physics*, 82:5991, 1997.

[50] MIT/EmSim: Electromigration Simulator. http://nirvana.mit.edu/emsim/.

[51] Z. S. Choi, J. H. Lee, J. K. White, and C. V. Thompson. CuEmSim: Simulation of electromigration and electromigration-induced damage in Cu-based interconnects. Ongoing research, Massachusetts Institute of Technology, 2003.

[52] NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/. May 2004.

[53] E. A. Amerasekera and F. N. Najm. *Failure Mechanisms in Semiconductor Devices*, chapter 2.3, page 32. John Wiley and Sons, 1998.

[54] J. Cho. *Effect of MicroStructure of Aluminum Alloys on the Electromigration-limited Reliability of VLSI Interconnects*. PhD Dissertation, Massachusetts Institute of Technology, Department of Materials Science and Engineering, September 1990.

[55] A. N. Kolmogorov. On A Logarithmic Normal Distribution Law Of The Dimensions Of Particles Under Pulverization. *Dokl. Akad Nauk, USSR 31*, 2:99–101, 1941.

[56] L. P. Murray, L. C. Rathbun, and E. D. Wolf. New Technique and Analysis of Accelerated Electromigration Life Testing in Multilevel Metallization. *Applied Physics Letter*, 53:1414, 1988.

[57] O. C. Oh, H. Haznedar, M. Gall, A. Grinshpon, V. Zolotov, K. Pon, and R. Panda. A methodology for chip-level electromigration risk assessment and product qualification. In *Proceedings of the ISQED*, pages 232–237, March 2004.

[58] C. K. Hu, R. Rosenberg, and K. Y. Lee. *Applied Physics Letter*, 74:2945, 1999.

[59] Y. J. Park, V. K. Andleigh, and C. V. Thompson. *Applied Physics Letter*, 85:3546, 1999.

[60] V. T. Srikar and C. V. Thompson. *Applied Physics Letter*, 74:37, 1999.

[61] F. L. Wei. *The Electromigration Drift Velocity and the Reliability of Dual-Damascene Copper Interconnect Trees.* SM Thesis, Massachusetts Institute of Technology, Department of Materials Science and Engineering, February 2004.

[62] J. R. Lloyd and J. J. Clement. *Thin Solid Films*, 262:135, 1995.

[63] Java 2 Platform Standard Edition v1.3.1 API Specification. http://java.sun.com/j2se/1.3/docs/api/. This is a web document on Java Platform core API provided by Sun Microsystems.

[64] J. K. Ousterhout. Corner-Stitching: A Data Structuring Technique for VLSI Layout Tools. *UC Berkeley Technical report CSD-83-154*, December 1983.

[65] J. Grad and J. E. Stine. A Standard Cell Library for Student Projects. In *Proceedings of the International Conference on Microelectronic Systems Education*, pages 98–99, 2003.

[66] S. M. Alam, F. Wei, C. L. Gan, C. V. Thompson, and D. E. Troxel. Impact of Non-blocking Vias on Electromigration and Circuit-level Reliability Assessments of Cu Interconnects. In *Proceedings of the Advanced Metallization Conference*, October 2004.

[67] C. K. Hu, L. Gignac, S. G. Malhotra, R. Rosenberg, and S. Boettcher. *Applied Physics Letter*, 78:904, 2001.

[68] F. Wei, C. L. Gan, C. V. Thompson, J. J. Clement, S. P. Hau-Riege, K. L. Pey, W. K. Choi, H. L. Tay, B. Yu, and M. K. Radhakrishnan. In *Proceedings of the Spring MRS*, page 645, 2002.

[69] B. S. Lall, B. M. Guenin, and R. J. Molnar. Methodology for Thermal Evaluation of Multichip Modules. In *Proceedings of the 11th IEEE SEMI-THERM Symposium*, pages 72–79, 1995.

[70] T. M. de Oca, B. Joiner, and Z. Johnson. Thermal Sub-modeling of High Density Interconnect Substrates. In *Proceedings of the Inter Society Conference of Thermal Phenomena*, pages 618–625, 2002.

[71] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu. On Thermal Effects in Deep Sub-Micron VLSI Interconnects. In *Proceedings of the 36th ACM Design Automation Conference*, pages 885–891, 1999.

[72] D. Chen, E. Li, E. Rosenbaum, and S. Kang. Interconnect Thermal Modeling for Accurate Simulation of Circuit Timing and Reliability. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 19:197, February 2000.

[73] J. H. Lienhard IV and J. H. Lienhard V. *A Heat Transfer Textbook*. Phlogiston Press, 2003.

[74] Y. K. Cheng, R. Paha, C. C. Teng, E. Rosenbaum, and S. M. Kang. ILLIADS-T: An electrothermal timing simulator for temperature-sensitive reliability diagnosis of CMOS VLSI chips. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 17:668, August 1998.

[75] ANSYS Theory Reference from ANSYS Documentation 7.1. Ansys 7.1. A license has been obtained for research work from Professor Alex Slocum's group, Massachusetts Institute of Technology.

[76] T. Chiang, K. Banerjee, and K. C. Saraswat. Compact modeling and SPICE-based simulation for electrothermal analysis of multilevel ULSI interconnects. In *Proceedings of the ICCAD*, pages 165–172, 2001.

[77] Thermal Properties of Materials. http://www.users.globalnet.co.uk/d̃mprice /ta/materials.htm.

[78] R. V. Joshi, S. S. Kang, and C. T. Chuang. 3D Thermal Analysis for SOI and its impact on Circuit Performance. IBM, T. J. Watson Research Center. Working paper.

[79] M. B. Kleiner, S. A. Kuhn, P. Ramm, and W. Weber. Thermal Analysis of Vertically Integrated Circuits. In *Proceedings of the IEDM*, pages 487–490, 1995.

[80] H. Nakayama et al. In *Proceedings of the CICC*, pages 381–384, 2000.

[81] K. A. Jenkins and R. L. Franch. Impact of Self-Heating on Digitial SOI and Strained-Silicon CMOS Circuits. In *Proceedings of the SOI Conference*, pages 161–163, 2003.

[82] J. Behm and J. Huttunen. Heat Spreading and Conduction in Compressed Heatsink. In *Proceedings of the International Flotherm Users Conference*, 2001.

[83] Y. Cheng, C. Tsai, C. Teng, and S. Kang. *Electrothermal Analysis of VLSI Systems*, chapter 4.3. Kluwer Academic Publishers, 200.

[84] S. Im and K. Banerjee. Full Chip Thermal Analysis of Planar (2-D) and Vertically Integrated (3-D) High Performance ICs. In *Proceedings of the IEDM*, pages 727–730, 2000.

[85] Tiny64. A simple configurable microprocessor in VHDL, June 2004. http://www.opencores.org.

[86] C. S. Hau-Riege, A. P. Marathe, and V. Pham. In *Proceedings of the 41st International Reliability Physics Symposium*, page 173, 2003.

[87] C. S. Hau-Riege, S. Hau-Riege, and A. P. Marathe. The Effect of Inter-level Dielectric on the Critical Tensile Stress to Void Nucleation for the Reliability of Cu Interconnects. In *Proceedings of the Stress Workshop*, 2004.

[88] M. Hatano, T. Usui, Y. Shimooka, and H. Kaneko. EM lifetime improvement of Cu damascene interconnects by p-SiC cap layer. In *Proceedings of the IITC*, page 212, 2002.

[89] J. M. Rabaey. *Digital Integrated Circuits, A Design Perspective*, chapter 4.4, page 234. Prentice Hall, 1996.

[90] L. T. Su and et al. *IEEE Transactions on Electron Devices*, 41:69–75, 1994.

[91] G. G. Shahidi. SOI technology for the GHz era. *IBM Journal of Research and Development*, 46:121–131, 2002.

[92] Java World. Performance tests show Java as fast as C++, February 1998. http://www.javaworld.com/javaworld/jw-02-1998/jw-02-jperf_p.html.

[93] Java profiler available from ej-technologies JProfiler. http://www.ej-technologies.com/products/jprofiler/overview.html.

[94] T. Singh. Things you can do to speed up things, June 2004. http://www10.brinkster.com/geniusportal/speedjava.html.

[95] J. Hardwick. Optimizing Java for Speed, March 1998. http://www-2.cs.cmu.edu/j̃ch/java/speed.html.

[96] K. D. Lee, X. Lu, E. T. Ogawa, H. Matsuhashi, P. S. Ho, V. A. Vlaschke, and R. Augur. In *Proceedings of the 40th International Reliability Physics Symposium*, page 322, 2002.

[97] D. O. Ouma, D. S. Boning, J. E. Chung, W. G. Easter, V. Saxena, S. Misra, and A. Crevasse. Characterization and Modeling of Oxide Chemical-Mechanical Polishing Using Planarization Length and Pattern Density Concepts. *IEEE Transactions on Semiconductor Manufacturing*, 15:232, May 2002.

[98] T. Vandenberge. Designing for process variation at 130 nm , EE Times, September 2003. http://www.eetimes.com/story/OEG20030908S0052.

[99] D. E. Troxel, troxel@mit.edu, C. V. Thompson, cthomp@mtl.mit.edu, and Syed M. Alam, salam@alum.mit.edu. Massachusetts Institute of Technology.