

**PLACEMENT FOR FAST AND RELIABLE  
THROUGH-SILICON-VIA (TSV) BASED 3D-IC LAYOUTS**

A Dissertation  
Presented to  
The Academic Faculty

by

Krit Athikulwongse

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

Georgia Institute of Technology  
December 2012

Copyright © 2012 by Krit Athikulwongse

**PLACEMENT FOR FAST AND RELIABLE  
THROUGH-SILICON-VIA (TSV) BASED 3D-IC LAYOUTS**

Approved by:

Dr. Sung Kyu Lim, Advisor  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Saibal Mukhopadhyay  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Muhannad Bakir  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Madhavan Swaminathan  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Hyesoon Kim  
College of Computing  
*Georgia Institute of Technology*

Date Approved: 25 July 2012

*Dedicated to my father, Piyah Athikulwongse, who passed away since my childhood,  
my mother, Kesaraporn Athikulwongse, who raised me by herself since then,  
and my sisters and brothers for their unconditional love and support.*

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Prof. Sung Kyu Lim, for his time, patience, and encouragement throughout my Ph.D. career under his guidance. I deeply appreciate all the supports he provided and the trust he put in me to do the research. I have learned much from him not only how to work on research but also how to work with people professionally. I am also indebted to Prof. D. Scott Wills, my former advisor, who provided invaluable supports throughout my study at Georgia Tech even after we no longer worked together. I am grateful to Prof. David Z. Pan from the University of Texas at Austin for his insightful suggestions on experiments during many hours of discussions on my research. I would like to express my appreciation to Prof. Saibal Mukhopadhyay for serving as my Reading Committee member, explaining to me the in-depth theories, and providing intriguing feedbacks on my research. I am thankful to Prof. Muhannad Bakir, my Reading Committee member; Prof. Madhavan Swaminathan and Prof. Hyesoon Kim, my Defense Examination Committee members; and Prof. Paul Hasler, my Proposal Review Committee member, for their valuable time serving on my committees and useful comments they provided. I am also thankful to Prof. Gabriel H. Loh and Prof. Hsien-Hsin S. Lee.

I would like to express my appreciation to GTCAD members for giving me a wonderful time and experiences, especially Dr. Dae Hyun Kim for his friendship, countless assistances, and encouraging discussions; Xin Zhao for her constant support and friendship, especially during my first year with GTCAD; Young-Joon Lee for his several helps on usage of EDA tools and coding GTCADesigner; Moongon Jung for his frank comments, helpful feedbacks, and insightful suggestions; Mohit Pathak for practical discussions and his help on temperature-aware placer; Dr. Michael B. Healy for sharing his insight on usage of Intel Math Kernel Library; Taigon Song for his numerous assistances, especially in equipment issues; Shreepad Panth for his help on block-level designs and useful discussions; Chang Liu; Dr. Mongkol Ekpanyapong; Rohan Goel; Hemant Sane; and Dr. Ismail F. Baskaya. I would



also like to thank GREEN members, Subho Chatterjee, Jeremy Tolbert, Minki Cho, Amit R. Trivedi, and Kwanyeob Chae; MARS and STING members, Dr. Dong Hyuk Woo, Dean L. Lewis, Tzu-Wei Lin, Mohammad M. Hossain, and Guanhao Shen; and UTDA members from the University of Texas at Austin, Dr. Jae-Seok Yang, Dr. Ashutosh Chakraborty, Jiwoo Pak, and Joydeep Mitra, for our fruitful discussions and productive collaboration.

I would like to express my gratitude to ECE personnel, especially Pamela F. Halverson and Beverly J. Scheerer for their fabulous support and helping me on numerous administrative issues. I am also thankful to Peter L. Huynh, David S. Webb, and Keith L. May for their great IT support.

I am indebted to my parents for the love and encouragement they have provided throughout my life. I am also grateful to my sisters and brothers, especially Raweewan Athikulwongse and Mongkol Athikulwongse, who have constantly motivated and supported me. I would like to express my sincere appreciation to wonderful families I have known during my time in Atlanta, especially Pakinee and Boonluer Chungwatana, Ling and Donald Goh, and Elizabeth and Lipton Chinloy for their warm friendship and numerous assistances. I am thankful to L. Suzanne Marger, my landlady, for the kindness and generosity she has provided during my stay at her place. I would like to thank many valuable friends I have made in Atlanta, especially Dr. Vichai Meemongkolkiat, Dr. Nattapon Chayopitak, and Sampan Nettayanun for their useful advices and countless helps.

Lastly, I would like to thank the National Electronics and Computer Technology Center (NECTEC) and the National Science and Technology Development Agency (NSTDA), the Royal Thai Government, for providing support and opportunity to pursue a Ph.D. degree in the United States to me.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . .	iv
LIST OF TABLES . . . . .	x
LIST OF FIGURES . . . . .	xii
LIST OF SYMBOLS OR ABBREVIATIONS . . . . .	xvi
SUMMARY . . . . .	xvii
CHAPTER I INTRODUCTION . . . . .	1
1.1 Objective . . . . .	2
1.2 Contributions . . . . .	2
1.3 Organization . . . . .	4
CHAPTER II ORIGIN AND HISTORY OF THE PROBLEM . . . . .	6
2.1 Wirelength . . . . .	6
2.2 Stress-induced Carrier Mobility Variation . . . . .	7
2.3 Temperature . . . . .	8
2.4 Quality Trade-offs . . . . .	9
CHAPTER III WIRELENGTH-DRIVEN PLACEMENT FOR GATE-LEVEL DESIGN OF 3D ICS . . . . .	11
3.1 3D-IC Design Flow . . . . .	12
3.2 3D Placement Algorithm . . . . .	14
3.2.1 Overview of Force-directed Placement . . . . .	14
3.2.2 Overview of the 3D Placement Algorithm . . . . .	15
3.2.3 Placing Cells in 3D ICs . . . . .	16
3.2.4 Placing TSVs in TSV Coplacement Scheme . . . . .	17
3.2.5 Net Splitting . . . . .	17
3.2.6 Preplacing TSVs in TSV-Site Scheme . . . . .	17
3.3 TSV Assignment . . . . .	18
3.4 Experimental Results . . . . .	19
3.4.1 Effectiveness of Net Splitting . . . . .	20
3.4.2 Wirelength and Runtime Comparison . . . . .	20

3.4.3	Metal Layers and Silicon Area . . . . .	22
3.4.4	On Wirelength vs. Number of TSVs . . . . .	22
3.4.5	On Wirelength and Die Area vs. Number of Dies . . . . .	23
3.4.6	TSV Coplacement vs. TSV-Site Schemes . . . . .	24
3.5	Summary . . . . .	25
CHAPTER IV IMPACT OF MECHANICAL STRESS AND PLACEMENT ON THE TIMING OF TSV-BASED 3D ICS . . . . .		27
4.1	Introduction . . . . .	28
4.2	Related Work and Motivation . . . . .	30
4.3	Modeling and Design Flow . . . . .	33
4.4	Carrier Mobility Variation . . . . .	34
4.4.1	Mobility Variation under TSV-induced Stress . . . . .	34
4.4.2	Mobility Variation under STI-induced Stress . . . . .	38
4.4.3	Mobility Variation under Both TSV and STI-induced Stress . . . . .	42
4.5	Timing Analysis with Stress Consideration . . . . .	44
4.5.1	Timing Analysis for 3D ICs . . . . .	44
4.5.2	Timing Library for Mobility Variation . . . . .	45
4.6	TSV-stress-driven Placement Optimization . . . . .	47
4.7	TSV-stress-driven Global Placement . . . . .	49
4.7.1	Carrier Mobility-Based Forces . . . . .	49
4.7.2	Convergence of TSV-stress-driven Global Placement . . . . .	52
4.8	Experimental Results . . . . .	53
4.8.1	Full-Chip Mobility Variation Map . . . . .	53
4.8.2	Full-Chip Timing Analysis Results . . . . .	54
4.8.3	Manual Placement Optimization Results . . . . .	58
4.8.4	Impact of KOZ on Carrier Mobility Variation . . . . .	61
4.8.5	Impact of KOZ on Area and Wirelength . . . . .	62
4.8.6	Impact of KOZ on TSV-stress-aware Timing . . . . .	63
4.8.7	TSV-stress-driven Placement Results . . . . .	64
4.9	Summary . . . . .	66

CHAPTER V	EXPLOITING DIE-TO-DIE THERMAL COUPLING IN 3D-IC PLACEMENT . . . . .	67
5.1	Motivation . . . . .	68
5.2	Global Placement Algorithms . . . . .	70
5.2.1	Design Flow . . . . .	70
5.2.2	Force-directed 3D Placement . . . . .	71
5.2.3	TSV Spread and Alignment . . . . .	72
5.2.4	Thermal Coupling-aware Placement . . . . .	73
5.3	Evaluation Flow . . . . .	79
5.3.1	Power Analysis for 3D ICs . . . . .	80
5.3.2	GDSII-Level Thermal Analysis . . . . .	80
5.4	Experimental Results . . . . .	83
5.4.1	Impact of TSV Density Uniformity . . . . .	84
5.4.2	Temperature-Wirelength Trade-off . . . . .	85
5.4.3	Comparison with State-of-the-Art . . . . .	86
5.4.4	Power and Thermal Maps . . . . .	89
5.4.5	Runtime Results . . . . .	91
5.5	Summary . . . . .	92
CHAPTER VI	BLOCK-LEVEL 3D-IC DESIGN QUALITY TRADE-OFFS STUDY FOR UNBALANCED DIE STACKING . . . . .	93
6.1	Background . . . . .	94
6.1.1	Die Bonding and Redistribution Layers . . . . .	94
6.1.2	The Goal of This Work . . . . .	96
6.2	Block-Level 3D-IC Design . . . . .	96
6.2.1	Partitioning . . . . .	97
6.2.2	TSV Insertion and Floorplanning . . . . .	97
6.2.3	Bonding Pad Assignment and RDL Routing . . . . .	98
6.3	Design Evaluation . . . . .	98
6.3.1	Traditional Metrics . . . . .	98
6.3.2	Thermal Analysis . . . . .	99
6.3.3	Mechanical Stress Analysis . . . . .	100

6.4	Experimental Results . . . . .	100
6.4.1	Baseline Designs . . . . .	100
6.4.2	Impact of TSV Size . . . . .	106
6.4.3	Impact of TSV Pitch . . . . .	107
6.5	Summary . . . . .	108
CHAPTER VII CONCLUSIONS . . . . .		109
REFERENCES . . . . .		111
PUBLICATIONS . . . . .		117
VITA . . . . .		119

## LIST OF TABLES

Table 1	Benchmark circuits. . . . .	19
Table 2	Wirelength from TSV coplacement scheme with and without net splitting. . . . .	20
Table 3	Comparison of wirelength (WL) and runtime for placement for IWLS 2005 benchmarks and industrial circuits. Cell occupancy is 80 %, and the number of 3D nets is set to 3 % to 5 % of the number of total nets during partitioning. The numbers in parentheses are ratios to 2D. . . . .	21
Table 4	Comparison of the minimum number of metal layers (ML) and total silicon area for 2D and 3D (4 dies) designs for IWLS 2005 benchmarks and industrial circuits. The numbers in parentheses are ratios to 2D. . . . .	23
Table 5	Comparison of wirelength of TSV coplacement and TSV-site scheme with MST-based TSV assignment and placement-based TSV assignment [1]. The numbers in the parentheses are ratios to TSV coplacement. . . . .	25
Table 6	Benchmark circuits. . . . .	55
Table 7	Comparison of hole mobility variation range. . . . .	55
Table 8	Comparison of electron mobility variation range. . . . .	56
Table 9	Longest path delay (LPD) comparison. (Percentage of changes is shown in parenthesis.) . . . . .	56
Table 10	Total negative slack (TNS) comparison. (Percentage of changes is shown in parenthesis.) . . . . .	57
Table 11	Gate optimization considering only TSV stress on the target path with perturbation. . . . .	59
Table 12	Gate optimizations considering both TSV and STI stresses on the target path with perturbation. . . . .	61
Table 13	Benchmark circuits. . . . .	62
Table 14	Impact of KOZ on carrier mobility variation for ckt5. . . . .	63
Table 15	Impact of KOZ on area and wirelength for ckt5. . . . .	63
Table 16	Impact of KOZ on TSV-stress-aware timing for ckt5. . . . .	64
Table 17	Timing comparison for regular and irregular TSV position with 2-row TSVs. . . . .	65
Table 18	Notations used for thermal coupling-aware placement. . . . .	74
Table 19	Benchmark circuits. . . . .	84
Table 20	Routed wirelength, longest path delay, and power of placements with uniform [1] and non-uniform [1] TSV position. . . . .	85

Table 21	Temperature ( $^{\circ}\text{C}$ ) of placements with uniform [1] and non-uniform [1] TSV position. ( $\Delta T_{ja} = T_{ja,max} - T_{ja,min}$ ) . . . . .	85
Table 22	Comparison with state-of-the-art temperature-aware placers [2, 3, 4, 5, 1]. The proposed placers are TSA (TSV spread and alignment) and CA (Coupling-aware placement). The routed wirelength, delay, and power values are normalized to the non-uniform TSV placement [1] shown in Table 20. The temperature values are normalized to the uniform TSV placement [1] shown in Table 21. . . . .	88
Table 23	Runtime comparison of uniform TSV placement [1], non-uniform TSV placement [1], state-of-the-art temperature-aware placers [2, 3, 4, 5] and the proposed placers. The proposed placers are TSA (TSV spread and alignment) and CA (Coupling-aware placement). . . . .	91
Table 24	Characteristics of the test circuit (reconfigurable computing array) and baseline design. . . . .	101
Table 25	Comparison of area, footprint, and wirelength in different layouts. TSV-f, TSV-d, and TSV-w are TSV-farm, TSV-distributed, and TSV-whitespace, respectively. The numbers in parenthesis after design style are TSV size and pitch in $\mu\text{m}$ . . . . .	102
Table 26	Comparison of longest path delay (LPD), with and without optimization, and number of buffers in different layouts. TSV-f, TSV-d, and TSV-w are TSV-farm, TSV-distributed, and TSV-whitespace, respectively. The numbers in parenthesis after design style are TSV size and pitch in $\mu\text{m}$ . . . . .	103
Table 27	Comparison of power and temperature of different layouts. TSV-f, TSV-d, TSV-w, and TSV-p are TSV-farm, TSV-distributed, and TSV-whitespace, respectively. The numbers in parenthesis after design style are TSV size and pitch in $\mu\text{m}$ . . . . .	104
Table 28	Comparison of stress of different layouts. . . . .	105

## LIST OF FIGURES

Figure 1	Via-first and via-last TSVs. . . . .	12
Figure 2	Two 3D-IC design flows developed in this work, (a) TSV coplacement and (b) TSV-site. . . . .	13
Figure 3	Splitting a 3D net into subnets (side view). . . . .	18
Figure 4	Wirelength distribution of (a) des_perf, where the die width is $572\mu\text{m}$ in 2D design and $311\mu\text{m}$ in 3D design (4 dies), and (b) b19, where the die width is $762\mu\text{m}$ in 2D design and $411\mu\text{m}$ in 3D design (4 dies). . . . .	21
Figure 5	Wirelength vs. number of TSVs of (a) des_perf and (b) b19 for 2D and 3D (4 dies) designs. . . . .	23
Figure 6	Wirelength vs. number of dies of des_perf in 3D design. . . . .	24
Figure 7	Die area and number of TSVs of des_perf in 3D design. . . . .	24
Figure 8	Cadence SoC Encounter snapshot of the bottommost die of Ind2 designed by (a) TSV coplacement and (b) TSV-site schemes. Routing for 3D nets is shown in blue. . . . .	25
Figure 9	Layout with small vs. large KOZ around TSVs. TSV landing pads are large yellow squares. . . . .	29
Figure 10	Thermal stress around TSV. . . . .	30
Figure 11	Thermal stress in active region caused by surrounding STIs. . . . .	31
Figure 12	Mobility change due to tensile stress. Top: $\Delta\mu/\mu$ for longitudinal tensile stress, bottom: $\Delta\mu/\mu$ for transverse tensile stress. . . . .	31
Figure 13	Buffer cell delay change due to TSV stress. (a) slower rising delay with longitudinal tensile stress, (b) faster rising delay with transverse tensile stress. . . . .	32
Figure 14	Overall flow for TSV/STI stress modeling and analysis flow. . . . .	34
Figure 15	Optimal orientation of MOSFET to maximize mobility for (001) surface and $\langle 110 \rangle$ channel. . . . .	36
Figure 16	Mobility contour map for a TSV. Top: contour map for hole mobility variation, bottom: contour map for electron mobility variation. . . . .	37
Figure 17	Contour of stress (FEA simulation) caused by TSVs nearby a cell. . . . .	38
Figure 18	Linear superposition of stress (FEA simulation) caused by TSVs nearby a cell. . . . .	38
Figure 19	Zigzag TSV placement has small $(\Delta\mu/\mu)_h$ between rows due to compensation. . . . .	39
Figure 20	Contour of stress (FEA simulation) caused by STI in horizontal direction. . . . .	39



Figure 21	Stress (FEA simulation) on a horizontal line across the center of the STI in Figure 20. . . . .	40
Figure 22	Setup for FEA simulations used to model STI stress. . . . .	40
Figure 23	STI stress (FEA simulation and model) at different distances. . . . .	41
Figure 24	Stress (FEA simulation and model) induced by STI with different widths. . . . .	41
Figure 25	Contour map of stress (model) for a single $4\mu\text{m}$ -wide STI. . . . .	42
Figure 26	Contour maps of mobility (model) for a STI. (a) hole mobility variation, (b) electron mobility variation. . . . .	42
Figure 27	Contour of stress (FEA simulation) caused by TSV on top of a cell with an STI on its right side. . . . .	43
Figure 28	Linear superposition of stress (FEA simulation) caused by a TSV and an STI. . . . .	43
Figure 29	Impact of the interaction between TSV and STI stress (model) on mobility variation. (a) hole, (b) electron. . . . .	44
Figure 30	Timing corner determination according to mobility variation. . . . .	45
Figure 31	Timing corner with TSV stress. . . . .	45
Figure 32	Extended timing corner with both TSV and STI stresses. . . . .	46
Figure 33	Inverter delay variation with different $(\Delta\mu/\mu)_h$ and $(\Delta\mu/\mu)_e$ . (a) Rising delay dependency on $(\Delta\mu/\mu)_h$ , (b) Falling delay dependency on $(\Delta\mu/\mu)_e$ . . . . .	47
Figure 34	Design flow for TSV-stress-driven placement optimization. . . . .	48
Figure 35	Carrier-mobility-variation surface surrounding TSVs. . . . .	50
Figure 36	All forces applied to a cell. . . . .	52
Figure 37	Mobility-variation contour map for $22 \times 21$ TSV array. (a) hole, (b) electron. . . . .	54
Figure 38	Mobility-variation contour maps for a layout considering both TSV and STI stresses. (a) hole, (b) electron. . . . .	54
Figure 39	Gate perturbation to take advantage of TSV-stress-induced mobility variation. (a) hole-mobility contour with original gate placement, (b) hole-mobility contour after gate perturbation, (c) electron-mobility contour with original gate placement, (d) electron-mobility contour after gate perturbation. . . . .	60
Figure 40	Gate perturbation to take advantage of TSV-STI-stress-induced mobility variation. (a) hole mobility contour with original gate placement, (b) hole mobility contour after gate perturbation, (c) electron mobility contour with original gate placement, (d) electron mobility contour after gate perturbation. . . . .	62

Figure 41	Zoom-up snapshots of TSV-stress-driven placement results for ckt3 using 2-row TSV cells. . . . .	65
Figure 42	Die-to-die heat coupling from TSVs. TSVs are shown in white. The top die is closer to heatsink. The cold spot C is caused by the TSVs in spot A on the same die. The hot spot D is caused by the TSVs in spot B from the bottom die. . . . .	69
Figure 43	Structure, thermal conductivity, and thermal profile of bulk silicon with and without TSVs. Dark shade represents low thermal conductivity. . .	69
Figure 44	Design flow for the 3D IC global placement. . . . .	71
Figure 45	TSV spread and TSV align forces. . . . .	72
Figure 46	Thermal conductivity-based vs power density-based forces. . . . .	75
Figure 47	Illustration of $B_d^{\text{cond}}$ . (a) $P_d^{\text{cell}}$ , (b) $s_d^{\text{cond}} \cdot K_d^{\text{sink}}$ , (c) $B_d^{\text{cond}}$ , (d) potential for $B_d^{\text{cond}}$ after solving Poisson's equation. . . . .	76
Figure 48	Computation of $K_d^{\text{sink}}$ . (a) $K_j^{\text{die}}$ , (b) $K_1^{\text{sink}}$ . . . . .	77
Figure 49	Evaluation flow for temperature-aware 3D-IC global placement. . . . .	80
Figure 50	Power analysis flow for 3D ICs. . . . .	81
Figure 51	Analyzed structure of a TSV-based 3D IC. Each die is modeled with 15 layers of different materials. The entire 4-die structure contains 62 layers. . . . .	81
Figure 52	GDSII layout-level thermal analysis flow. . . . .	82
Figure 53	Material composition inside a thermal cell. . . . .	83
Figure 54	Temperature-Wirelength trade-off. . . . .	86
Figure 55	Power and thermal profile of designs with uniform [1] (left) and non-uniform [1] (right) TSV position. (TSVs are in white in the layout. Area with low power density or temperature is in blue.) . . . . .	90
Figure 56	Temperature of ckt3 placed by different placement algorithms. Die 1 is close to PCB, and Die 4 is close to heatsink. . . . .	91
Figure 57	Side view of a 3D IC. (a) with RDLs and (b) without RDLs. . . . .	94
Figure 58	RDL wires connecting TSVs on bottom die to bonding pads on top die. . . . .	95
Figure 59	Timing and power analysis flow for die-to-wafer stacked 3D ICs. . . . .	99
Figure 60	GDSII layout-level thermal analysis flow. . . . .	100
Figure 61	Layout of bottom die of the circuit in (a) TSV-farm, (b) TSV-distributed, and (c) TSV-whitespace styles. TSVs are in white. . . . .	101
Figure 62	RDL routing in TSV-whitespace style. . . . .	102
Figure 63	Temperature of bottom die of the circuit in (a) TSV-farm, (b) TSV-distributed, and (c) TSV-whitespace styles. . . . .	105

Figure 64	Stress of bottom die of the circuit with 10- $\mu$ m TSVs in (a) TSV-farm, (b) TSV-distributed, and (c) TSV-whitespace styles. . . . .	106
-----------	--	-----

## LIST OF SYMBOLS OR ABBREVIATIONS

<b>CA</b>	Thermal coupling-aware placement.
<b>CTE</b>	Coefficient of thermal expansion.
<b>FEA</b>	Finite element analysis.
<b>KOZ</b>	Keep-out zone.
<b>LPD</b>	Longest path delay.
<b>RDL</b>	Redistribution layer.
<b>SA</b>	Stress-aware.
<b>SD</b>	Stress-driven placer.
<b>STA</b>	Static timing analysis.
<b>STI</b>	Shallow trench isolation.
<b>TD</b>	Timing-driven placer.
<b>TNS</b>	Total negative slack.
<b>TSA</b>	TSV spread and alignment.
<b>TSV</b>	Through-silicon via.
<b>WL</b>	Wirelength.
<b>WLD</b>	Wirelength-driven placer.
<b>WNS</b>	Worst negative slack.

## SUMMARY

The objective of this research is to explore the feasibility of addressing the major performance and reliability problems or issues, such as wirelength, stress-induced carrier mobility variation, temperature, and quality trade-offs, found in three-dimensional integrated circuits (3D ICs) that use through-silicon vias (TSVs) at placement stage. Four main works that support this goal are included. In the first work, wirelength of TSV-based 3D ICs is the main focus. In the second work, stress-induced carrier mobility variation in TSV-based 3D ICs is examined. In the third work, temperature inside TSV-based 3D ICs is investigated. In the final work, the quality trade-offs of TSV-based 3D-IC designs are explored.

In the first work, a force-directed, 3D, and gate-level placement algorithm that efficiently handles TSVs is developed. The experiments based on synthesized benchmarks indicate that the developed algorithm helps generate GDSII layouts of 3D-IC designs that are optimized in terms of wirelength. In addition, the impact of TSVs on other physical aspects of 3D-IC designs is also studied by analyzing the GDSII layouts.

In the second work, the model for carrier mobility variation caused by TSV and STI stresses is developed as well as the timing analysis flow considering the stresses. The impact of TSV and STI stresses on carrier mobility variation and performance of 3D ICs is studied. Furthermore, a TSV-stress-driven, force-directed, and 3D placement algorithm is developed. It exploits carrier mobility variation, caused by stress around TSVs after fabrication, to improve the timing and area objectives during placement. In addition, the impact of keep-out zone (KOZ) around TSVs on stress, carrier mobility variation, area, wirelength, and performance of 3D ICs is studied.

In the third work, two temperature-aware global placement algorithms are developed. They exploit die-to-die thermal coupling in 3D ICs to improve temperature during placement. In addition, a framework used to evaluate the results from temperature-aware global

placements is developed. The main component of the framework is a GDSII-level thermal analysis that considers all structures inside a TSV-based 3D IC while computing temperature. The developed placers are compared with several state-of-the-art placers published in recent literature. The experimental results indicate that the developed algorithms help improve the temperature of 3D ICs effectively.

In the final work, three block-level design styles for TSV-based die-to-wafer bonded 3D ICs are discussed. Several 3D-IC layouts in the three styles are manually designed. The main difference among these layouts is the position of TSVs. Finally, the area, wirelength, timing, power, temperature, and mechanical stress of all layouts are compared to explore the trade-offs of layout quality.

# CHAPTER I

## INTRODUCTION

Modern lifestyles demand high functionality of electronic devices, but depend on high mobility. This demand never stops growing, whereas the high mobility constrains the form factor of the devices. The electronic industry has been relying on reducing feature size to put additional transistors into the devices. This approach, however, requires not only advanced researches in technology scaling but also high investment. Every new process generation makes this approach increasingly hard to be technologically and economically feasible. Three-dimensional (3D) integration is a viable approach that allows designers to add functionality to the devices while maintaining the same footprint without the need for new process generation. Stacking dies in 3D integrated circuits (ICs) also reduces the footprint of designs. In addition, footprint reduction helps decrease wirelength, thus improving the performance of the designs. Stacking memory dies and core dies together and using short, on-chip, and wide-I/O interconnections between them help increase the memory bandwidth and decrease the memory latency. Furthermore, dies from different processes, such as logic, memory, analog, and sensor, can be stacked in 3D ICs, enabling additional functionalities of the devices.

Stacking dies in 3D ICs requires through-silicon vias (TSVs) for interconnections between the dies. Because TSVs occupy silicon area of the dies, using too many TSVs increases die area, and thus diminishes wirelength reduction. The position of TSVs and logic gates must also be carefully determined so that the wirelength is minimized. In addition, after TSV fabrication process, tensile stress can build up in region surrounding TSVs. Besides TSVs, shallow trench isolation (STI) causes compressive stress on silicon surface. These mechanical stresses can change carrier mobility inside transistors, resulting in unpredictable change in performance of 3D ICs. Therefore, the position of TSVs and STIs relative to logic

gates must be carefully determined so that the performance of 3D ICs is not negatively impacted. Stacking thinned dies in 3D ICs also increases power density, and thus results in rising temperature. High temperature leads to other reliability problems. Because TSVs are usually filled with copper, they have high thermal conductivity. The position of TSVs must be carefully determined so that they help remove heat from high-power logic gates and reduce temperature of 3D ICs. It is clear that the position of TSVs and logic gates plays an important role in the quality of 3D ICs. A good placement result for one problem may be a bad placement result for another problem. This quality trade-offs must also be considered while determining the position of TSVs and logic gates.

Placement is one of the most important stages in physical design because it is performed early in the design flow. Altering the layout of an IC, i.e., changing position of logic gates after placement, can dramatically affect performance and reliability of the IC. TSVs used in 3D ICs pose additional challenges to placement because of their physical, mechanical, and thermal impacts on the 3D ICs. The problems and issues in 3D ICs mentioned above create the need for research in placement for fast and reliable TSV-based 3D-IC layouts.

### ***1.1 Objective***

The objective of this research is to explore the feasibility of addressing the major performance and reliability problems or issues, such as wirelength, stress-induced carrier mobility variation, temperature, and quality trade-offs, found in 3D ICs that use TSVs at placement stage. In this research, the impact of placement on the major performance and reliability problems or issues found in TSV-based 3D ICs is studied. As outcome of this research, new wirelength-driven, TSV-stress-driven, and temperature-aware placement algorithms for TSV-based 3D ICs are developed.

### ***1.2 Contributions***

In this dissertation, a set of placement algorithms that address the major performance and reliability problems or issues found in 3D ICs and the studies of the impact of placement on the problems are presented. The main contributions of this research are as follows:



- **A wirelength-driven placement algorithm for gate-level design of 3D ICs:** A force-directed and 3D placement algorithm is developed. The algorithm takes the area occupied by TSVs into account while optimizing for wirelength during placement. Because the developed algorithm physically inserts TSVs into layouts, the layouts are realistic.
- **A study of the physical impact of TSVs on the layout of 3D ICs:** Based on fully validated GDSII-level layouts generated from the results obtained from the developed wirelength-driven placement algorithm, various experiments on the physical impact of TSVs on the layout of 3D ICs are performed.
- **A study of the impact of mechanical stress on the timing of 3D ICs:** A compact TSV-STI-stress-induced carrier-mobility-variation model and stress-aware, 3D, and static timing analysis (SA 3D STA) are developed. The SA 3D STA takes the carrier mobility change caused by TSV and STI stresses into account while analyzing timing of 3D ICs. Various experiments on the impact of mechanical stress on the timing of 3D ICs are performed. The impact of keep-out zone (KOZ) on stress, carrier mobility variation, and timing of 3D ICs is also evaluated. A demonstration on stress-aware performance optimization is provided by adjusting gate positions manually.
- **A TSV-stress-driven placement algorithm for gate-level design of 3D ICs:** A force-directed and 3D placement algorithm is developed to exploit hole and electron mobility variation caused by TSV stress for TSV-stress-aware performance optimization. Because the developed algorithm considers TSV stress during placement, it improves the TSV-stress-aware performance of the layouts significantly.
- **Two temperature-aware placement algorithms for gate-level design of 3D ICs:** Two effective heuristics that exploit the die-to-die thermal coupling in 3D ICs in force-directed temperature-aware placement are developed. By considering both power density and thermal conductivity inside 3D ICs during placement, both algorithms improve the temperature significantly. Results from the developed algorithms are compared with the results from several state-of-the-art placers.

- **A study of the impact of TSVs on the temperature of 3D ICs:** A framework that takes all structures inside a TSV-based 3D IC, including adhesive, TSV, landing pad, and liner, into account while computing temperature is developed. Extensive experiments are performed to show the trade-off among wirelength, delay, power, and temperature results obtained from GDSII layouts.
- **A study of design quality trade-offs of block-level placements of die-to-wafer bonded 3D ICs:** Several 3D-IC layouts, composed of two dies with different die sizes, are manually designed in three different block-level design styles. They are compared with respect to area, wirelength, timing, power, temperature, and mechanical stress.

### 1.3 *Organization*

This dissertation is organized into seven chapters as follows:

- **Chapter I:** In this chapter, the thesis of this dissertation is introduced, the contributions of this research are also summarized, and the organization of this dissertation is explained.
- **Chapter II:** In this chapter, problems or issues related to the research presented in this dissertation are described along with previous works related to the problems or issues.
- **Chapter III:** In this chapter, two different TSV-handling schemes, named TSV coplacement and TSV-site, for gate-level 3D-IC design are presented. In TSV coplacement scheme, gates and TSVs are simultaneously placed, whereas TSVs are placed at regular positions before placing gates in TSV-site scheme. The wirelength-driven placement algorithm for gate-level design of 3D ICs that supports both schemes is explained. Area, wirelength, and number of metal layers of 3D layouts are compared with the results of 2D layouts. In addition, the layouts designed in TSV coplacement and TSV-site schemes are compared.

- **Chapter IV:** In this chapter, an introduction on mechanical stress in 3D ICs, its impact on carrier mobility variation and timing, and KOZ are given. Then, related works and motivation are described. The modeling and design flow is presented. The model for carrier mobility variation caused by TSV and STI stresses is explained. A stress map according to TSV and STI positions is generated, and hole and electron mobility variations are estimated from the map. Timing analysis considering the stresses is then explained in detail. After that, TSV-stress-driven placement optimization and TSV-stress-driven global placement are described. The developed TSV-stress-aware 3D STA is performed during placement iterations to guide the placement algorithm. Finally, the experimental results are reported.
- **Chapter V:** In this chapter, a motivation is given by an example of die-to-die thermal coupling. The two temperature-aware global placement algorithms, named TSV spread and alignment method (TSA) and thermal coupling-aware placement (CA), are then explained in detail. A framework used to evaluate the results from temperature-aware global placements is described. The main components of the framework are power analysis and GDSII-level thermal analysis for 3D ICs. Finally, the experimental results, including the comparison with several state-of-the-art placers published in recent literature, are reported.
- **Chapter VI:** In this chapter, a background on the block-level placements of die-to-wafer bonded 3D ICs is given. Then, the three block-level design styles, named TSV-farm, TSV-distributed, and TSV-whitespace, are explained. The design evaluation, including traditional metrics, thermal analysis, and mechanical stress analysis, is then described. Several 3D-IC layouts in the three styles are manually designed. Finally, the area, wirelength, timing, power, temperature, and mechanical stress of all layouts are compared.
- **Chapter VII:** In this chapter, the researches presented in this dissertation are summarized, and concluding remarks are provided.

## CHAPTER II

### ORIGIN AND HISTORY OF THE PROBLEM

Three-dimensional integrated circuits (3D ICs) have received a great interest in the past recent years because of promising benefits, such as performance improvement and power reduction, they offer. Dies in the 3D-IC stack are usually thinned, and bonded together. Several bonding schemes are available. In most of the bonding schemes, TSV is the major component that connects signal, power/ground, and clock nets across the adjacent dies. Therefore, the position of TSVs is an important factor that determines the performance and reliability of 3D ICs. Many works have been proposed to address the performance and reliability problems during placement. Four broad categories of problems or issues related to the research presented herein are wirelength, stress-induced carrier mobility variation, temperature, and quality trade-offs.

#### **2.1 Wirelength**

Wirelength is a traditional performance metric for layouts obtained from placement algorithms. Long total wirelength indirectly represents high wire capacitive load to driving logic gates, thus low switching speed. Many placement algorithms have been proposed to optimize wirelength in 2D ICs. Current state-of-the-art are nonlinear placers and force-directed quadratic placers. Nonlinear placers [6, 7] use nonlinear optimization methods to minimize a nonlinear cost function for wirelength. Quadratic placers use a quadratic cost function for wirelength, which can be efficiently minimized by solving systems of linear equations. Minimizing only wirelength can result in high cell overlap. Force-directed quadratic placers [8, 9] introduce forces to the systems to reduce the overlap.

Stacking thinned dies in 3D ICs adds complexity to wirelength optimization. Inherently, it helps reduce wirelength [10] because of reduced footprint and vertical connections provided by TSVs. Although increasing number of TSVs used in a design results in decreasing wirelength, TSVs themselves require placement area. Typical size of TSVs ranges from

1  $\mu\text{m}$  to 20  $\mu\text{m}$  [11]. Because of TSV size, footprint area increases with the number of TSVs used in a design, indirectly diminishing wirelength reduction. Increasing number of TSVs beyond a certain point for a design results in wirelength increase [12].

Few works [13, 4] have been proposed to optimize wirelength in TSV-based 3D ICs. In [13], two major algorithms, folding and stacking, are used to transform 2D placement results to 3D placement results. The work starts from a result obtained from a wirelength-driven placer for 2D ICs, which should have optimized wirelength. To form an initial 3D placement result, the 2D placement result is folded in the first algorithm, whereas, in the second algorithm, cells are locally stacked after the 2D placement result is linearly shrunk. Further optimization is performed after transformation. In [4], a min-cut placer for 3D ICs recursively divides the netlist and placement area. The cut direction is determined by comparing the scaled length of the three dimensions of the placement area. The cost function used in this work is weight assigned to nets based on wire and TSV parasitics. Although TSVs are considered in both works, TSV size is neglected, leading to unrealistically high number of TSVs, which reduces validity of their results.

## ***2.2 Stress-induced Carrier Mobility Variation***

The performance of ICs depends on the output current of transistors, which in turn depends on mobility of holes and electrons. Carrier mobility depends on electrical factors, e.g., doping concentration and electric field [14], as well as mechanical factors, e.g., temperature [15] and stress [16]. Change in any of these factors results in carrier mobility variation, resulting in unpredictable logic-gate switching delay, thus unreliable operation of ICs. Doping concentration and electric field basically depend on device and process technology. Temperature depends mainly on IC operation and partly on IC design. Stress depends highly on the interaction of physical structures inside ICs. Engineered stress sources, such as stress liner, have been used to enhance carrier mobility of devices [17]; however, unintended stress sources, such as shallow trench isolation (STI), cause performance degradation in ICs.

STIs cause stress on silicon surface because of mismatch between coefficient of thermal expansion (CTE) of silicon dioxide ( $0.5 \times 10^{-6} \text{ K}^{-1}$ ), a widely used STI fill material, and

silicon ( $3 \times 10^{-6} \text{ K}^{-1}$ ). TSVs used in 3D ICs complicate the problem because they are usually filled with copper, which has much higher CTE ( $17 \times 10^{-6} \text{ K}^{-1}$ ) than both materials. Both STIs and TSVs are fabricated at high temperature. After cooling down to room temperature, silicon dioxide contracts slower than surrounding silicon, and pushes its surface, causing compressive stress in the area. On the other hand, copper contracts the fastest, and pulls silicon surface, causing tensile stress in the area [18]. Besides being opposite kinds, stress caused by both structures may interact with each other, resulting in unpredictable stress in 3D ICs. Because of its magnitude, stress caused by both structures is not negligible.

Few works [19, 20] have been proposed to consider stress-induced carrier mobility variation during placement. In [19], STI fabrication process is simulated to obtain mobility model for STI stress. The model is used to perform STI-stress-aware delay analysis of critical paths. A detail placement perturbation is used to improve the performance. In [20], stress caused by silicon-germanium source/drain is exploited to improve performance. Silicon-germanium source/drain is created by etching out source/drain regions of conventional transistor and filling them with silicon-germanium alloy. Large lattice constant of the alloy causes compressive stress in the channel. Placement is perturbed to facilitate sizing the area of source/drain so that the change in stress improves critical path delay. Stress caused by TSVs used in 3D ICs, however, has never been considered by any placement work. In addition, the interaction of stress caused by TSVs and other stress sources has never been studied.

### ***2.3 Temperature***

High operating temperature is a major cause that leads to other reliability problems. Problems related to mechanical reliability, such as delamination or cracking, are caused by thermal cycling and thermomechanical stress, resulting from mismatch between CTE of materials used in ICs and packages. Problems related to electrical reliability, such as electromigration [21] and negative-bias-temperature instability [22], are also accelerated by high operating temperature. Unfortunately, rising demand on IC functionality forces increasing

number of devices integrated in an IC, leading to power and temperature increase.

Stacking thinned dies in 3D ICs helps exacerbate the problem because it increases power density, and thus elevates chip temperature. In addition, polymer adhesive, a popular material used to bond thinned dies, worsens the situation because of its low thermal conductivity [23]. Moreover, if the thinned dies are silicon on insulator (SOI) [24], which is prevalent in the industry, extremely high temperature can be expected [25]. TSVs used in 3D ICs may help mitigate temperature increase because of their high thermal conductivity; however, the difference between thermal conductivity of TSVs and silicon can lead to another problem, thermal variation.

Several works [3, 2, 13, 4, 5] have been proposed to consider temperature during placement. In [3], a force based on power density is used in a force-directed quadratic placer to flatten power density and thus reduce temperature in 2D ICs. In [2], a force based on temperature is introduced to a force-directed quadratic placer to improve temperature. The algorithm alternates between performing finite element analysis (FEA) for temperature and solving linear equations for new placement. Although the work focuses on 3D ICs, it does not consider TSVs. In [13], a 2D placement result is transformed to an initial 3D placement result. Additional thermal optimization based on thermal resistive model is performed. To reduce temperature in [4], a min-cut placer for 3D ICs recursively divides the placement area, and cuts the netlist based on net weight that is calculated from switching activity and parasitics of wires and TSVs of each net. Although TSVs are considered in both [13] and [4], their thermal conductivity is not considered. In [5], TSV thermal properties are considered during placement; however, the work assumed that adhesive is an ideal insulator. In reality, heat can still flow through (silicon and) adhesive because of its thinness. Based on the assumption, the work balanced only the number of TSVs in a bin to heat dissipated from logic cells in the same bin and bins vertically below.

## **2.4 Quality Trade-offs**

The approaches to solve the problems mentioned earlier typically focus on only one aspect of the problems. Focusing on only one problem during placement may exacerbate the other

problems. For example, trying to optimize only wirelength may result in a placement with high mechanical stress because TSVs are placed too close to each other. It may also result in a placement with high temperature because high-power logic cells are placed close to each other but far from any TSV. Therefore, a good 3D-IC design must balance these quality trade-offs.

Many works [1, 26, 27] on global placement have been proposed for 3D-IC design. In these works, logic cells of a flattened netlist and TSV cells are placed together during 3D-IC design at gate level. Other works [28, 29, 30] on floorplanning have also been proposed for 3D-IC design. In these works, functional blocks and TSVs are floorplanned together during 3D-IC design at block level. Because block-level 3D-IC design allows the reuse of optimized blocks in current 2D ICs, designs under this methodology are likely to be early 3D-IC products on the market.

To commercialize 3D-IC products, the manufacturing process must allow for low cost and high yield for fabrication of 3D ICs. Dies inside a 3D-IC stack generally come from different processes that are optimized for different parts, e.g., logic, memory, analog, sensor. Therefore, these dies may come in different sizes. Out of the three bonding processes (die-to-die, die-to-wafer, and wafer-to-wafer), die-to-wafer bonding is the most suitable for this kind of 3D-IC integration. Die-to-die bonding provides high yield (by choosing only known-good dies for bonding), but it is an expensive process. Wafer-to-wafer bonding is a low-cost process (by bonding multiple dies in the wafers at the same time), but it may result in low yield if dies in the wafers are defective. Besides, it requires that all dies in the 3D-IC stack have the same size.

All design methodologies proposed in the above papers [1, 26, 27, 28, 29, 30] assume that all dies have the same footprint area. Therefore, they can not be used to design a die-to-wafer bonded 3D IC at block level. With scarcity of supporting design methodologies, the quality trade-offs of 3D ICs designed with different die sizes have not been well studied.



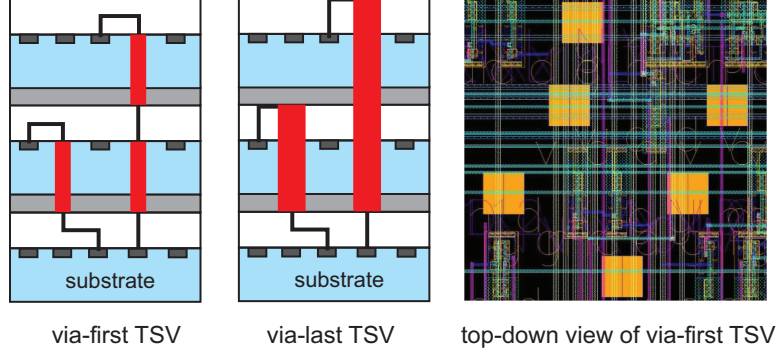
## CHAPTER III

### WIRELENGTH-DRIVEN PLACEMENT FOR GATE-LEVEL DESIGN OF 3D ICs

Three-dimensional (3D) integrated circuits (ICs) are emerging as a natural way to overcome interconnect-scaling problems in two-dimensional (2D) ICs. 3D ICs benefit from smaller footprint area than 2D ICs and from vertical (z-direction) interconnections between different dies [10, 12]. Small footprint area of 3D ICs allows gates to be placed close to each other, thereby leading to shorter wirelength than 2D ICs. Vertical interconnections by through-silicon vias (TSVs) also help shorten wirelength because gates can be placed on top of each other in different dies, eliminating the need of long cross-chip interconnects existing in 2D ICs. This short wirelength helps alleviate routing congestion as well as crosstalk and noise problems. Therefore, 3D ICs are expected to replace 2D ICs in the coming future.

Although TSVs can alleviate congestion, reduce wirelength, and improve performance, they occupy nonnegligible silicon area. Excessive or ill-placed TSVs not only increase die area, but they also have negative impact on these objectives in 3D ICs [12]. Therefore, computer-aided design (CAD) tools for 3D ICs should carefully consider the impact of TSVs during placement and routing. Depending on their type as shown in Figure 1, via-first TSVs interfere with device layer, whereas via-last TSVs interfere with both device and metal layers. A typical size of via-first TSVs ranges from  $1\mu\text{m}$  to  $5\mu\text{m}$ , whereas that of via-last TSVs ranges from  $5\mu\text{m}$  to  $20\mu\text{m}$  [11]. These TSVs are much larger than wires, local vias, and gates. Thus, care must be taken to consider the impact of TSV usage on the layout of each die in a 3D-IC stack. Most previous works on 3D-IC CAD tools [13, 4], however, ignore either the sheer size of TSVs or the fact that TSVs interfere with gates and wires.

In this work, a force-directed and 3D placement algorithm is developed. It can support two different TSV-handling schemes, namely “TSV coplacement” and “TSV-site.” In



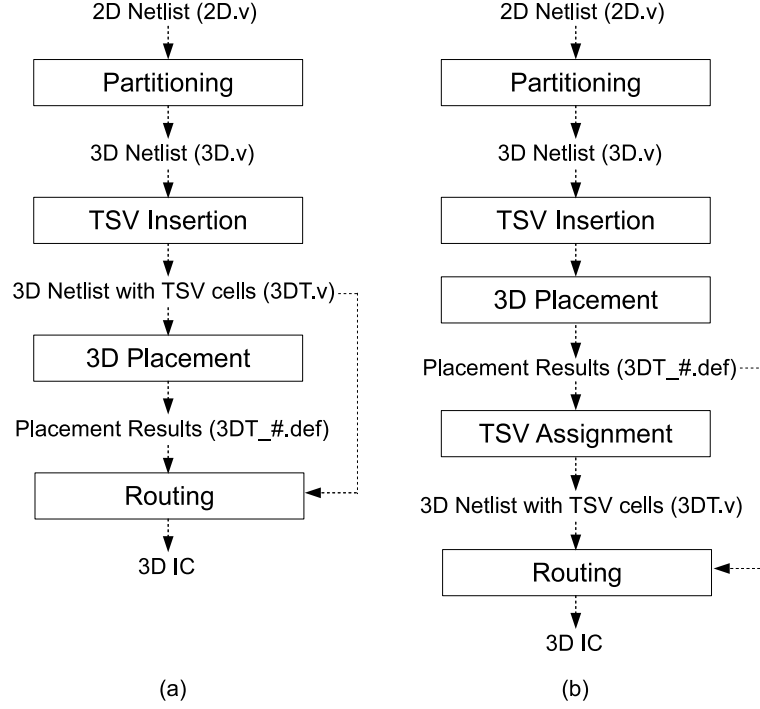
**Figure 1:** Via-first and via-last TSVs.

TSV coplacement scheme, TSVs and gates are simultaneously placed, whereas, in TSV-site scheme, TSVs are placed at regular positions before placing gates. Since many excellent 2D routers have been developed, they can be used to complete routing in 3D ICs. Using 2D routers in TSV coplacement scheme is easy because TSVs are inserted into the netlist, whereas using 2D routers in TSV-site scheme requires an additional step, which is “TSV assignment” [1]. The placement algorithm is integrated into a commercial tool. This new tool flow generates GDSII-level 3D-IC layouts that are fully validated. Based on these GDSII-level layouts, various studies are performed, and how TSVs affect 3D-IC layouts is demonstrated.

### 3.1 3D-IC Design Flow

Two 3D-IC design flows are devised for comparisons in this work, namely TSV coplacement and TSV-site, as shown in Figure 2. These flows are developed in such a way that existing 2D routing tools can be used while handling TSVs efficiently. By utilizing existing 2D routing tools, GDSII-level layouts of 3D ICs can be easily generated for in-depth analysis.

**Partitioning:** In the first stage of both design schemes, gates in the 2D netlist are distributed into  $N_{\text{die}}$  dies by a modified FM partitioning. During the partitioning, the cutsizes is controlled to obtain the desired number of TSVs. The output of this stage is the 3D netlist in which some of the 2D nets (nets having all their gates in a die) of the original design become 3D nets (nets having their gates in different dies). After partitioning is completed, the minimum number of TSVs to be inserted can be computed. Although



**Figure 2:** Two 3D-IC design flows developed in this work, (a) TSV coplacement and (b) TSV-site.

multiple TSVs can be used for a 3D net to connect gates in two adjacent dies, only one TSV is used for a 3D net between two adjacent dies in this work.

**TSV insertion and placement in TSV coplacement scheme:** In TSV coplacement scheme, TSVs are added into the 3D netlist during TSV insertion stage, and then TSVs and gates are simultaneously placed during 3D placement. The 3D placer is explained in Section 3.2. The output of the 3D placer is a DEF file for each die.

**TSV insertion and placement in TSV-site scheme:** In TSV-site scheme, TSVs are uniformly preplaced on each die in TSV insertion stage, and then gates are placed in 3D placement stage. During 3D placement, preplaced TSVs are treated as placement obstacles because a TSV should not overlap any gate. An additional stage, TSV assignment [1], is needed after 3D placement to determine which preplaced TSV belongs to which 3D net. Then, the 3D netlist is updated to reflect the assigned TSVs.

**Routing:** After the DEF and the netlist files for each die are generated, Cadence SoC Encounter [31] is used to route each die. Routing is separately performed on each die

because each die has its own netlist and cell (TSV or gate) positions. To facilitate TSV manipulation by Cadence SoC Encounter, a “TSV cell” is defined as if it is a standard cell.

### 3.2 3D Placement Algorithm

The 3D placement algorithm used in this work is based on a force-directed quadratic placement algorithm [9]. The algorithm is modified to place cells (TSVs or gates) in 3D.

#### 3.2.1 Overview of Force-directed Placement

In quadratic placement, a placement result is computed by minimizing the quadratic wirelength function  $\Gamma$ , which can be expressed as

$$\Gamma = \Gamma_x + \Gamma_y, \quad (1)$$

where  $\Gamma_x$  and  $\Gamma_y$  are wirelength along x- and y-axis. Because  $\Gamma_x$  and  $\Gamma_y$  are independent, they can be separately minimized to obtain the minimum of  $\Gamma$ . The following description for x-dimension similarly applies to y-dimension. Here,  $\Gamma_x$  can be written in a matrix form as

$$\Gamma_x = \frac{1}{2} \mathbf{x}^T \mathbf{C}_x \mathbf{x} + \mathbf{x}^T \mathbf{d}_x + \text{constant}, \quad (2)$$

where  $\mathbf{x} = [x_1 \cdots x_N]^T$  is a vector representing the x-position of  $N$  cells being placed,  $\mathbf{C}_x$  is an  $N \times N$  matrix representing the connection among the cells along x-axis, and  $\mathbf{d}_x = [d_{x,1} \cdots d_{x,N}]^T$  is a vector representing the connection to fixed pins along x-axis. Element  $c_{x,ij}$  of matrix  $\mathbf{C}_x$  is the weight of connection between cell  $i$  and cell  $j$ , and element  $d_{x,i}$  is the negative weighted position of fixed pins connected to cell  $i$ . The minimum of  $\Gamma_x$  can be obtained by setting its derivative to zero. Therefore, the cell placement along x-axis is computed by solving

$$\mathbf{C}_x \mathbf{x} + \mathbf{d}_x = \mathbf{0}. \quad (3)$$

Quadratic placement can be viewed as an elastic spring system when  $\Gamma$  is treated as the total spring energy of the system. Because the derivative of a spring energy is a force, the derivative of  $\Gamma_x$  in Equation (2) can be viewed as a net force  $\mathbf{f}_x^{\text{net}}$  as

$$\mathbf{f}_x^{\text{net}} = \nabla_x \Gamma_x = \mathbf{C}_x \mathbf{x} + \mathbf{d}_x, \quad (4)$$

where  $\nabla_x = [\partial/\partial_{x_1} \cdots \partial/\partial_{x_N}]^T$  is the vector differential operator. At equilibrium,  $\mathbf{f}_x^{\text{net}}$  is zero, resulting in minimum  $\Gamma_x$ , but cells can be crowded in few areas of the chip, resulting in high cell overlap.

Density-based force  $\mathbf{f}_x^{\text{den}}$  spreads cells away from high-cell-density area to low-cell-density area to reduce cell overlap. Density-based force in [9] is defined for 2D ICs. It is modified to support cell overlap removal in 3D ICs. The modification is explained in Section 3.2.3. Hold force  $\mathbf{f}_x^{\text{hold}}$  is used to decouple each placement iteration from the previous iteration. It cancels out net force that pulls cells back to the placement in initial iteration, and can be written as

$$\mathbf{f}_x^{\text{hold}} = -(\mathbf{C}_x \mathbf{x}' + \mathbf{d}_x), \quad (5)$$

where  $\mathbf{x}' = [x'_1 \cdots x'_N]^T$  is a vector representing the x-position of cells from the previous placement iteration. When no density-based force is applied, hold force holds cells being placed into their position.

Total force  $\mathbf{f}_x$  is the summation of net force, density-based force, and hold force. The total force is set to zero,

$$\mathbf{f}_x = \mathbf{f}_x^{\text{net}} + \mathbf{f}_x^{\text{den}} + \mathbf{f}_x^{\text{hold}} = \mathbf{0}, \quad (6)$$

to obtain the placement result with minimal wirelength and some cell overlap reduction for each placement iteration.

### 3.2.2 Overview of the 3D Placement Algorithm

The 3D placement algorithm is divided into the following three phases: initial placement, global placement, and detail placement.

In the first phase, the initial placement is computed by solving Equation (3). The initial placement result contains high cell overlap, which will be reduced in each global placement iteration in the second phase by introducing density-based force and hold force to Equation (6) and solving the equation.

Global placement continues until the amount of remaining cell overlap is low. Then, detail placement starts in the third phase to legalize the result from global placement using a greedy algorithm.

### 3.2.3 Placing Cells in 3D ICs

It is not possible to extend a 2D and force-directed placement algorithm to a 3D placement algorithm simply by adding z-axis variable in Equation (1). The reason is that all the fixed pins in 3D ICs are on the C4-bump side, resulting in placing all the cells at the same z-position in the initial placement [4], i.e.,  $\mathbf{z} = \mathbf{0}$ . In this work, the force-directed quadratic placement algorithm in [9] is extended by exploiting the fact that cells are already assigned to dies by the partitioner and not moving them across dies during placement. Therefore,  $\Gamma_z$  is not included in Equation (1), allowing the placer to focus on wirelength minimization along x- and y-axis.

Density-based force in [9] is modified to support placing cells in 3D ICs. Because cell overlap on all dies are different, density-based force for a cell is computed based on the cell overlap of the die on which the cell is being placed.

The placement problem is formulated as a global electrostatic problem by treating cell area as positive charge and chip area as negative charge. The placement density  $D$  on die  $d$  can be computed by

$$D(x, y) \Big|_{z=d} = D^{\text{cell}}(x, y) \Big|_{z=d} - D^{\text{chip}}(x, y) \Big|_{z=d}, \quad (7)$$

where  $D^{\text{cell}}(x, y) \Big|_{z=d}$  is the cell density at position  $(x, y)$  computed by using only cells being placed on die  $d$ , and  $D^{\text{chip}}(x, y) \Big|_{z=d}$  is the chip capacity scaled to match total area of cells being placed on the die.

After  $D$  is computed, placement potential  $\Phi$  can be obtained by solving Poisson's equation,

$$\Delta \Phi(x, y) \Big|_{z=d} = -D(x, y) \Big|_{z=d}. \quad (8)$$

The negative gradient of  $\Phi$  indicates in which direction and how fast the cell at that position should move. Density-based force is modeled by connecting cell  $i$  to its target point  $\hat{x}_i^d$  with a spring of spring constant  $\hat{w}_{x,i}^d$ . The target point is computed by

$$\hat{x}_i^d = x'_i - \frac{\partial}{\partial x} \Phi(x, y) \Big|_{(x'_i, y'_i), z=d}, \quad (9)$$

where  $x'_i$  is the x-position of cell  $i$  being placed on die  $d$  from the previous placement

iteration. Therefore, for cell  $i$ , density-based force  $f_{x,i}^{\text{den}} = \hat{w}_{x,i}^{\text{d}}(x_i - \hat{x}_i^{\text{d}})$ , where  $x_i$  is the x-position of cell  $i$  being placed. Density-based force  $\mathbf{f}_x^{\text{den}}$  is finally defined for 3D ICs by

$$\mathbf{f}_x^{\text{den}} = \hat{\mathbf{C}}_x^{\text{d}}(\mathbf{x} - \hat{\mathbf{x}}^{\text{d}}), \quad (10)$$

where  $\hat{\mathbf{C}}_x^{\text{d}}$  is a diagonal matrix of  $\hat{w}_{x,i}^{\text{d}}$ ,  $\mathbf{x} = [x_1 \cdots x_N]^{\text{T}}$  is a vector representing the x-position of  $N$  cells being placed, and  $\hat{\mathbf{x}}^{\text{d}} = [\hat{x}_1^{\text{d}} \cdots \hat{x}_N^{\text{d}}]^{\text{T}}$  is a vector representing the target x-position of the cells.

### 3.2.4 Placing TSVs in TSV Coplacement Scheme

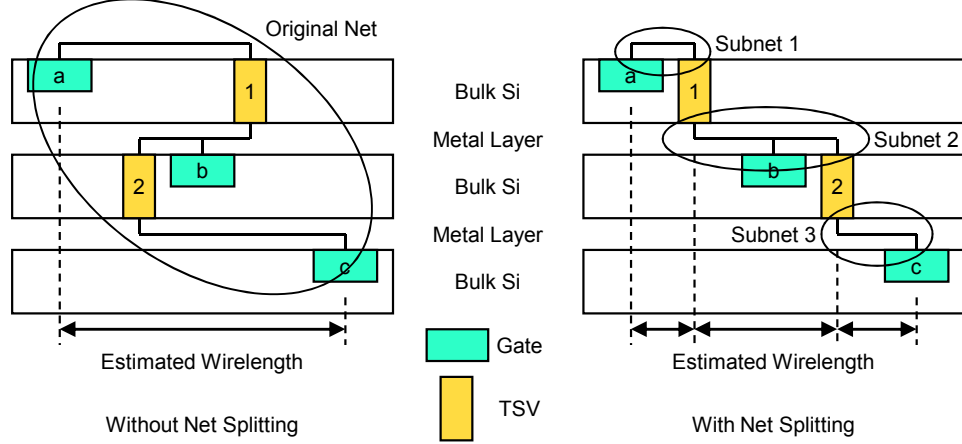
In TSV coplacement scheme, a TSV is treated as a cell being placed by the 3D placement algorithm. Therefore, it is called a TSV cell in this subsection, and an original cell in the design is explicitly called a gate cell. The 3D placement algorithm is modified to place TSV cells in TSV coplacement scheme. After adding the minimum number of TSV cells into the netlist, the total number of cells being placed is updated. The area of TSV cells is also used to compute  $D^{\text{cell}}(x, y)|_{z=d}$  and  $D^{\text{chip}}(x, y)|_{z=d}$  in Equation (7). The resulting vector  $\mathbf{x}$  obtained from solving Equation (3) and (6) includes the x-position of both TSV cells and gate cells.

### 3.2.5 Net Splitting

During wirelength computation, “net splitting” is used to compute accurate wirelength as shown in Figure 3. Wirelength computation without net splitting is based on the projection of the cell positions in all dies onto a 2D plane. On the other hand, wirelength computation with net splitting is based on the projection of the cell positions in each die onto its own 2D plane. Therefore, wirelength computation with net splitting gives more accurate wirelength estimation than wirelength computation without net splitting. The comparison of these two approaches is presented in Section 3.4.1.

### 3.2.6 Preplacing TSVs in TSV-Site Scheme

In TSV-site scheme, TSVs are preplaced in placement area before the original cells are placed. Therefore, they are treated as placement obstacles. Although the total number



**Figure 3:** Splitting a 3D net into subnets (side view).

of cells being placed is not updated, and the resulting vector  $\mathbf{x}$  obtained from solving Equation (3) and (6) still includes only the x-position of the original cells in the design, the area of preplaced TSVs is included when computing  $D^{\text{cell}}(x, y)|_{z=d}$  and  $D^{\text{chip}}(x, y)|_{z=d}$  in Equation (7).

TSVs are evenly preplaced as placement obstacles in rows and columns in this scheme. Placement obstacles can be naturally handled by the mean of placement density in [9]. By including the area of preplaced TSVs when computing placement density, density-based force is altered in such a way that it drives cells being placed away from preplaced TSVs.

### 3.3 TSV Assignment

The TSV assignment used in this work is adopted from [1]. TSV assignment problem is to assign 3D nets to TSVs for given sets of dies, 3D nets, placed gates, and preplaced TSVs so that the total wirelength of 3D nets is minimized. The constraints are as follows: (1) a TSV cannot be assigned to more than one 3D net, and (2) a 3D net should use one TSV between two adjacent dies.

For a 3D net spanning more than two dies in a 3D-IC stack, all combinations of TSVs on different dies assigned to the 3D net should be considered to find the optimum solution for TSV assignment. The number of possible combinations increases dramatically with the number of 3D nets. Although restricting TSVs assigned to a 3D net to a small window helps reduce the number of combinations, the solution space still grows exponentially with



the number of dies a 3D net spans. Therefore, two heuristic algorithms, minimum spanning tree (MST) and placement-based TSV assignments, were proposed in [1].

MST-based TSV assignment starts by constructing an MST for a 3D net, and then the nearest TSV to the shortest edge is chosen. The process continues for the next shortest edge until all TSVs of the 3D net are assigned. MST-based TSV assignment is a sequential (net-by-net) method. The order of 3D nets for assignment is important. In this work, 3D nets are sorted in the ascending order of bounding-box size because small bounding-box 3D nets have limited number of TSVs to choose.

Placement-based TSV assignment starts by treating placed gates and unassigned TSVs as fixed cells and movable cells, respectively, and then the problem is solved by using a placement algorithm in two steps. TSVs are placed by a force-directed quadratic method regardless of TSV-site positions in global assignment step. Then, each TSV is assigned (or snapped) to each TSV-site position in detail assignment step.

### 3.4 *Experimental Results*

IWLS 2005 benchmarks [32] and several industrial circuits as listed in Table 1 are used for 3D placement. TSV cell size of  $2.47\ \mu\text{m} \times 2.47\ \mu\text{m}$  and 45-nm technology are also used for experiments in this work.

**Table 1:** Benchmark circuits.

Circuit	# Gates	# Transistors	# Nets	Profile
Ind 1	16K	137K	12K	Microprocessor
Ind 2	15K	106K	15K	Inverse DCT
Ind 3	16K	134K	16K	Microprocessor
Ind 4	20K	146K	20K	Microprocessor
Ind 5	30K	317K	30K	Arithmetic Unit
ethernet	77K	729K	77K	Ethernet IP Core
RISC	88K	775K	89K	Microprocessor
b18	104K	728K	104K	Microprocessor Cores
des_perf	109K	823K	109K	DES (Data Encryption Standard)
b19	169K	1.29M	169K	Microprocessor Cores

### 3.4.1 Effectiveness of Net Splitting

In the first experiment, the effectiveness of net splitting for TSV coplacement scheme is studied. The wirelength of 3D placement results from TSV coplacement scheme without net splitting and with net splitting are compared in Table 2. Although TSV coplacement without net splitting is better than TSV coplacement with net splitting for two circuits, TSV coplacement with net splitting is generally better than TSV coplacement without net splitting. The average improvement is 5.59 %. The reason that TSV coplacement with net splitting generates shorter wirelength than TSV coplacement without net splitting is that wirelength is estimated more accurately in a 3D view with net splitting than without net splitting. Therefore, the placer can reduce the total wirelength efficiently. For the rest of this work, net splitting is used for wirelength estimation in TSV coplacement scheme.

**Table 2:** Wirelength from TSV coplacement scheme with and without net splitting.

Circuit	Without net splitting ( $\mu\text{m}$ )	With net splitting ( $\mu\text{m}$ )	Difference
Ind 1	444,867	408,713	-8.13 %
Ind 2	309,936	288,143	-7.03 %
Ind 3	305,961	308,006	+0.67 %
Ind 4	405,010	393,215	-2.91 %
Ind 5	658,886	584,024	-11.36 %
ethernet	1,538,792	1,406,073	-8.62 %
RISC	2,225,730	2,025,187	-9.01 %
b18	2,610,358	2,683,424	+2.80 %
des_perf	2,362,977	2,199,149	-6.93 %
b19	4,612,405	4,364,694	-5.37 %
Average			-5.59 %

### 3.4.2 Wirelength and Runtime Comparison

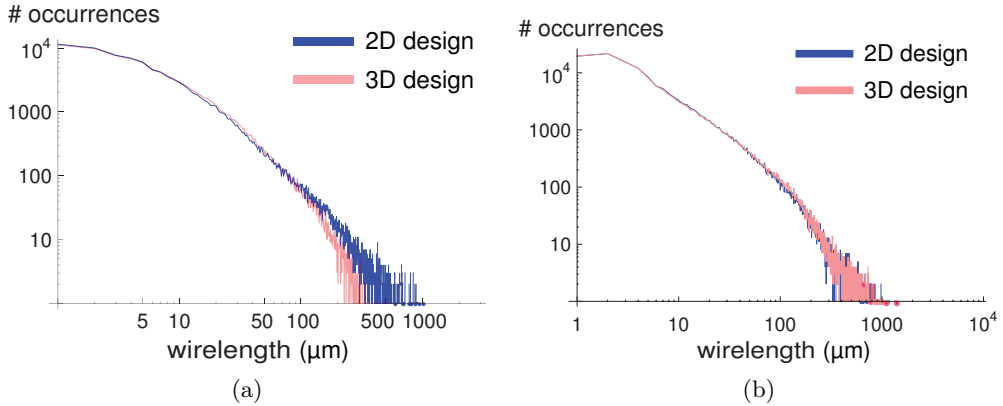
Wirelength of 2D and 3D placement results and runtimes are shown in Table 3. The wirelength reduction for nonmicroprocessor circuits is 10 % to 20 % in 3D design; however, microprocessor circuits can not benefit from 3D design in terms of wirelength.

To find the causes of discrepancy in wirelength reduction of the two circuit types, the wirelength distribution of des\_perf, a nonmicroprocessor circuit, and b19, a set of microprocessors, are plotted in Figure 4. As shown in Figure 4, long interconnections of des\_perf in

**Table 3:** Comparison of wirelength (WL) and runtime for placement for IWLS 2005 benchmarks and industrial circuits. Cell occupancy is 80 %, and the number of 3D nets is set to 3 % to 5 % of the number of total nets during partitioning. The numbers in parentheses are ratios to 2D.

Circuit	2D Design		3D Design	
	WL ( $\mu\text{m}$ )	Runtime (s)	WL ( $\mu\text{m}$ )	Runtime (s)
Ind 1	397,015 (1.0)	85 (1.0)	399,924 (1.01)	93 (1.10)
Ind 2	334,648 (1.0)	72 (1.0)	284,340 (0.85)	53 (0.73)
Ind 3	287,587 (1.0)	71 (1.0)	300,781 (1.05)	81 (1.14)
Ind 4	411,993 (1.0)	157 (1.0)	388,315 (0.94)	101 (0.64)
Ind 5	703,461 (1.0)	189 (1.0)	582,603 (0.83)	188 (1.00)
ethernet	1,534,386 (1.0)	1,289 (1.0)	1,401,059 (0.91)	1,287 (1.00)
RISC	1,976,549 (1.0)	880 (1.0)	2,001,986 (1.01)	727 (0.83)
b18	2,415,867 (1.0)	1,459 (1.0)	2,683,424 (1.11)	1,134 (0.78)
des_perf	2,445,398 (1.0)	1,367 (1.0)	1,911,731 (0.78)	950 (0.69)
b19	3,986,586 (1.0)	2,642 (1.0)	3,945,515 (0.99)	2,173 (0.82)

2D design are shortened in 3D design. The longest wire of des\_perf in 2D design is about 1000- $\mu\text{m}$  long, whereas the longest wire in 3D design is about 320- $\mu\text{m}$  long. This effect obviously comes from small footprint area compared with 2D design and connections in z-direction.



**Figure 4:** Wirelength distribution of (a) des\_perf, where the die width is 572  $\mu\text{m}$  in 2D design and 311  $\mu\text{m}$  in 3D design (4 dies), and (b) b19, where the die width is 762  $\mu\text{m}$  in 2D design and 411  $\mu\text{m}$  in 3D design (4 dies).

On the other hand, long interconnections of b19 in 2D design are not shortened in 3D design. Since partitioning is used as a preprocess for 3D placement, the cut size of the 4-way min-cut partitioning can be counted. The cut size of des\_perf is 1,613 (1.47 %) out of 109,415 nets, whereas the cut size of b19 is only 253 (0.15 %) out of 169,470 nets. This

small cut size indicates that b19 is so highly modularized that the total wirelength cannot be reduced much if min-cut partitioning is used.

As shown in Table 3, runtime of 3D placement is smaller than 2D placement. The reason is that, in each global placement iteration, 3D placement results have smaller number of cell overlaps than 2D placement results because each die in 3D ICs contains less number of cells to be placed than 2D ICs. Since a force-directed quadratic placement algorithm spends a significant portion of its runtime in removing overlaps, having reduced number of cells in a die improves runtime.

### 3.4.3 Metal Layers and Silicon Area

Since a 3D design has smaller footprint area than its 2D-design counterpart, and each die in a 3D design has less number of cells than the 2D design, the number of metal layers required for a 3D design can be smaller than that for the 2D design. Therefore, an attempt to find the minimum number of metal layers that leads to a successful routing is made. For fair comparisons, the cell occupancy is fixed at 80 %. The comparison of the minimum number of metal layers in 2D and 3D designs is shown in Table 4. With four metal layers, some of the 2D designs can not be routed because of congestion (design-rule-check errors), whereas all the 3D designs can be routed. The benefit of the decreased number of metal layers in 3D design comes from TSV insertion, which results in the unintentional increase of silicon area. The area increase in 3D designs are also shown in Table 4.

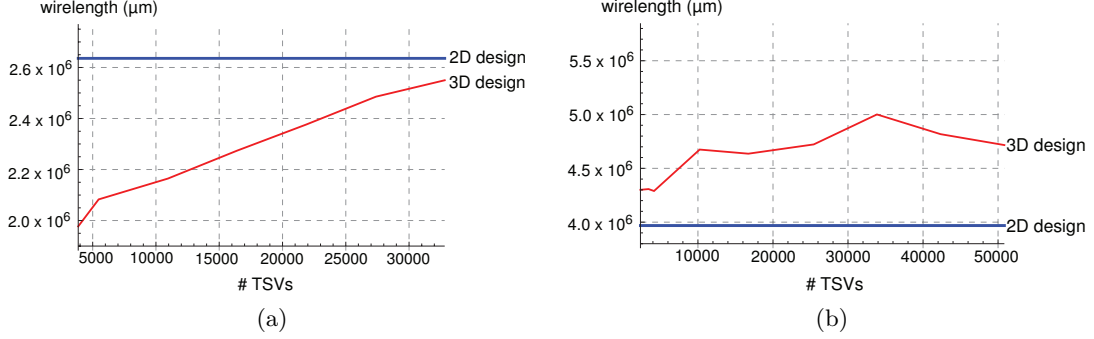
### 3.4.4 On Wirelength vs. Number of TSVs

In this experiment, the relationship between wirelength and the number of TSVs is studied. The results for des\_perf and b19 are shown in Figure 5. The wirelength of des\_perf in 3D design monotonically increases as the TSV count increases. This result indicates that the additional TSVs do not help wirelength reduction much. They rather increase die area thereby increasing the wirelength. On the other hand, the wirelength of b19 in 3D design generally increases at first as the TSV count increases, but it saturates after all. Although a clear and obvious conclusion on the relationship between wirelength and the number of TSVs cannot be drawn from these observations, using too many TSVs will eventually

**Table 4:** Comparison of the minimum number of metal layers (ML) and total silicon area for 2D and 3D (4 dies) designs for IWLS 2005 benchmarks and industrial circuits. The numbers in parentheses are ratios to 2D.

Circuit	2D Design		3D Design		
	# ML	Area ( $\mu\text{m}^2$ )	# ML	Area ( $\mu\text{m}^2$ )	# TSVs
Ind 1	5	44,944 (1.0)	4	69,696 (1.55)	1,700
Ind 2	4	44,944 (1.0)	4	58,564 (1.30)	1,302
Ind 3	4	48,841 (1.0)	4	69,696 (1.43)	798
Ind 4	4	63,001 (1.0)	4	80,656 (1.28)	1,016
Ind 5	5	103,684 (1.0)	4	147,456 (1.42)	2,789
ethernet	4	293,764 (1.0)	4	341,056 (1.16)	3,866
RISC	4	314,721 (1.0)	4	386,884 (1.23)	4,438
b18	5	338,724 (1.0)	4	495,616 (1.46)	10,404
des_perf	5	327,184 (1.0)	4	386,884 (1.18)	3,856
b19	5	580,644 (1.0)	4	712,336 (1.23)	8,497

increase the die area, which will result in wirelength increase.

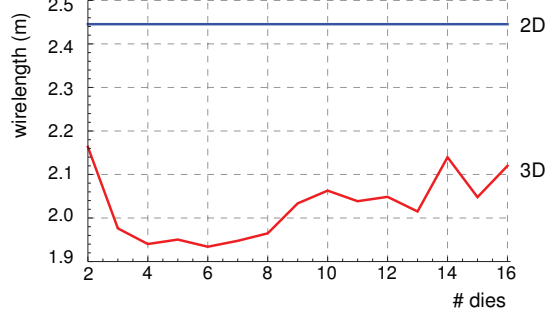


**Figure 5:** Wirelength vs. number of TSVs of (a) des\_perf and (b) b19 for 2D and 3D (4 dies) designs.

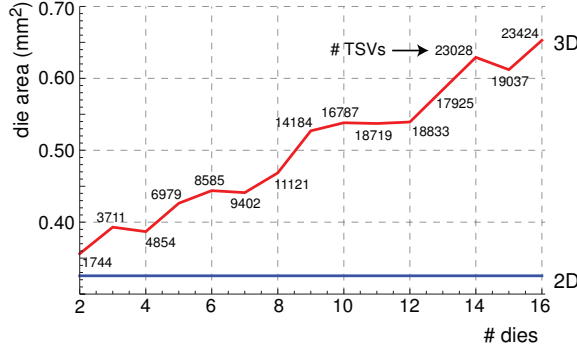
### 3.4.5 On Wirelength and Die Area vs. Number of Dies

In this experiment, the number of dies ( $N_{\text{die}}$ ) is varied from 2 to 16, and wirelength, die area, and the number of TSVs are observed. The wirelength of des\_perf in 3D design dramatically decreases as  $N_{\text{die}}$  increases up to four, then it saturates or slightly increases as shown in Figure 6. If  $N_{\text{die}}$  is increased further, the TSV count and die area will increase as shown in Figure 7. In other words, increasing  $N_{\text{die}}$  is helpful at first, but becomes not helpful after  $N_{\text{die}}$  increases beyond a certain point because 1) the TSV count increases, 2) the increased TSV count leads to the increase of die area, and 3) some of the 2D nets do

not need to become 3D nets. This trend may not be applicable to all the 3D designs. Using a small number of TSVs, however, is helpful if partitioning is used as a pre-process for 3D placement.



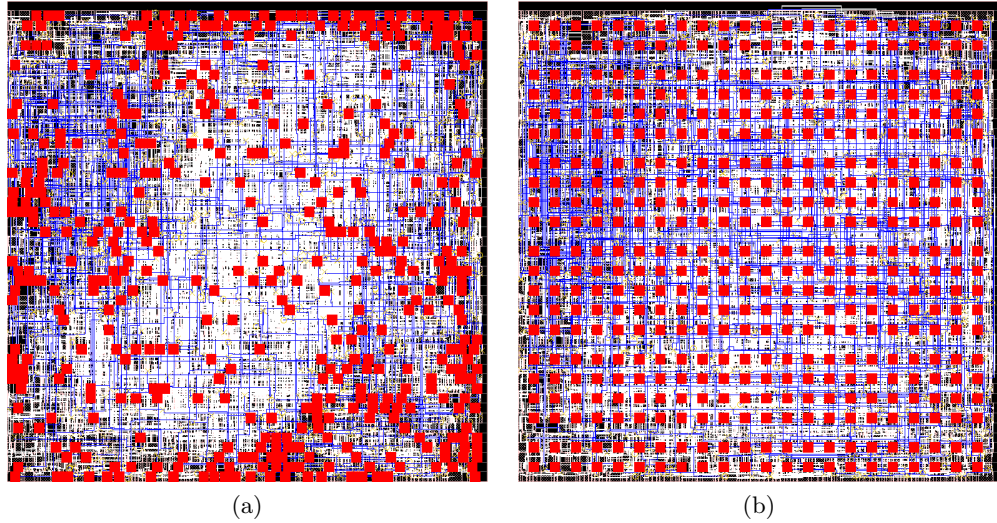
**Figure 6:** Wirelength vs. number of dies of des\_perf in 3D design.



**Figure 7:** Die area and number of TSVs of des\_perf in 3D design.

### 3.4.6 TSV Coplacement vs. TSV-Site Schemes

The placement and routing result of 3D placement in TSV coplacement and TSV-site schemes are shown in Figure 8. The comparison between TSV coplacement and TSV-site schemes is shown in Table 5. The wirelength increase of TSV-site scheme with MST-based TSV assignment compared to TSV coplacement is 8% to 15%, whereas the wirelength increase of TSV-site scheme with placement-based TSV assignment is 10% to 17%. Although TSV coplacement is better than TSV-site scheme with respect to wirelength, TSV-site scheme has its own advantages, which are “better heat dissipation and stronger package bonding” according to [33].



**Figure 8:** Cadence SoC Encounter snapshot of the bottommost die of Ind2 designed by (a) TSV coplacement and (b) TSV-site schemes. Routing for 3D nets is shown in blue.

**Table 5:** Comparison of wirelength of TSV coplacement and TSV-site scheme with MST-based TSV assignment and placement-based TSV assignment [1]. The numbers in the parentheses are ratios to TSV coplacement.

Circuit	Wirelength ( $\mu\text{m}$ )		
	TSV Coplacement	TSV-Site	
		MST-based	Placement-based
Ind 2	284,340 (1.0)	310,677 (1.09)	312,423 (1.10)
ethernet	1,401,059 (1.0)	1,513,381 (1.08)	1,554,960 (1.11)
des.perf	1,911,731 (1.0)	2,197,209 (1.15)	2,228,375 (1.17)

### 3.5 Summary

In this chapter, two 3D-IC design flows, TSV coplacement and TSV-site, are proposed. In TSV coplacement design flow, gates and TSVs are placed simultaneously, whereas, in TSV-site design flow, TSVs are uniformly preplaced, and gates are placed while the preplaced TSVs are treated as placement obstacles. A forced-directed placement algorithm for 2D ICs is extended to support placement for 3D ICs designed in both design flows. The proposed flows allow the study of the impact of TSVs on the 3D stacked IC layouts. The experimental results indicate that the proposed design flows place gates and TSVs effectively. Despite increasing die area for TSV insertion, the layout of 3D ICs generated by the proposed algorithms have shorter wirelength and use fewer metal layers than the layout of 2D designs. In addition, the layouts from TSV coplacement design flow have shorter wirelength than

the layouts from TSV-site design flow.



## CHAPTER IV

### IMPACT OF MECHANICAL STRESS AND PLACEMENT ON THE TIMING OF TSV-BASED 3D ICS

Stacking dies in 3D ICs requires through-silicon vias (TSVs) for interconnection between the dies. The fabrication process for TSVs causes tensile stress surrounding the TSVs. Because of the prevalence of shallow trench isolation (STI) in deep submicron technology, STI is also a major source of compressive stress in ICs. Mechanical stresses caused by these structures affect carrier mobility of transistors, thus leading to significant timing variation. Keep-out zone (KOZ) is a conservative way to prevent any devices/cells from being impacted by TSV stress. However, owing to already large TSV size, large KOZ can significantly reduce the placement area available for cells. Although, without control, both TSV and STI stresses may have negative impact on timing, they can actually be exploited for timing optimization because they are strongly layout-dependent, and their effect is systematic.

In this work, a compact TSV-STI-stress-induced carrier-mobility-variation model and stress-aware, 3D, and static timing analysis (SA 3D STA) are developed. First, a stress map according to TSV and STI positions is generated. Stress calculation is based on analytical model for TSV, a model developed from finite element analysis (FEA) simulation for STI, and linear superposition. The map is then used to estimate hole and electron mobility variation. During SA 3D STA, each gate is substituted by another gate having timing characteristics according to the estimated hole and electron mobility change caused by TSV and STI stresses. How TSV and STI stresses play an important role in performance optimization is then demonstrated by adjusting gate positions manually.

In addition, a placement algorithm is proposed to exploit hole and electron mobility variation caused by TSV stress. Carrier-mobility-based forces are introduced to a force-directed, 3D, and gate-level placement algorithm, and how to balance them against original placement forces is described. The placement algorithm is integrated into commercial tools.

The design flow enables trial or detail routing, parasitic extraction, and, finally, TSV-SA 3D STA to be performed on GDSII-level 3D-IC layouts. The accurate information on critical paths and critical nets/gates on them is used to guide the placer. Using the above mentioned design flow, the impact of KOZ on stress, carrier mobility variation, area, wirelength, and performance of 3D ICs is also studied.

#### **4.1 Introduction**

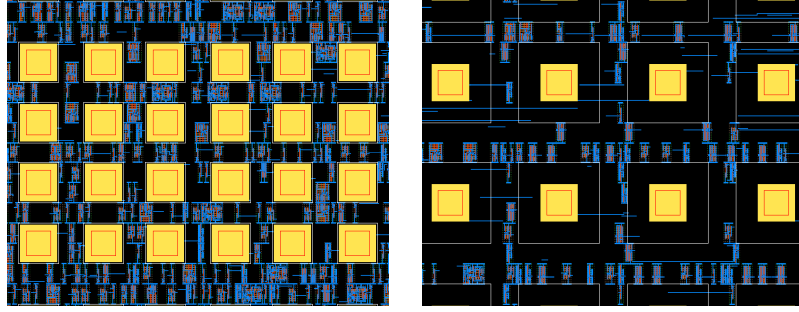
TSV fabrication causes tensile mechanical stress around TSVs because of the mismatch in the coefficient of thermal expansion (CTE) between silicon ( $3 \times 10^{-6} \text{ K}^{-1}$ ) and copper ( $17 \times 10^{-6} \text{ K}^{-1}$ ), a widely used material for TSV fill [34]. After cooling down from copper electroplating and annealing temperature to room temperature, copper contracts much faster than surrounding silicon, and pulls its surface, causing tensile stress in the area [18]. Severe stress can result in cracking and damage in substrate and devices on top [35]. Moreover, stress causes hole and electron mobility variation in devices, which can result in performance degradation without proper control. Longitudinal (with respect to carrier flow) tensile stress reduces hole mobility, whereas transverse tensile stress increases the mobility [17]. If a PMOS on a timing critical path experiences longitudinal tensile stress from TSVs, it can cause unexpected setup and hold time violation.

Another major stress source in ICs is shallow trench isolation (STI). The CTE of silicon dioxide, the widely used material for STI fill, is  $0.5 \times 10^{-6} \text{ K}^{-1}$  at  $20^\circ\text{C}$ . Because it is lower than the CTE of both silicon and copper, STI causes compressive stress on active region it surrounds. Oxidation and oxide densification for STI also take place at much higher temperature than TSV annealing temperature. Therefore, the compressive stress caused by STIs is also not negligible. Longitudinal compressive stress enhances hole mobility, but degrades electron mobility. If an NMOS on a critical path experiences compressive stress, it can also cause unexpected setup and hold time violation as well.

Even though several papers have been published regarding the impact of TSV stress [36] or STI stress [37] on performance of IC layouts, their impact is studied separately. The impact of combined stresses on performance is studied in this work. Because TSV/STI

stresses are layout dependent, a design flow is proposed to analyze timing variation caused by both stresses and show its implications for layout optimizations during 3D-IC design.

Traditionally, to avoid the impact of TSV stress on carrier mobility variation, keep-out zone (KOZ) is introduced. KOZ is the area surrounding each TSV from which all gates must “keep out” so that they are not influenced by the TSV stress. To determine the size of KOZ in [38], the magnitude of stress caused by TSVs was studied, and analyzed. KOZ is usually large because it is defined such that TSV stress outside itself is under preset tolerance. In real designs, the presence of abundant TSVs in use already has tremendous impact on 3D-IC layout. As illustrated in Figure 9, large KOZ only worsens the situation because it reduces the TSV-stress-induced carrier mobility variation in surrounding gates at the cost of increasing die size.



**Figure 9:** Layout with small vs. large KOZ around TSVs. TSV landing pads are large yellow squares.

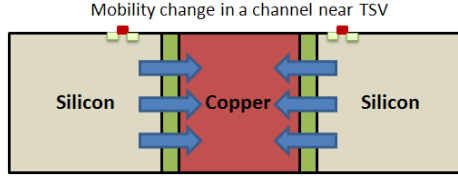
To reduce KOZ without adverse electrical effect, placers must also consider the effect of TSV stress on carrier mobility variation. Gates on critical paths must be placed in the position where the carrier mobility inside their p- and n-type metal-oxide semiconductors (PMOS/NMOS) is not degraded (if not enhanced) by TSV stress. Engineered stress has been widely used in industry to improve chip performance [17]. Few academic works also proposed placement perturbation techniques to use shallow trench isolation (STI) stress [37] and strained silicon [39] for the same purpose. By considering the effect of TSV stress on carrier mobility variation during placement, the necessity to keep large KOZ for electrical reason starts becoming obsolete.

## 4.2 Related Work and Motivation

The relation between stress and strain is shown in Equation (11). In the equation,  $E$  is Young's Modulus (169 GPa for silicon [40]),  $\sigma$  is the applied stress, and  $\epsilon$  is the deformation rate. For example, 169-MPa stress in silicon results in 0.1 % strain in silicon.

$$\sigma = E \times \epsilon. \quad (11)$$

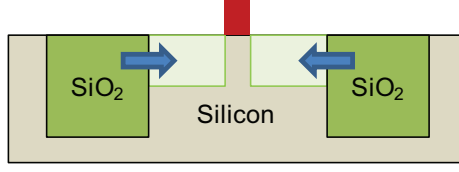
During 3D-IC manufacturing, stress is caused by CTE mismatch between copper TSV and silicon as shown in Figure 10. Investigations [41] indicate that, at 200 °C, an anneal time of 30-60 minutes is required to achieve reasonable copper layer properties. Since CTE of copper is larger than silicon, at room temperature, copper has less volume compared with that during annealing process because of contraction. Several papers have been published to simulate the TSV stress [34, 18] using finite element analysis (FEA) simulation. They show that TSV can cause tensile stress of more than 200 MPa.



**Figure 10:** Thermal stress around TSV.

Strained silicon has been used to enhance  $I_{on}$  of a transistor [42]. Unlike TSV stress, its impact on performance, however, is not layout dependent. Several unwanted stress sources are largely layout dependent, and should be considered during the design step. Shallow trench isolation (STI) is one of the unintentional stress sources [37, 43] because silicon dioxide used for STI fill pushes out silicon atoms near STI as shown in Figure 11. Silicon dioxide in STI is generally grown and densified at temperature as high as 1000 °C [44]. Since CTE of silicon dioxide is smaller than silicon, at room temperature, silicon dioxide contracts slower than silicon when cooling down to room temperature. Results from FEA simulation indicate that STI can also cause compressive stress of more than 200 MPa.

Mobility change  $\Delta\mu$  as a function of applied stress  $\sigma$  has been proposed by the following



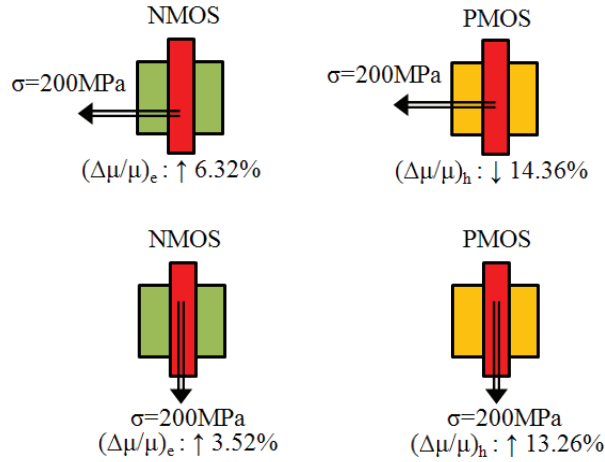
**Figure 11:** Thermal stress in active region caused by surrounding STIs.

equation [16]:

$$\frac{\Delta\mu}{\mu} = -\Pi \times \sigma, \quad (12)$$

where  $\Pi$  is the tensor of piezoresistive coefficient, and  $\sigma$  is the applied stress in silicon. Positive  $\sigma$  means tensile stress, whereas compressive stress is represented by negative  $\sigma$ .

Since tensile stress increases mean free path for electron, it enhances NMOS performance. However, longitudinal tensile stress degrades PMOS performance as shown in Figure 12(a) [45]. With longitudinal stress, piezoresistive coefficient for electrons is  $-3.16 \times 10^{-10} \text{ Pa}^{-1}$ , and the coefficient for holes is  $7.18 \times 10^{-10} \text{ Pa}^{-1}$  for (001) wafer surface and  $\langle 110 \rangle$  channel which are the most popular scheme for semiconductor manufacturing [46, 16]. For example, when TSV stress is 200 MPa,  $(\Delta\mu/\mu)_e$  is +6.32% for NMOS, and  $(\Delta\mu/\mu)_h$  is -14.36% for PMOS.

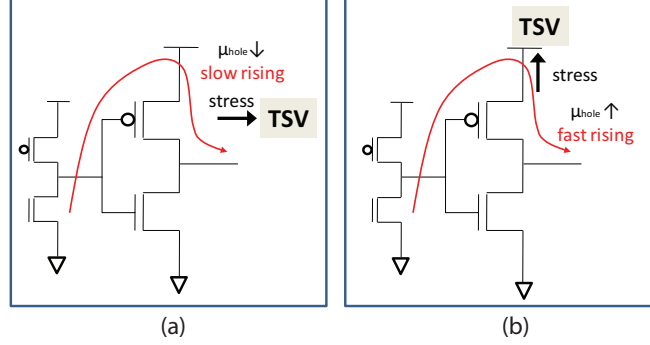


**Figure 12:** Mobility change due to tensile stress. Top:  $\Delta\mu/\mu$  for longitudinal tensile stress, bottom:  $\Delta\mu/\mu$  for transverse tensile stress.

However, if TSV is placed perpendicular to a transistor channel, mobility for both holes and electrons is enhanced by adding space in silicon lattice for carriers to move fast. For transverse stress, piezoresistive coefficient for electrons is  $-1.76 \times 10^{-10} \text{ Pa}^{-1}$ , and the

coefficient for holes is  $-6.63 \times 10^{-10}$  for (001) surface and  $\langle 110 \rangle$  channel. Similarly,  $(\Delta\mu/\mu)_e = +3.52\%$ ,  $(\Delta\mu/\mu)_h = +13.26\%$  can be expected with  $\sigma = 200$  MPa. Empirically, it is known that  $(\Delta I_{on}/I_{on})_{pmos}$  is 0.5~0.9 times of  $(\Delta\mu/\mu)_h$ , and  $(\Delta I_{on}/I_{on})_{nmos}$  is 0.4~0.6 times of  $(\Delta\mu/\mu)_e$  [47, 48] because  $I_{on}$  of a transistor is determined by the sum of source, drain, and channel resistance.

Transistor variation due to the stress can change cell timing characteristics. In Figure 13(a), buffer rising delay increases because of longitudinal tensile stress. Even though buffer size is the same in Figure 13(b), rising delay decreases in Figure 13(b) because of transverse tensile stress. Therefore, TSV-stress-aware timing analysis and layout optimization are essential steps for 3D-IC design.



**Figure 13:** Buffer cell delay change due to TSV stress. (a) slower rising delay with longitudinal tensile stress, (b) faster rising delay with transverse tensile stress.

Although stress in the direction perpendicular to a transistor channel can affect performance of the transistor, the major stress variation caused by STI is in horizontal direction. Because of standard cell structure, as suggested by [37, 43], STI stress in horizontal direction is the main STI stress variation that affects mobility. The trend for STI is also different from the trend for TSV because the stress caused by STI is compressive instead of tensile. Longitudinal compressive stress enhances PMOS performance, but degrades NMOS performance.

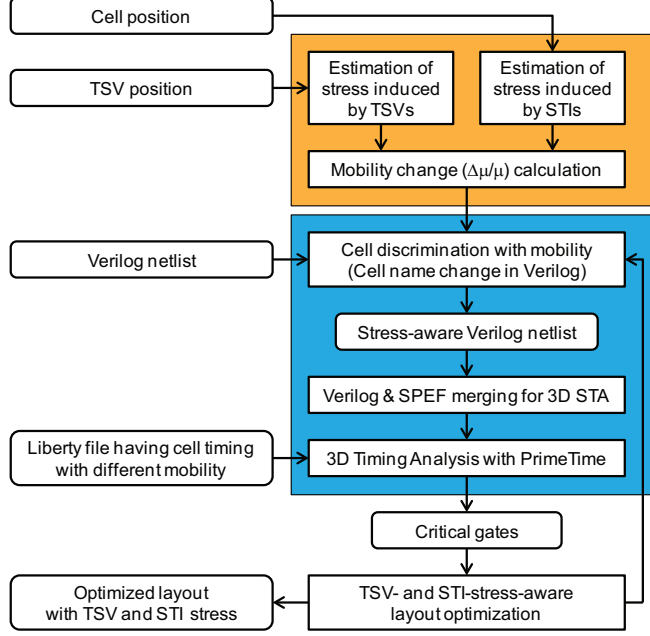
A transistor-level STI-stress-aware delay analysis was also proposed in [37]. The method is not suitable for combined stress-aware timing analysis and optimization for two reasons. First, in the paper, only mobility-variation model is provided after converting stress obtained

from TCAD simulation. Because TSV stress and STI stress interact with each other, stress from both structures should be combined before converted to mobility variation. Second, the method is based on SPICE simulation. Because of the amount of time required for simulation, the proposed method is suitable for only a small number of critical paths in a layout, thus limiting its application to the late stages of design flow. During early stages of design flow, e.g., global and detail placement, delay analysis may be required iteratively, and quick gate-level STI-stress-aware timing analysis is desirable.

### 4.3 *Modeling and Design Flow*

The overall flow of the 3D-IC design methodology is shown in Figure 14. The stress-driven design flow is consisted of three steps. The first step is to calculate TSV and STI stress and mobility change. Since FEA simulation which provides an accurate solution takes several hours to simulate stress for one TSV, the analytical model proposed in [18] is used. Mobility change can be calculated by extension of Equation (12). The process and device modeling for a single TSV is explained, and extended to consider multiple TSVs in Section 4.4.1. For STI stress, a model is developed from results obtained from FEA simulations of STI stress. Mobility change caused by STI stress can be calculated in the same way as mobility change caused by TSV stress. The process and device modeling for a single STI is explained, and extended to consider STIs on both sides of each cell in Section 4.4.2.

The second step is stress-aware, 3D, and static timing analysis (SA 3D STA). PrimeTime is used as a static-timing analysis (STA) engine. In Section 4.5, how to deal with verilog netlist and timing library to consider mobility variation is explained. The timing result can be used for layout optimization. Intuitively, if a PMOS in a cell is on a critical path, the cell should be moved to the region of a TSV that has positive  $(\Delta\mu/\mu)_h$ , or moved in such a way that surrounding STIs cause positive  $(\Delta\mu/\mu)_h$ . Finally, timing analysis can be run iteratively to verify the layout optimization effect.



**Figure 14:** Overall flow for TSV/STI stress modeling and analysis flow.

## 4.4 Carrier Mobility Variation

### 4.4.1 Mobility Variation under TSV-induced Stress

#### 4.4.1.1 Mobility Variation under Single TSV

In this work, cylindrical TSV shape, which is widely used for good manufacturability, is assumed. Finite element analysis (FEA) based TSV simulation has been proposed in [34, 18]. The simulation approaches provide an accurate solution with long runtime, which is not acceptable for design flows that calculate stress for several thousands of TSVs iteratively after each optimization. Therefore, an analytical model proposed in [18] is used. Assuming 2D and radial plain stress, the analytical solution, which is known as *Lamé* stress solution in [18], is expressed as follows:

$$\sigma_{rr} = -\frac{B\Delta\alpha\Delta T}{2} \left(\frac{R}{r}\right)^2. \quad (13)$$

The analytical stress model provides a relatively accurate solution [18]. In Equation (13),  $B$  is biaxial modulus,  $\Delta\alpha$  is CTE difference between copper and silicon,  $\Delta T$  is the temperature difference between copper annealing and operating temperature,  $R$  is TSV radius, and  $r$  is a distance from the center of a TSV. Here,  $\Delta T$  is assumed to be 250 °C, which is the case of 25 °C for the room temperature and 275 °C for the copper annealing temperature



which is relatively low annealing temperature [41]. The equation indicates that the thermal stress near a TSV depends on the ratio of TSV radius and a distance from the TSV center.

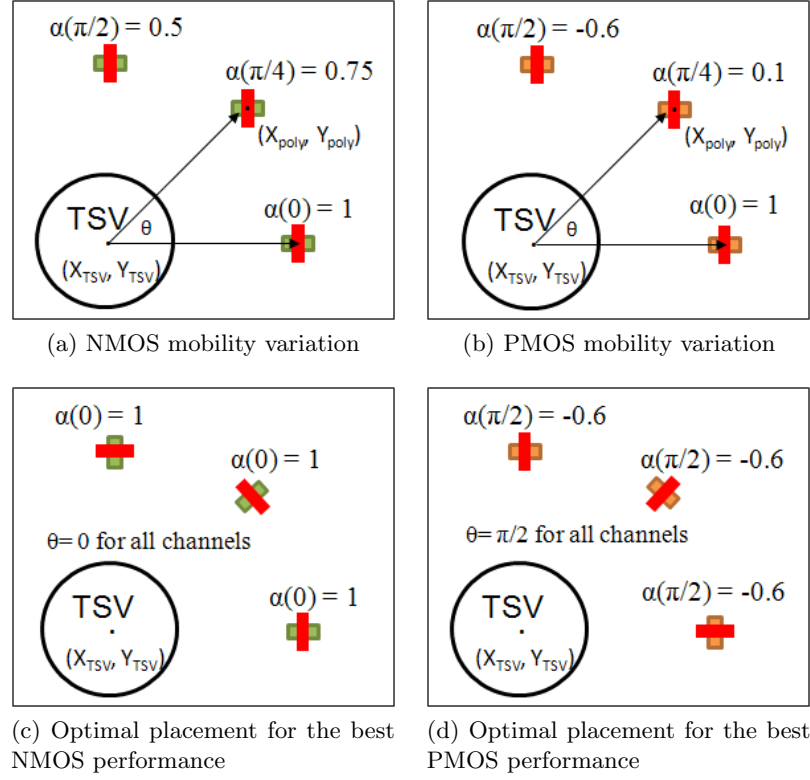
Equation (12) provides an efficient way to calculate mobility variation caused by  $\sigma_{rr}$ . As observed in Section 4.2, mobility change depends on not only  $\sigma_{rr}$  but also the orientation between applied force and a transistor channel. An empirical value for showing the relation between mobility change and a channel direction has been proposed in [45]. Equation (12) is extended to consider stress and channel direction as follows:

$$\begin{aligned}\frac{\Delta\mu}{\mu}(\theta) &= -\Pi \times \sigma_{rr} \times \alpha(\theta), \\ \theta &= \tan^{-1} \left| \frac{y_{\text{TSV}} - y_{\text{poly}}}{x_{\text{TSV}} - x_{\text{poly}}} \right|,\end{aligned}\tag{14}$$

where  $\alpha(\theta)$  is an orientation factor as a function of  $\theta$ , which is defined as the angle between a line connecting TSV center to a channel and the channel direction as shown in Figure 15(a) and (b). Here,  $\Pi$  is the piezoresistive coefficient at  $\theta = 0$ , which corresponds to longitudinal stress.

In Figure 15(a), if an NMOS is on the right side of a TSV,  $\theta$  becomes zero, and  $\alpha(0)$  becomes one, which enhances the NMOS mobility at its maximum. However, if an NMOS is on the top of a TSV,  $\alpha(\pi/2)$  is 0.5, which means that the NMOS mobility increase is half of the enhancement at  $\theta = 0$ . PMOS experiences the opposite trend, which has the best mobility enhancement at  $\theta = \pi/2$ . If  $\theta$  is zero, then, PMOS becomes slower than the case of no TSV stress. The transistor orientations for the best performance are shown in Figure 15(c) and (d). Even though mixing channel directions is not allowed because of patterning difficulty, the observation provides a way to optimize layout for 3D ICs.

A contour map for hole mobility variation is shown in the top of Figure 16. In the contour map, hole mobility decreases in a horizontal direction, whereas it increases in a vertical direction. In  $45^\circ$  direction, hole mobility does not change. Contour map for electron mobility variation is presented in the bottom of Figure 16. As shown in Figure 15(a), mobility enhancement zone in horizontal direction is larger than that in vertical direction.



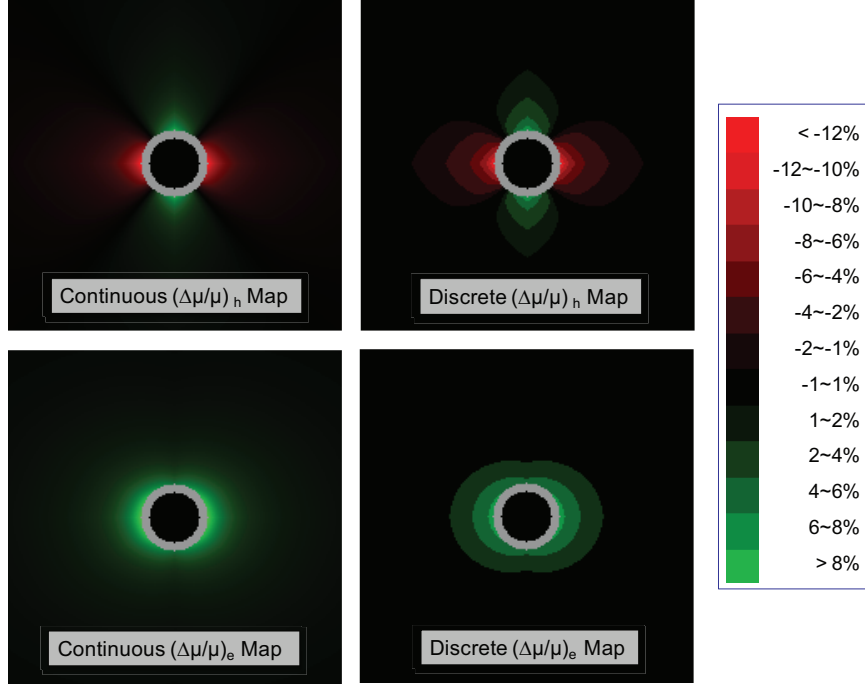
**Figure 15:** Optimal orientation of MOSFET to maximize mobility for (001) surface and  $\langle 110 \rangle$  channel.

#### 4.4.1.2 Mobility Variation under Multiple TSVs

Since many TSVs for signaling, power/ground, and clock network are used in 3D ICs, the variation model needs to be extended to stress the effect of multiple TSVs. Each TSV works as a stress source to silicon. When a position on a wafer is strained by multiple stress sources, linear superposition can provide the multiple-stress-source solution [18, 49].

An example of result from simulations of a cell with two TSVs nearby is shown in Figure 17. The contour of stress caused by each TSV is shown in Figure 17(a) and (b), respectively. When the stress caused by both TSVs is simulated, the stress from both TSVs linearly combines.

The horizontal stress caused by both TSVs on a horizontal line across the center of the cell is shown in Figure 18. An additional plot, linear superposition, is created by adding the stress individually caused by each TSV together. The result from this linear superposition is close to the stress caused by both TSVs. Therefore, it is possible to use linear superposition



**Figure 16:** Mobility contour map for a TSV. Top: contour map for hole mobility variation, bottom: contour map for electron mobility variation.

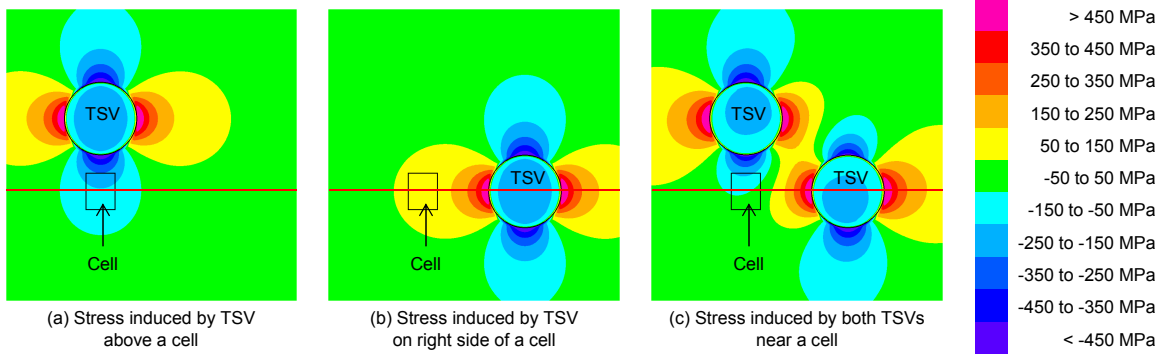
to estimate stress caused by TSVs nearby a cell.

The mobility variation for multiple TSVs is proposed as follows:

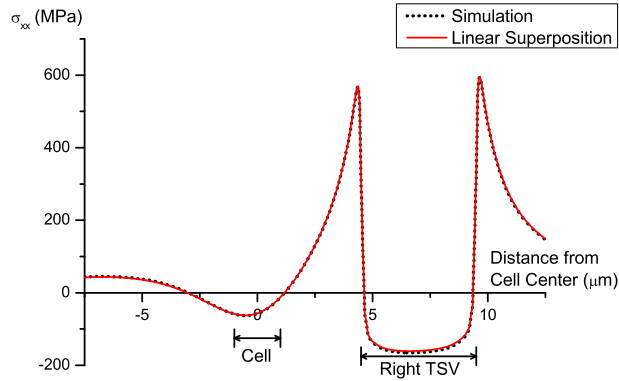
$$\frac{\Delta\mu}{\mu}_{\text{TSV}} = \sum \frac{\Delta\mu}{\mu}(\theta) = -\Pi \sum_{i \in \text{TSVs}} (\sigma_i \times \alpha(\theta_i)), \quad (15)$$

where  $\sigma_i$  is the tensile stress caused by  $i^{\text{th}}$  TSV,  $\alpha(\theta_i)$  is the orientation factor of  $i^{\text{th}}$  TSV, and  $\theta_i$  is the angle between the horizontal axis and a line connecting the center of  $i^{\text{th}}$  TSV to a point that mobility variation is calculated for.

In the top of Figure 19, the  $(\Delta\mu/\mu)_h$  for two different TSV placement schemes having the same TSV density is compared. Since zigzag TSV placement has compensation effect of positive and negative hole mobility changes between adjacent rows, it has more hole-mobility variation-free zone than regular TSV placement even if the mobility degradation effect within a row remains the same. In the bottom of Figure 19, electron-mobility contour for zigzag and regular TSV placement is shown. They do not have compensation effect. From Figure 19, zigzag TSV placement is preferred for small region of PMOS variation, whereas regular TSV placement is preferred for large hole-mobility enhancement zone.



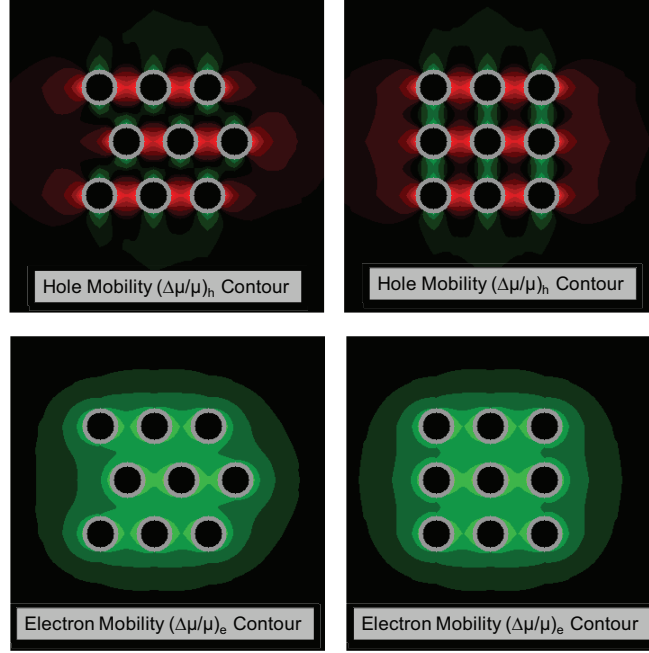
**Figure 17:** Contour of stress (FEA simulation) caused by TSVs nearby a cell.



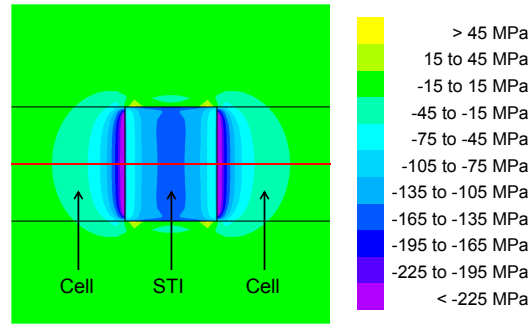
**Figure 18:** Linear superposition of stress (FEA simulation) caused by TSVs nearby a cell.

#### 4.4.2 Mobility Variation under STI-induced Stress

ANSYS [50], a commercial FEA-based simulator, is used to simulate stress caused by STI in this work. An example of simulation result is shown in Figure 20. In the figure, the contour of stress in horizontal direction caused by an STI in a plane of silicon is illustrated. Note that negative stress values represent compressive stress. Compressive stress caused by the STI can be higher than 100 MPa on silicon surface close to the STI, or even higher than 200 MPa on silicon surface adjacent to the STI. The contour lines in the area close to the STI are observed to be parallel to the left and right edges of the STI. Therefore, the magnitude of horizontal stress caused by an STI can be approximated as uniform stress in vertical direction. This approximation results in some error at a position far from an STI and off its center; however, the actual magnitude of the stress at the position is relatively small, and so is its impact on the mobility variation.



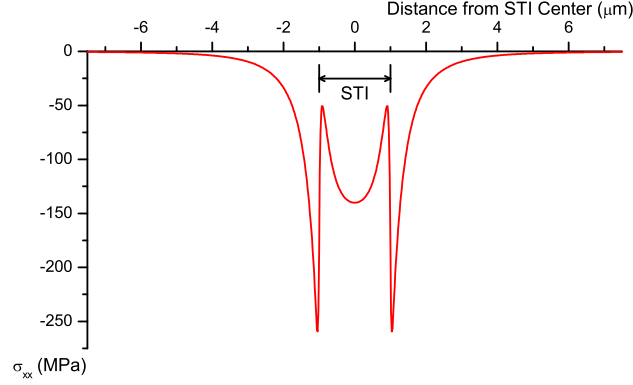
**Figure 19:** Zigzag TSV placement has small  $(\Delta\mu/\mu)_h$  between rows due to compensation.



**Figure 20:** Contour of stress (FEA simulation) caused by STI in horizontal direction.

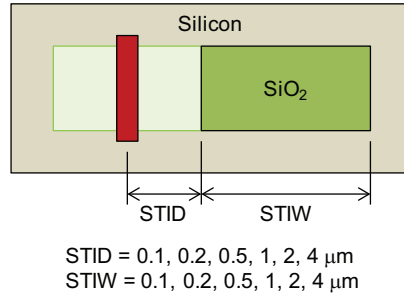
The horizontal stress, caused by the STI, on a horizontal line across the center of the STI is shown in Figure 21. The far left area of the STI is stress free. The stress magnitude increases as the distance from the center of the STI decreases, and rapidly increases in the area adjacent to the STI. Inside STI, the stress is still compressive although its magnitude drops sharply. The trend reverses when moving away from STI center to the right side of the STI.

The simulation setup used to develop a model for STI stress is shown in Figure 22. A patch of STI made of silicon dioxide is deposited on the surface of a silicon plane. STI stress mainly depends on two major parameters [37], distance to STI edge (STID) and



**Figure 21:** Stress (FEA simulation) on a horizontal line across the center of the STI in Figure 20.

width of STI (STIW). The values of these two parameters used for FEA simulations are listed under the figure. Other dimensions of the STI are from NCSU 45-nm cell library. The combinations of these two parameters result in 36 simulations in total. Because STI stress in horizontal direction is the main stress that affects mobility [37, 43], stress along x-axis  $\sigma_{xx}$  at the channel is measured.



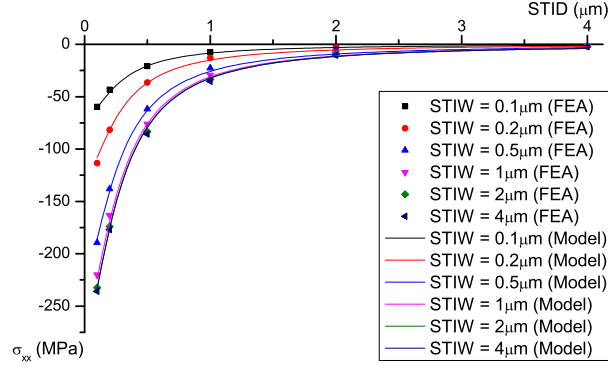
**Figure 22:** Setup for FEA simulations used to model STI stress.

The simulation results are shown in Figure 23 and 24. In Figure 23, the magnitude of STI stress rapidly decreases with the distance from the edge of STI. In Figure 24, the magnitude of STI stress rapidly increases with the width of STI initially, but does not change much after the width is higher than a certain value. The two observations lead to the model of STI stress in the following form:

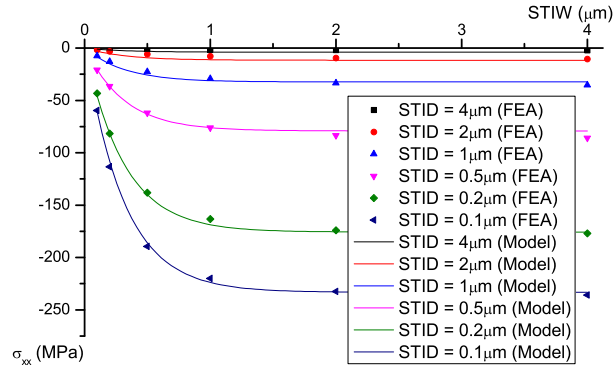
$$\sigma_{xx} = \frac{\alpha(1 - e^{\beta \cdot \text{STIW}}) + \chi}{\text{STID}^\delta + \epsilon} \quad (16)$$

where  $\alpha$ ,  $\beta$ ,  $\chi$ ,  $\delta$ , and  $\epsilon$  are curve-fitting constants, and their value are -37.51, -3.24, 0.8601,

1.594, and 0.1317, respectively. The coefficient of determination for this model is 0.9987, and the root mean square error is 2.843 MPa.



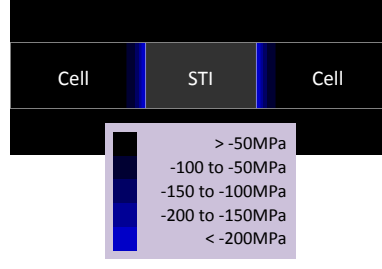
**Figure 23:** STI stress (FEA simulation and model) at different distances.



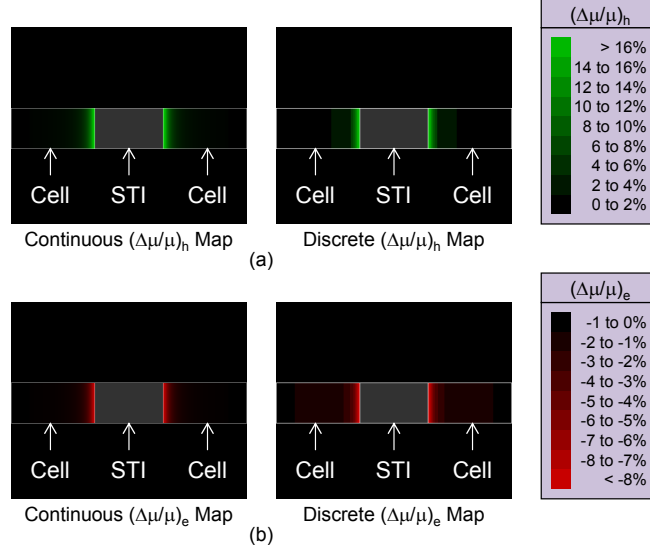
**Figure 24:** Stress (FEA simulation and model) induced by STI with different widths.

Based on the model, contour map of stress caused by an STI is generated. The contour for a  $4\mu\text{m}$ -wide STI is shown in Figure 25. In the figure, compressive stress of more than 200 MPa is close to the STI edge, but the stress magnitude rapidly drops below 100 MPa in horizontal direction. Note that the area occupied by STI is shown in gray in the contour. By using  $\sigma_{xx}$  obtained from Equation (16) in Equation (12), the contour map for hole and electron mobility variation can be generated, and is shown in Figure 26(a) and (b), respectively. From the contour, hole mobility is only enhanced by STI stress, whereas electron mobility is only degraded. Both hole mobility enhancement and electron mobility degradation take place on left and right sides of the STI.

Similar to the case for multiple TSVs in Section 4.4.1.2, it is possible to use linear



**Figure 25:** Contour map of stress (model) for a single 4μm-wide STI.



**Figure 26:** Contour maps of mobility (model) for a STI. (a) hole mobility variation, (b) electron mobility variation.

superposition to estimate stress caused by STIs on both left and right sides of a cell. The mobility variation for multiple STIs is proposed as follows:

$$\frac{\Delta\mu}{\mu}_{\text{STI}} = \sum \frac{\Delta\mu}{\mu} = -\Pi \sum_{i \in \text{STIs}} \sigma_{xx_i} \quad (17)$$

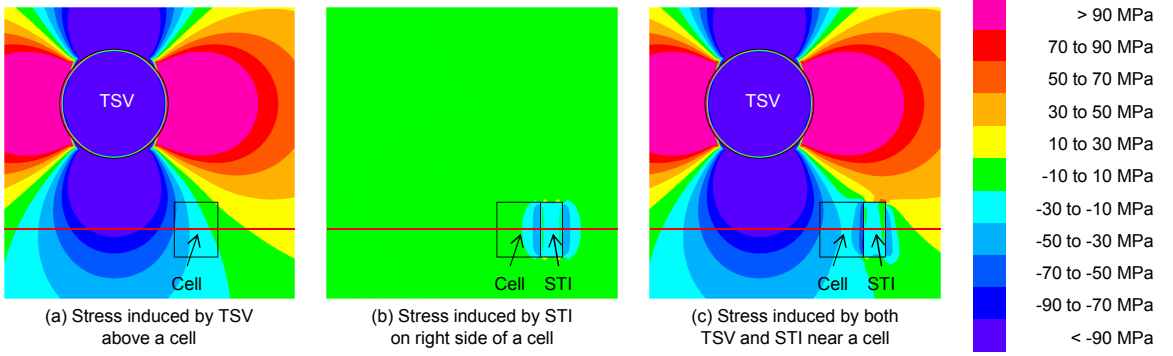
where  $\sigma_{xx_i}$  is the compressive stress caused by  $i^{\text{th}}$  STI on left or right side of the cell.

#### 4.4.3 Mobility Variation under Both TSV and STI-induced Stress

In 3D ICs, stress effect from both TSV and STI must be considered. An example of result from simulations of a cell with a TSV on top and an STI on its right side is shown in Figure 27. The contour of stress caused by the TSV on the top of the cell and the STI on the right side of the cell is shown in Figure 27(a) and (b), respectively. When the stress



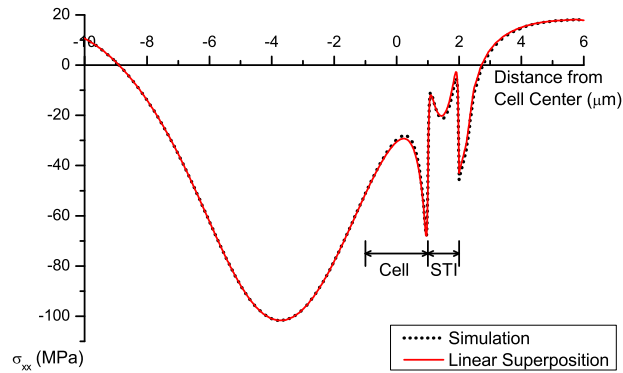
caused by both TSV and STI is simulated, the stress from both TSV and STI interact with each other, and results in change in stress contour in the cell.



**Figure 27:** Contour of stress (FEA simulation) caused by TSV on top of a cell with an STI on its right side.

The horizontal stress caused by the TSV and STI is shown in Figure 28. An additional plot, linear superposition, is created by adding the stress caused by the TSV and STI together. The result from this linear superposition is close to the stress caused by the TSV and STI obtained from FEA simulation. Therefore, it is possible to use linear superposition to estimate stress caused by both TSV and STI during early design stages. Then, the mobility variation for both TSV and STI is proposed as follows:

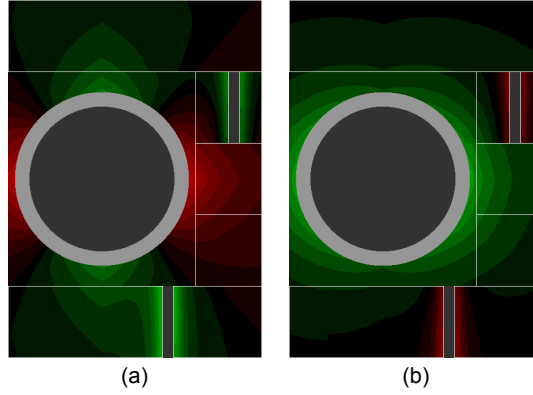
$$\frac{\Delta\mu}{\mu_{\text{total}}} = \frac{\Delta\mu}{\mu_{\text{TSV}}} + \frac{\Delta\mu}{\mu_{\text{STI}}} \quad (18)$$



**Figure 28:** Linear superposition of stress (FEA simulation) caused by a TSV and an STI.

The model allows study of the impact of the interaction between TSV and STI stress on circuit performance. The mobility-variation contour of a TSV and two STIs is shown in

Figure 29. As shown in Figure 29(a), the impact of STI stress can compensate the impact of TSV stress on hole mobility variation (the area between the TSV and the right STI), or even increase it (the area under the TSV and on the left of the bottom STI). On the other hand, the impact of STI stress can only reduce the improvement from the impact of TSV stress on electron mobility variation as shown in Figure 29(b).



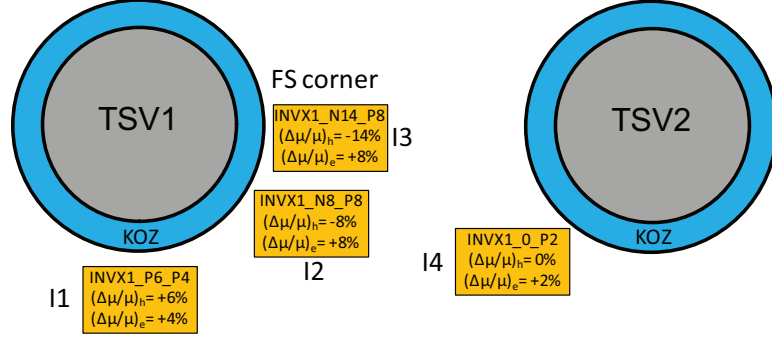
**Figure 29:** Impact of the interaction between TSV and STI stress (model) on mobility variation. (a) hole, (b) electron.

#### 4.5 Timing Analysis with Stress Consideration

In this section, how to incorporate the mobility variation into cell-level STA flow is explained.

##### 4.5.1 Timing Analysis for 3D ICs

Even though two gates have the same topology, their timing characteristic can be different depending on stress amount and stress direction to transistor channel. An example is shown in Figure 30. Gates having the same topology and size are in different timing corners systematically determined by TSVs (and STIs). When two TSVs are near three inverters, gate characteristics at different positions are different. For a given layout,  $\Delta\mu/\mu$  at any point can be determined by Equation (18). After mobility calculation, the proposed framework renames gates such that mobility variation is included in verilog netlist. For example, I2 is renamed to INVX1\_P-8\_N+8 , which means  $-8\%$  hole mobility and  $+8\%$  electron mobility in Figure 30.

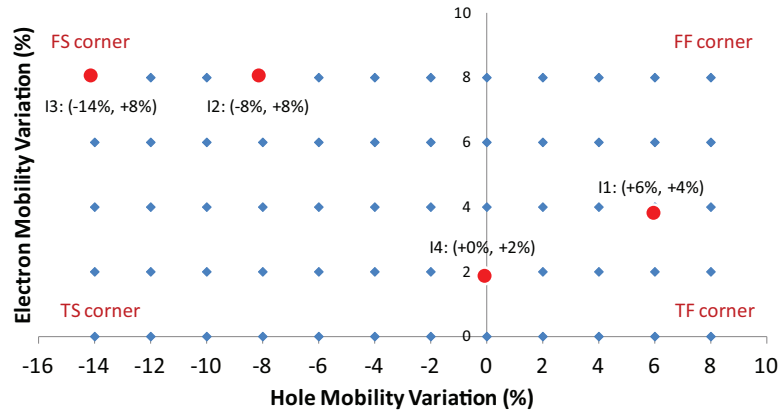


**Figure 30:** Timing corner determination according to mobility variation.

A verilog netlist and a parasitic extraction file (SPEF) for each die are prepared. In addition, a top-level verilog netlist is generated. It instantiates the dies in a 3D-IC stack, and connects them using wires, which correspond to TSV connections. Then, a top-level SPEF file for the TSV connections is generated. With a proper timing constraint file, PrimeTime [51] can be executed, and provide the results of SA 3D STA.

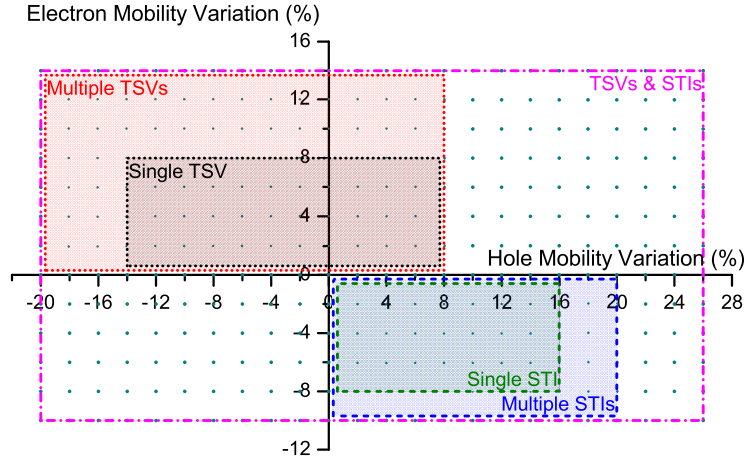
#### 4.5.2 Timing Library for Mobility Variation

To consider the systematic variation during timing analysis, a gate is characterized with different mobility corners as shown in Figure 31. Hole mobility variation ranges from  $-14\%$  to  $+8\%$ , and electron mobility variation ranges up to  $+8\%$  to cover stress caused by TSVs in Figure 30. Inverter I1 in Figure 30 matches the corner near FF corner, whereas I3 is in FS corner. With mobility-variation-aware library and verilog netlist containing renamed gates, PrimeTime performs timing analysis considering TSV stress.



**Figure 31:** Timing corner with TSV stress.

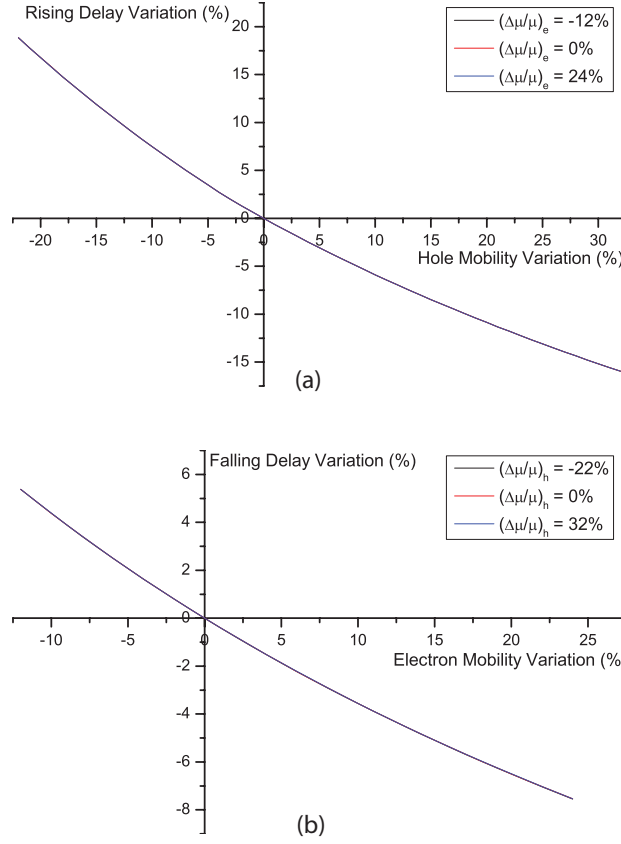
To cover mobility variation caused by multiple TSVs, the mobility variation range needs to be extended to  $-20\%$  to  $+8\%$  for PMOS and  $0\%$  to  $+14\%$  for NMOS. In addition, to consider both TSV and STI stresses, the mobility variation range needs to be extended even further. The mobility variation ranges needed to be covered for different stress sources are illustrated in Figure 32. Because of their opposite kinds of stress, the mobility variation range needed to be covered for TSV and STI hardly overlaps with each other. The interaction between both TSV and STI stresses requires more than merely adding the covered mobility variation range for both of them.



**Figure 32:** Extended timing corner with both TSV and STI stresses.

If mobility step is  $2\%$ ,  $312 (=24 \times 13)$  libraries need to be characterized with different mobility values, which is prohibitive. However, rising delay variation only depends on  $(\Delta\mu/\mu)_h$ , and falling delay variation only depends on  $(\Delta\mu/\mu)_e$  as shown in Figure 33. When an inverter rising delay with mobility variation is simulated, electron mobility variation does not contribute to the delay. Similarly, falling delay only depends on electron mobility variation. In addition, as shown in Figure 33, hole mobility variation can cause more than  $20\%$  PMOS performance variation depending on device technology, and electron mobility variation can enhance NMOS performance up to  $7.5\%$ . The inverter in NCSU library and PTM SPICE model [52] are used to obtain Figure 33. Therefore,  $(\Delta\mu/\mu)_e$  can be fixed while sweeping  $(\Delta\mu/\mu)_h$ . Characterizing  $37 (=24+13)$  libraries is enough to cover the entire mobility set. If mobility step is  $4\%$ ,  $20 (=13+7)$  libraries are required. Since delay variation

has semilinear dependency on mobility variation, interpolation can be used for the mobility value between two libraries.

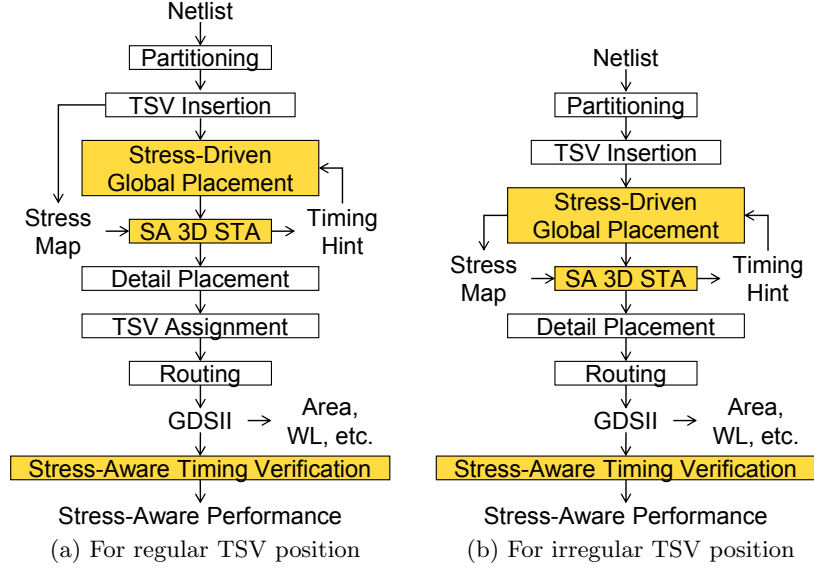


**Figure 33:** Inverter delay variation with different  $(\Delta\mu/\mu)_h$  and  $(\Delta\mu/\mu)_e$ . (a) Rising delay dependency on  $(\Delta\mu/\mu)_h$ , (b) Falling delay dependency on  $(\Delta\mu/\mu)_e$ .

#### 4.6 TSV-stress-driven Placement Optimization

In this section, an overview of the TSV-stress-driven timing-optimization methodology is presented. Basically, the placement styles in [1] are used. The developed 3D placer for TSV-stress-driven timing optimization is shown in Figure 34. The framework in [1] supports two different TSV placements, namely, regular and irregular TSV position. In the case of regular TSV position, TSVs are placed at regular gridlike sites over the die area, and any net that needs to span multiple dies must connect to these TSVs. In the case of irregular TSV position, TSV and gate cell positions are determined simultaneously. In this work, the global placement stage is modified for TSV-stress-driven timing optimization because of the flexibility to move cells to improve TSV-stress-aware timing. No TSV-stress-driven timing

optimization of any kind is performed during routing stage because TSV stress mainly affects gate delay based on gate position, which is not changed during routing.



**Figure 34:** Design flow for TSV-stress-driven placement optimization.

For design with regular TSV position, called TSV-site in [1], the flow starts by partitioning gate cells into dies of a 3D-IC stack using a min-cut approach. Then, the minimum number of required signal TSVs is estimated, and preplaced on the dies. Knowing position of preplaced signal TSVs, TSV stress map on all dies can be calculated for use during SA 3D STA. Then, TSV-stress-driven global placement, which is presented in Section 4.7, is performed to obtain placement result. Note that the placer calls SA 3D STA to obtain the sets of nets and gates on critical paths to be optimized after every predefined iterations. Then, detail placement is performed, and TSVs are assigned to multiple-die nets in the 3D-IC stack using the same method as in [1]. After routing, TSV-stress-aware performance can be evaluated from GDSII layout.

For design with irregular TSV position, called TSV coplacement in [1], the flow differs from that for design with regular TSV position in a few ways. After partitioning, TSVs are included into netlist as part of placement cells of multiple-die nets using the same heuristic, called net splitting, as in [1], and TSV assignment stage is not needed. Because TSV position is changed in every placement iteration, TSV stress map needs to be regularly

updated.

The presented design flow allows study of the impact of KOZ on TSV stress, carrier mobility variation, area, wirelength, and performance of 3D ICs. The result of the study is analyzed, and reported in detail in Section 4.8.

#### 4.7 TSV-stress-driven Global Placement

In this section, the TSV-stress-driven global-placement algorithm is described. It is based on a forced-directed quadratic placement [9], which was extended to support 3D-IC design in [1]. Carrier-mobility-based forces are introduced, and how to balance them against original placement forces in both works is described. The convergence of the algorithm when placing design with many TSVs with large KOZ is also discussed.

##### 4.7.1 Carrier Mobility-Based Forces

To consider the effect of TSV-stress-induced carrier mobility variation during global placement, two additional forces, one for hole mobility variation  $\mathbf{f}_x^{\text{mobil,h}}$  and another for electron mobility variation  $\mathbf{f}_x^{\text{mobil,e}}$ , need to be introduced into total force  $\mathbf{f}_x$ .

Here,  $\mathbf{f}_x^{\text{mobil,h}}$  and  $\mathbf{f}_x^{\text{mobil,e}}$  can be separately defined because they aim to optimize delay of different devices, e.g., PMOS and NMOS. For brevity, only description related to hole mobility is given because it similarly applies to electron mobility. The force can be represented by hole-mobility-based springs connected to cells, and defined as

$$\mathbf{f}_x^{\text{mobil,h}} = \mathring{\mathbf{C}}_x^{\text{m,h}}(\mathbf{x} - \mathring{\mathbf{x}}^{\text{m,h}}), \quad (19)$$

where vector  $\mathring{\mathbf{x}}^{\text{m,h}} = [\mathring{x}_1^{\text{m,h}} \cdots \mathring{x}_N^{\text{m,h}}]^T$  represents the x-position of target points to which  $N$  cells are connected by hole-mobility-based springs, and diagonal matrix  $\mathring{\mathbf{C}}_x^{\text{m,h}}$  collects spring constants  $\mathring{w}_{x,i}^{\text{m,h}}$  of hole-mobility-based spring connected to cell  $i$ .

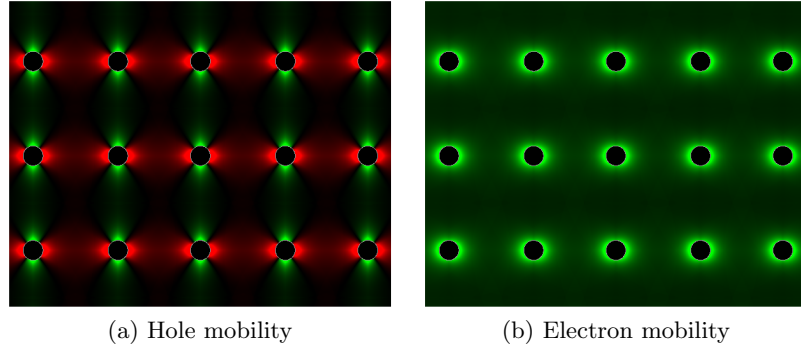
Hole-mobility-based target point  $\mathring{x}_i^{\text{m,h}}$  on die  $d$  is defined as

$$\mathring{x}_i^{\text{m,h}} = x'_i + l_i^{\text{m,h}} \cdot \left. \frac{\frac{\partial}{\partial x} \Phi^{\text{m,h}}(x, y)}{\|\nabla \Phi^{\text{m,h}}(x, y)\|} \right|_{(x'_i, y'_i), z=d}, \quad (20)$$

where vector  $\mathbf{x}' = [x'_1 \cdots x'_N]^T$  represents the x-position of  $N$  cells from last iteration,  $\Phi^{\text{m,h}}(x, y) = \frac{\Delta\mu}{\mu_{\text{total}}}(x, y)$  is hole-mobility-variation surface charted by using the model

described in Section 4.4.1,  $\nabla\Phi^{m,h}(x,y)$  is its gradient, and  $l_i^{m,h}$  is a length along the gradient direction of the surface. The gradient is added to current position in this equation because a cell should move in the direction of hole mobility increase.

The carrier-mobility-variation surfaces are shown in Figure 35. The green area in the figures indicates carrier mobility enhancement caused by TSV stress, and the red area indicates degradation. Unlike placement density, carrier-mobility-variation surface is smooth (except at TSV edge because mobility variation is not defined inside TSVs). Therefore, the direction of target points to which cells are connected by mobility-based springs can be determined from the surface gradient directly.



**Figure 35:** Carrier-mobility-variation surface surrounding TSVs.

#### 4.7.1.1 Balancing Forces

The newly introduced  $\mathbf{f}_x^{\text{mobil,h}}$  needs to be balanced against  $\mathbf{f}_x^{\text{den}}$  and  $\mathbf{f}_x^{\text{net}}$  (no need to balance against  $\mathbf{f}_x^{\text{hold}}$ ). The force-directed quadratic placement in [9] already has a mechanism to balance  $\mathbf{f}_x^{\text{den}}$  against  $\mathbf{f}_x^{\text{net}}$  so that the speed of cell spreading is regulated across placement iterations. The same mechanism can be used, and, therefore,  $\mathbf{f}_x^{\text{mobil,h}}$  is balanced against only  $\mathbf{f}_x^{\text{den}}$ . The parameters that need adjustment are the length along the gradient direction of hole-mobility-variation surface  $l_i^{m,h}$  and hole-mobility-based spring constant  $\hat{w}_{x,i}^{m,h}$ .

In this work,  $l_i^{m,h}$  is chosen so that hole mobility at the target point is higher than that at the current cell position. The length starts at  $1/8 \times$  average cell size, and increases to  $1/4$ ,  $1/2$ , and  $1 \times$  average cell size while hole mobility increases. The length is limited to average cell size so that wirelength does not increase too much. If hole mobility at even  $1/8$



$\times$  average cell size is lower than that at the current cell position, hole-mobility-based force is not applied to that cell at all in that iteration.

Compared to density-based gradient, which directly defines the length to density-based target point for  $\mathbf{f}_x^{\text{den}}$ ,  $l_i^{\text{m,h}}$  is relatively constant. Density-based gradient is extremely high in early placement iterations because of cell overlap, and decreases to almost zero as overlap is resolved in late iterations [9]. By limiting  $l_i^{\text{m,h}}$  to average cell size, it is naturally balanced against the length to density-based target point. The length to density-based target point dominates during early iterations, and the effect of  $l_i^{\text{m,h}}$  becomes pronounced when the length to density-based target point drops below  $l_i^{\text{m,h}}$  during late iterations.

During global placement, SA 3D STA is performed periodically. The results from SA 3D STA include the set of cells whose rise- and/or fall-time slack is negative. Then, hole-mobility-based spring constant  $\hat{w}_{x,i}^{\text{m,h}}$  is balanced against density-based spring constant  $\hat{w}_{x,i}^{\text{d}}$  of diagonal matrix  $\hat{\mathbf{C}}_x^{\text{d}}$  by defining it as

$$\hat{w}_{x,i}^{\text{m,h}} = c_i^{\text{h},j} \times \hat{w}_{x,i}^{\text{d}}, \quad (21)$$

where  $c_i^{\text{h},j}$  is rise-time criticality of cell  $i$  after  $j^{\text{th}}$  SA 3D STA, and defined as

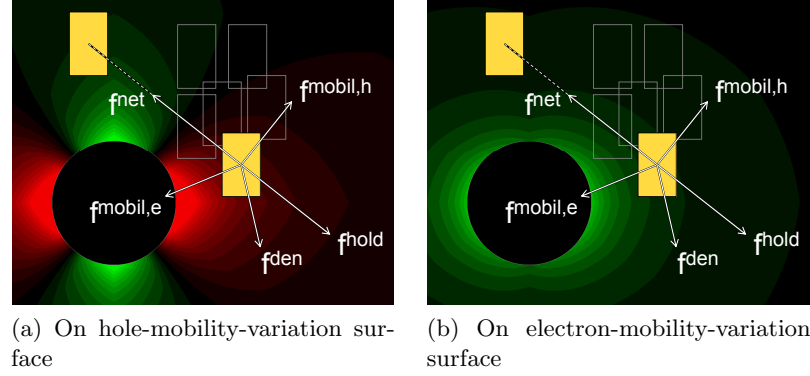
$$c_i^{\text{h},j} = \begin{cases} (c_i^{\text{h},j-1} + s_i^{\text{h},j}/S_{\min}^j)/2 & \text{if } i \in \mathbb{C}_c^{\text{h},j} \\ c_i^{\text{h},j-1}/2 & \text{otherwise,} \end{cases} \quad (22)$$

where  $s_i^{\text{h},j}$  is rise-time slack of cell  $i$ ,  $S_{\min}^j$  is the minimum timing slack of the design, and  $\mathbb{C}_c^{\text{h},j}$  is the set of cells whose rise-time slack is negative and less than 90% of  $S_{\min}^j$ . In other words, cell rise-time criticality is determined based on its history and current rise-time slack. Therefore, the effect of hole-mobility-based spring is pronounced on a cell, whose rise-time is highly critical, that needs hole mobility enhancement.

#### 4.7.1.2 New Total Force

An illustration of all forces applied to a cell is shown in Figure 36. In the figure,  $\mathbf{f}^{\text{net}}$  tries to hold the yellow cells of a net together, but  $\mathbf{f}^{\text{hold}}$  tries to nullify its effect, allowing cells to be moved based on other forces. Because of high cell density on top of the right yellow cell,  $\mathbf{f}^{\text{den}}$  tries to move the cell down. If the cell is rise-time critical,  $\mathbf{f}^{\text{mobil,h}}$  tries to move

the cell toward top right, away from the TSV, where hole mobility degradation decreases as shown in Figure 36(a). If the cell is, however, fall-time critical,  $\mathbf{f}^{\text{mobil,e}}$  tries to move the cell left, toward the TSV, where electron mobility increases as shown in Figure 36(b). In the case that a cell is both rise- and fall-time critical, the result depends on which timing is more critical.



**Figure 36:** All forces applied to a cell.

With the newly introduced hole-mobility-based force  $\mathbf{f}_x^{\text{mobil,h}}$  and electron-mobility-based force  $\mathbf{f}_x^{\text{mobil,e}}$ , the total force becomes

$$\mathbf{f}_x = \mathbf{f}_x^{\text{net}} + \mathbf{f}_x^{\text{hold}} + \mathbf{f}_x^{\text{den}} + \mathbf{f}_x^{\text{mobil,h}} + \mathbf{f}_x^{\text{mobil,e}}. \quad (23)$$

By setting  $\mathbf{f}_x = \mathbf{0}$  and substituting equations, the new result for each placement iteration can be obtained by solving

$$(\mathbf{C}_x + \mathring{\mathbf{C}}_x^{\text{d}} + \mathring{\mathbf{C}}_x^{\text{m,h}} + \mathring{\mathbf{C}}_x^{\text{m,e}})\Delta\mathbf{x} = -\mathring{\mathbf{C}}_x^{\text{d}}\mathbf{\Phi}^{\text{d}} + \mathring{\mathbf{C}}_x^{\text{m,h}}\mathbf{\Phi}^{\text{m,h}} + \mathring{\mathbf{C}}_x^{\text{m,e}}\mathbf{\Phi}^{\text{m,e}} \quad (24)$$

for  $\Delta\mathbf{x}$ , where vector  $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}'$  indicates how far cells should be moved,  $\mathbf{\Phi}^{\text{d}}$  is the vector collecting density-based gradients, and  $\mathbf{\Phi}^{\text{m,h}}$  and  $\mathbf{\Phi}^{\text{m,e}}$  are the vectors collecting  $l_i^{\text{m,h}} \cdot \frac{\partial}{\partial x} \Phi^{\text{m,h}} / \|\nabla \Phi^{\text{m,h}}\|$  and  $l_i^{\text{m,e}} \cdot \frac{\partial}{\partial x} \Phi^{\text{m,e}} / \|\nabla \Phi^{\text{m,e}}\|$  from Equation (20).

#### 4.7.2 Convergence of TSV-stress-driven Global Placement

Introducing  $\mathbf{f}^{\text{mobil,h}}$  and  $\mathbf{f}^{\text{mobil,e}}$  to 3D and force-directed placement without proper monitoring may cause problem to its convergence. During the early iterations of designs with

irregular TSV position, highly overlapping TSVs in a region result in extremely high mobility variation, which can misguide the placer. Because TSVs are also moved in every placement iteration as well to resolve their overlap, carrier-mobility-variation surfaces change. Critical cells are pulled by overlapping TSVs, worsening wirelength, until the overlap is finally resolved, and realize that the mobility improvement already vanishes. To prevent this problem, an upper-bound limit is set on mobility variation from Equation (15).

Another problem arises when a cell is moved over the top of a TSV or its KOZ during placement iterations. When a cell is inside a TSV, Equation (15) is not defined. Also the mobility variation is not valid when a cell is inside KOZ because the cell is moved out of KOZ during legalization. Under these cases,  $\mathbf{f}^{\text{mob},h}$  and  $\mathbf{f}^{\text{mob},e}$  are not applied to the cell to prevent the placer from being misguided.

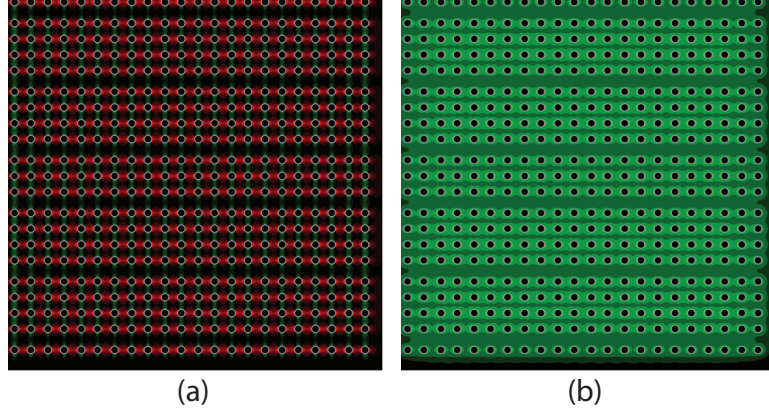
## 4.8 *Experimental Results*

### 4.8.1 Full-Chip Mobility Variation Map

The SA 3D STA flow is implemented in C++. The mobility-aware library is generated based on NCSU 45-nm cell library with 2% mobility step. TSV size of  $5\mu\text{m}$ , TSV parasitic capacitance of 70 fF, and resistance of  $0.1\Omega$  are used. The keep-out-zone (KOZ) size is set to  $0.5\mu\text{m}$ .

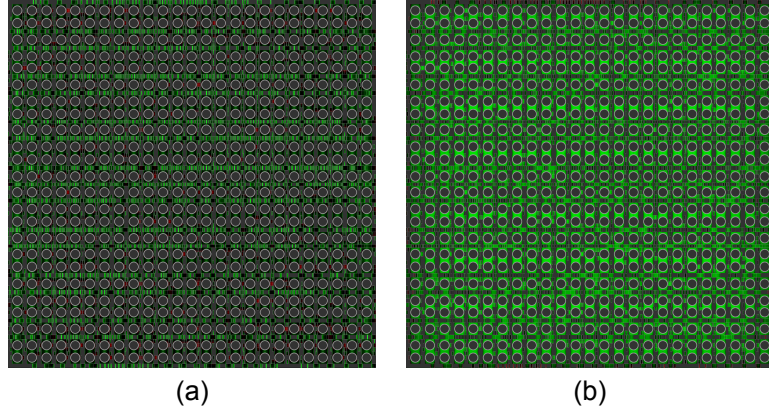
The compact stress and mobility modeling for TSV is efficient. The value of  $\Delta\mu/\mu$  at any point on a die can be obtained promptly. Generating mobility contour in Figure 37 (Die size:  $175^2\mu\text{m}^2$ , # TSVs: 462) takes only 14.9s. The proposed timing analysis with compact process/device model is fast enough to be used for iterative optimization purpose. An observation for layout optimization is shown in Figure 37(a). The leftmost and rightmost sides have wider hole mobility enhanced zone than the middle area because the regions have less mobility degradation by horizontally placed neighboring TSVs.

Next, mobility contours as shown in Figure 38 (Die size:  $220^2\mu\text{m}^2$ , # TSVs: 600, # cells: 3,422) are generated while considering stress from both TSV and STI. The contours are noticeably different from those in Figure 37 in two ways. First, because the stress inside STI is not the focus of this work, the area occupied by STI is shown in gray in the contour.



**Figure 37:** Mobility-variation contour map for  $22 \times 21$  TSV array. (a) hole, (b) electron.

Second, the trend of mobility variation changes, i.e., area of hole mobility enhancement increases, but area of electron mobility enhancement decreases. This phenomenon is largely due to the inclusion of STI stress.



**Figure 38:** Mobility-variation contour maps for a layout considering both TSV and STI stresses. (a) hole, (b) electron.

#### 4.8.2 Full-Chip Timing Analysis Results

In this experiment, SA 3D STA results are compared with the results neglecting stress. Ten benchmark circuits used to show the timing variation are listed in Table 6. The area utilization of each circuit is around 70 %. All whitespace is occupied by STI. The total amount of STI is in the same range (30 to 70 %) as in other STI-related works [53, 54]. The benchmark circuits are placed for wirelength minimization [1] with neither TSV nor STI stress consideration. Four-die 3D-IC stacks are assumed. The hole and electron mobility

variation of each benchmark circuit are shown in Table 7 and 8. The timing results are shown in Table 9 and 10.

**Table 6:** Benchmark circuits.

Circuits	# Cells	# Nets	# TSVs	Profile
ex	14,864	15,045	1,483	Execution unit
8051	15,712	15,755	1,575	Microcontroller
8086	19,895	19,909	1,987	Microprocessor
MAC2	29,706	29,980	2,971	Arithmetic unit
ethernet	77,234	77,381	7,748	Network controller
RISC	88,401	89,154	8,837	Microprocessor
b18	103,711	103,948	10,367	Multiprocessors
des_perf	109,181	109,416	10,916	Data encryption
vga_lcd	126,379	126,484	12,638	Display controller
b19	168,943	169,476	16,869	Multiprocessors

**Table 7:** Comparison of hole mobility variation range.

Circuit	Hole Mobility Variation (%)		
	With TSV Stress	With STI Stress	With TSV/STI Stresses
ex	-18.63 to +6.00	0.00 to +19.72	-14.88 to +25.21
8051	-18.88 to +6.36	0.00 to +19.72	-13.88 to +24.95
8086	-17.88 to +6.34	0.00 to +19.72	-13.70 to +25.53
MAC2	-17.46 to +6.31	0.00 to +19.72	-13.39 to +25.82
ethernet	-17.80 to +6.34	0.00 to +19.72	-14.06 to +25.93
RISC	-17.91 to +6.40	0.00 to +19.72	-14.25 to +26.05
b18	-18.63 to +6.32	0.00 to +19.72	-14.88 to +25.87
des_perf	-18.59 to +6.20	0.00 to +19.72	-14.49 to +25.85
vga_lcd	-18.65 to +6.37	0.00 to +19.72	-14.35 to +25.96
b19	-17.94 to +6.46	0.00 to +19.72	-14.43 to +25.83

When only TSV stress is considered, the hole and electron mobility variation of all benchmark circuits are in the same ranges as shown in Table 7 and 8. Hole mobility variation of cells in each circuit ranges from around  $-18\%$  to  $+6\%$ , whereas electron mobility variation ranges from 0 to around  $+13\%$ . Although the mobility variations of all benchmark circuits are in the same ranges, their timing variation is different. The change of longest path delay (LPD) of the benchmark circuits has variation from  $-5.65\%$  to  $+6.52\%$ . Some benchmark circuits have timing gain, whereas some benchmark circuits have timing penalty. On average, the impact of TSV stress on timing is  $2.82\%$ . For a random placement, because the average of carrier (both hole and electron) mobility variation is close to zero, the impact of

**Table 8:** Comparison of electron mobility variation range.

Circuit	Electron Mobility Variation (%)		
	With TSV Stress	With STI Stress	With TSV/STI Stresses
ex	0.00 to +13.28	-8.68 to 0.00	-8.68 to +10.86
8051	0.00 to +13.42	-8.68 to 0.00	-8.68 to +11.01
8086	0.00 to +12.73	-8.68 to 0.00	-8.68 to +11.15
MAC2	0.00 to +12.50	-8.68 to 0.00	-8.68 to +10.71
ethernet	0.00 to +12.61	-8.68 to 0.00	-8.68 to +11.00
RISC	0.00 to +12.85	-8.68 to 0.00	-8.68 to +11.30
b18	0.00 to +13.39	-8.68 to 0.00	-8.68 to +11.42
des_perf	0.00 to +13.44	-8.68 to 0.00	-8.68 to +11.49
vga_lcd	0.00 to +13.39	-8.68 to 0.00	-8.68 to +11.75
b19	0.00 to +13.02	-8.68 to 0.00	-8.68 to +11.49

**Table 9:** Longest path delay (LPD) comparison. (Percentage of changes is shown in parenthesis.)

Circuit	Longest Path Delay (ns)			
	Without Stress	With TSV Stress	With STI Stress	With TSV/STI Stresses
ex	12.009	11.881 (-1.06 %)	11.686 (-2.69 %)	11.577 (-3.59 %)
8051	5.041	5.370 (+6.52 %)	4.768 (-5.42 %)	4.761 (-5.56 %)
8086	9.283	9.423 (+1.50 %)	8.734 (-5.92 %)	8.888 (-4.26 %)
MAC2	7.797	7.905 (+1.38 %)	7.435 (-4.64 %)	7.525 (-3.49 %)
ethernet	9.294	9.484 (+2.05 %)	9.472 (+1.92 %)	9.562 (+2.89 %)
RISC	8.583	8.098 (-5.65 %)	8.434 (-1.73 %)	8.387 (-2.29 %)
b18	12.522	12.838 (+2.53 %)	12.013 (-4.06 %)	12.308 (-1.71 %)
des_perf	8.467	8.720 (+2.99 %)	8.026 (-5.21 %)	8.294 (-2.04 %)
vga_lcd	8.228	8.456 (+2.78 %)	7.835 (-4.77 %)	8.078 (-1.82 %)
b19	13.389	13.618 (+1.71 %)	12.760 (-4.70 %)	12.821 (-4.25 %)
Ave. Abs. Change		(2.82 %)	(4.11 %)	(3.19 %)

hole and electron mobility variation may compensate each other, resulting in low combined enhancement/degradation in timing for some cases. In many cases, however, the impact of hole and electron mobility variation is in the same direction, resulting in significant changes (either enhancement or degradation) of longest path delay. If TSV stress is considered while placing gates and TSVs, performance improvement can be expected for every benchmark circuit. The change of total negative slack (TNS) has variation from  $-28.48\%$  to  $+50.43\%$ , which is bigger than the variation of the change of delay. That result motivates the need of TSV-stress-aware layout optimization.

When only STI stress is considered, the hole and electron mobility variation of all

**Table 10:** Total negative slack (TNS) comparison. (Percentage of changes is shown in parenthesis.)

Circuit	Total Negative Slack (ns)			
	Without Stress	With TSV Stress	With STI Stress	With TSV/STI Stresses
ex	-8.815	-7.215 (-18.15 %)	-5.280 (-40.10 %)	-4.348 (-50.67 %)
8051	-144.035	-145.363 (+0.92 %)	-60.450 (-58.03 %)	-61.351 (-57.41 %)
8086	-19.317	-26.779 (+38.63 %)	-3.495 (-81.90 %)	-7.194 (-62.76 %)
MAC2	-87.337	-93.422 (+6.97 %)	-46.541 (-46.71 %)	-49.861 (-42.91 %)
ethernet	-474.917	-463.344 (-2.44 %)	-492.182 (+3.64 %)	-480.541 (+1.18 %)
RISC	-57.101	-40.840 (-28.48 %)	-27.864 (-51.20 %)	-17.779 (-68.86 %)
b18	-41.301	-62.128 (+50.43 %)	-22.024 (-46.67 %)	-30.331 (-26.56 %)
des_perf	-40.298	-45.054 (+11.80 %)	-10.090 (-74.96 %)	-11.513 (-71.43 %)
vga_lcd	-0.991	-1.191 (+20.25 %)	-0.671 (-32.27 %)	-0.875 (-11.71 %)
b19	-126.528	-145.533 (+15.02 %)	-43.996 (-65.23 %)	-48.795 (-61.44 %)
Ave. Abs. Change		(19.31 %)	(50.07 %)	(45.49 %)

benchmark circuits are exactly in the same ranges as shown in Table 7 and 8. Hole mobility variation of cells in each circuit ranges from 0 to +19.72%, whereas electron mobility variation ranges from -8.68 to 0%. The carrier mobility variation of all benchmark circuits are in the exact same range because STI stress depends heavily on the relative size of cell and its adjacent STIs. Wide cells or cells far away from narrow STIs have no carrier mobility variation, whereas the narrowest cell in the library having wide STIs on both of its sides has the highest carrier mobility variation. The change of longest path delay of the benchmark circuits has variation from -5.92% to +1.92%, and the average delay variation is 4.11%. Most benchmark circuits have timing gain because, for a random placement, the average of carrier (both hole and electron) mobility variation is much higher than zero. In addition, STI is pervasive on an IC layout. Without considering STI stress, STA only reports pessimistic timing result. Including STI stress, the pessimism in timing results decreases. TNS is significantly reduced by 50.07% on average because several violating paths become nonviolating by STI stress. Wide variation of the change of both delay and TNS suggests the importance of STI-stress-aware layout optimization.

Finally, when both TSV and STI stresses are considered, the range of hole and electron mobility variation of all benchmark circuits shifts from the range when only TSV or only STI stress is considered as shown in Table 7 and 8. Hole mobility variation of cells in each

circuit ranges from around  $-14\%$  to  $+25\%$ , whereas electron mobility variation ranges from  $-8.68$  to around  $+11\%$ . The change of longest path delay of the benchmark circuits has variation from  $-5.56\%$  to  $+2.89\%$ , and the average delay variation is  $3.19\%$ . The changes are in the same direction as the changes considering only STI stress. Compared to the changes considering only STI stress, some benchmark circuits have timing gain, whereas some benchmark circuits have timing penalty. This variation suggests that TSV stress still has significant impact on timing even after STI stress is considered. TNS is significantly reduced by  $45.49\%$  on average. Therefore, both TSV and STI stresses can be exploited together for performance improvement. The potential to exploit them to improve timing is revealed as shown in the next experiment.

#### 4.8.3 Manual Placement Optimization Results

The critical path in des\_perf is manually optimized to present the potential benefit of TSV-stress-aware layout optimization. Before optimization, the path delay is  $8.720\text{ ns}$  with TSV-stress-aware timing analysis. However, the delay could be reduced to  $8.138\text{ ns}$  with small layout perturbation, which is  $6.67\%$  improvement. It is even less than the path delay without stress, which is  $8.467\text{ ns}$  in Table 9. The gates on the path are shown in Table 11. The gates are renamed according to the mobility variation. Each gate position is adjusted with small perturbation so that the path has timing gain. The maximum timing gain in a gate is  $23.37\%$  improvement.

In Table 11, the timing of some gates is not improved even though the carrier mobility of the gates is enhanced. For example, the delay time of Gate #12 increases by  $1.11\%$  while electron mobility increases by  $4\%$  (from  $+8\%$  to  $+12\%$ ). The increase in delay is the result from moving a gate that it drives, Gate #13, to improve hole mobility of Gate #13. The impact of increase in wire capacitive load of Gate #12 outweighs the impact of increase in electron mobility of the gate, resulting in delay increase. If the electron mobility of the gate was not increased, the delay increase would be higher than  $1.11\%$ . Although the delay time of some gates on the path increases for this reason, the overall path delay decreases.

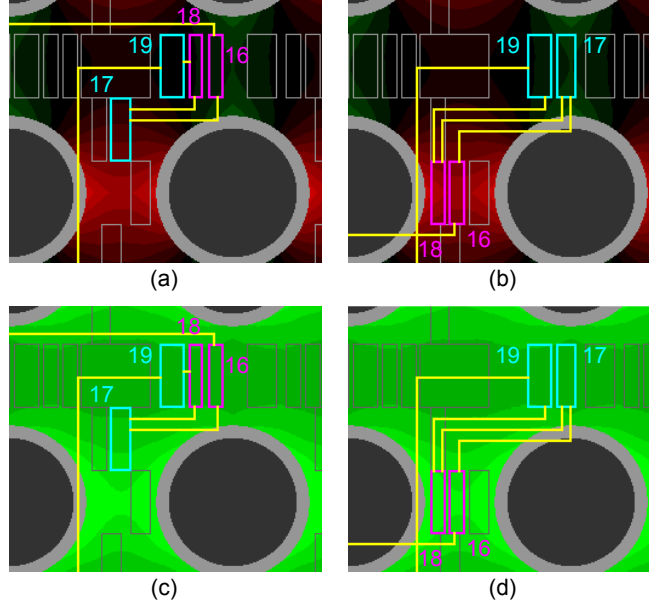


How gate reposition works for timing optimization is illustrated in Figure 39. The placement result on a die with TSV-stress-induced mobility-variation contours is captured. The gates in Logic depth 17 and 19 are hole-mobility critical gates because their timing arc is rising on the path. Therefore, the gates are perturbed to be placed close to green area in hole-mobility contour. However, the gates in Logic depth 16 and 18 are electron-mobility critical. Therefore, the gates are pushed to electron-mobility enhancement zone in Figure 39(c) and (d).

**Table 11:** Gate optimization considering only TSV stress on the target path with perturbation.

Logic Depth	Original Gate	Optimized Gate	Timing Arc	Original Delay (ps)	Optimized Delay (ps)	Reduction (%)
	Input port	Input port				
1	INVX1_P-6_N+10	INVX1_P-16_N+12	Fall	48.58	43.81	9.82
2	INVX1_P-16_N+12	INVX1_P+4_N+8	Rise	32.77	28.59	12.76
3	INVX1_P-16_N+12	INVX1_P-16_N+12	Fall	160.67	162.23	-0.97
4	INVX1_P-6_N+10	INVX1_P+2_N+8	Rise	469.81	422.60	10.05
5	INVX1_P-6_N+10	INVX1_P-16_N+12	Fall	183.39	168.95	7.87
6	INVX1_P-8_N+10	INVX1_P+4_N+8	Rise	789.49	717.44	9.13
7	INVX1_P+0_N+0	INVX1_P+0_N+0	Fall	896.10	939.16	-4.81
8	INVX1_P-16_N+12	INVX1_P+4_N+8	Rise	1,630.01	1,396.33	14.34
9	INVX1_P-2_N+6	INVX1_P-2_N+6	Fall	327.01	376.35	-15.09
10	INVX1_P+2_N+6	INVX1_P+2_N+6	Rise	221.73	182.14	17.86
11	INVX4_P+2_N+6	INVX4_P+2_N+6	Fall	112.48	95.60	15.01
12	MUX2X1_P+0_N+8	MUX2X1_P-16_N+12	Fall	401.91	406.39	-1.11
13	MUX2X1_P-16_N+12	MUX2X1_P+2_N+8	Rise	922.52	730.53	20.81
14	AOI21X1_P-16_N+12	AOI21X1_P-16_N+12	Fall	528.11	610.77	-15.65
15	NAND3X1_P+2_N+8	NAND3X1_P+2_N+8	Rise	826.82	941.80	-13.91
16	INVX1_P+2_N+8	INVX1_P-16_N+12	Fall	840.11	643.74	23.37
17	NOR2X1_P-6_N+10	NOR2X1_P+4_N+8	Rise	262.29	212.49	18.99
18	INVX1_P+2_N+8	INVX1_P-16_N+12	Fall	49.16	39.85	18.94
19	OAI21X1_P+0_N+8	OAI21X1_P+2_N+8	Rise	17.06	19.20	-12.54
	DFFPOSX1	DFFPOSX1	Rise	0.11	0.12	-9.09
Path Delay				8,720.13	8,138.09	6.67

Finally, the same critical path in des\_perf is manually optimized when considering both TSV and STI stresses to reveal the impact of the interaction between both stresses on performance. Before optimization, the path delay is 8.294ns with TSV-STI-stress-aware



**Figure 39:** Gate perturbation to take advantage of TSV-stress-induced mobility variation. (a) hole-mobility contour with original gate placement, (b) hole-mobility contour after gate perturbation, (c) electron-mobility contour with original gate placement, (d) electron-mobility contour after gate perturbation.

timing analysis. However, the delay could be reduced to 7.867 ns with small layout perturbation which is 5.15 % improvement. The gates on the path are shown in Table 12. The gates are renamed according to the mobility variation. Each gate position is adjusted with small perturbation so that the path has timing gain. The maximum timing gain in a gate is 17.63 % improvement.

How gate reposition works for this timing optimization is illustrated in Figure 40. The placement result on a die with TSV-STI-stress-induced mobility-variation contours is captured. Like in previous experiment, the gates in Logic depth 17 and 19 are hole-mobility critical gates because their timing arc is rising on the path. Besides moving them to the area that TSVs provide improvement on rise time, surrounding them by STIs improves hole mobility. However, the gates in Logic depth 16 and 18 are electron-mobility critical. Therefore, the gates are pushed to electron-mobility enhancement zone provided by TSVs as shown in Figure 40(c) and (d). Note that manual optimization when considering both TSV and STI stresses is more difficult than when only TSV stress is considered. When only TSV stress is considered, moving gates does not change mobility-variation contour. When both TSV

and STI stresses are considered, moving gates to exploit TSV stress changes dimension of STI surrounding it, thus their delay may not be improved as much as expected.

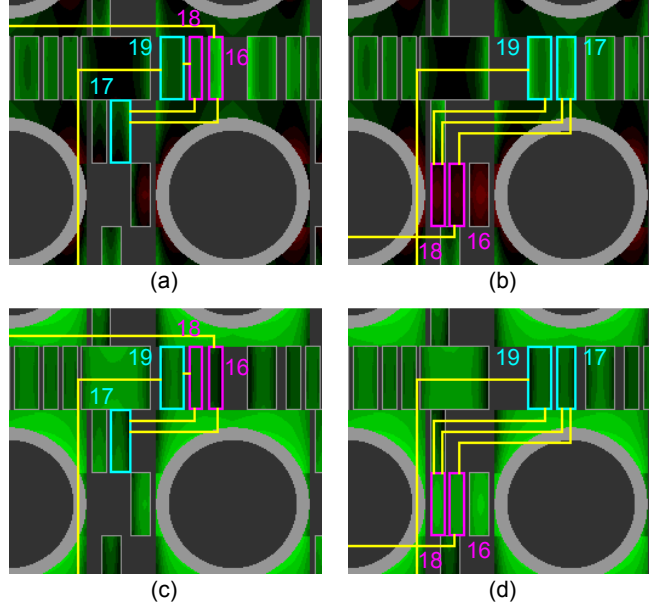
**Table 12:** Gate optimizations considering both TSV and STI stresses on the target path with perturbation.

Logic Depth	Original Gate	Optimized Gate	Timing Arc	Original Delay (ps)	Optimized Delay (ps)	Reduction (%)
	Input port	Input port				
1	INVX1_P+6_N+4	INVX1_P-8_N+10	Fall	49.85	44.62	10.49
2	INVX1_P+4_N+4	INVX1_P+20_N+0	Rise	29.15	24.01	17.63
3	INVX1_P+4_N+4	INVX1_P-6_N+8	Fall	167.51	166.46	0.63
4	INVX1_P+6_N+4	INVX1_P+16_N+2	Rise	428.15	374.05	12.64
5	INVX1_P+6_N+4	INVX1_P-2_N+6	Fall	178.81	163.86	8.36
6	INVX1_P+10_N+2	INVX1_P+18_N+2	Rise	677.35	614.72	9.25
7	INVX1_P+20_N-8	INVX1_P+6_N-2	Fall	914.05	949.16	-3.84
8	INVX1_P+4_N+4	INVX1_P+16_N+2	Rise	1,489.17	1,261.57	15.28
9	INVX1_P+18_N-2	INVX1_P+18_N-2	Fall	352.66	391.58	-11.04
10	INVX1_P+20_N-4	INVX1_P+20_N-4	Rise	175.74	147.00	16.35
11	INVX4_P+18_N-2	INVX4_P+18_N-2	Fall	102.76	89.22	13.18
12	MUX2X1_P+2_N+8	MUX2X1_P-10_N+10	Fall	409.04	410.91	-0.46
13	MUX2X1_P-10_N+10	MUX2X1_P-2_N+8	Rise	874.10	800.69	8.40
14	AOI21X1_P-10_N+8	AOI21X1_P-10_N+8	Fall	538.98	553.49	-2.69
15	NAND3X1_P+10_N+4	NAND3X1_P+12_N+4	Rise	749.84	702.28	6.34
16	INVX1_P+16_N+2	INVX1_P-6_N+8	Fall	852.80	888.45	-4.18
17	NOR2X1_P+4_N+6	NOR2X1_P+12_N+4	Rise	240.80	222.78	7.48
18	INVX1_P+10_N+4	INVX1_P-8_N+8	Fall	46.81	44.56	4.81
19	OAI21X1_P+6_N+6	OAI21X1_P+10_N+4	Rise	16.58	17.53	-5.73
	DFFPOSX1	DFFPOSX1	Rise	0.11	0.12	-9.09
Path Delay				8,294.26	7,867.06	5.15

#### 4.8.4 Impact of KOZ on Carrier Mobility Variation

For the remaining experiments, IWLS 2005 benchmarks [32] and several industrial circuits, as listed in Table 13, are used. TSV size is  $3\mu\text{m}$ . The TSV parasitic capacitance and resistance are  $50\text{ fF}$  and  $0.2\Omega$ , respectively. The KOZ around TSVs is expanded to make TSV cells (= TSV + KOZ) fit inside two to seven standard-cell rows (one standard-cell row =  $2.47\mu\text{m}$ ). All experiments are based on 4-die 3D-IC stacks with constant cell area density. Min-cut partitioner is used, and the target clock period of each circuit is set to the value reported after synthesis. All reported timing results come from SA 3D STA.

In this experiment, TSV cell size is increased from 2-row to 7-row while carrier mobility



**Figure 40:** Gate perturbation to take advantage of TSV-STI-stress-induced mobility variation. (a) hole mobility contour with original gate placement, (b) hole mobility contour after gate perturbation, (c) electron mobility contour with original gate placement, (d) electron mobility contour after gate perturbation.

**Table 13:** Benchmark circuits.

Circuit	# Gates	# Nets	# TSVs	Profile
ckt1	20K	20K	634	Microprocessor
ckt2	33K	33K	3,554	Arithmetic Unit
ckt3	50K	51K	5,352	Connection Bus
ckt4	80K	80K	2,846	Network Controller
ckt5	119K	119K	5,341	Data Encryption

variation caused by TSV stress is observed. The results are shown in Table 14. The results indicate that carrier mobility variation decreases as KOZ size increases, and starts becoming negligible (1 % or less) when TSV cell size reaches 6-row. Mobility variation in design with irregular TSV position is larger than that in design with regular TSV position. TSV cells in design with irregular TSV position can be crowded in some area, causing high TSV stress and mobility variation.

#### 4.8.5 Impact of KOZ on Area and Wirelength

The main purpose of KOZ is to prevent gates from being placed so close to TSV that they experience carrier mobility variation. The side effect of enforcing large KOZ to have

**Table 14:** Impact of KOZ on carrier mobility variation for ckt5.

TSV Cell	Mobility Variation (%)			
	Regular TSV Position		Irregular TSV Position	
	Hole	Electron	Hole	Electron
2-row	-4.56 – 2.81	0.33 – 3.62	-8.76 – 4.74	0.02 – 7.47
3-row	-4.05 – 2.45	0.30 – 2.35	-6.04 – 2.55	0.02 – 4.10
4-row	-2.07 – 1.51	0.26 – 1.37	-2.87 – 1.75	0.02 – 2.17
5-row	-1.55 – 0.93	0.18 – 0.89	-2.30 – 1.26	0.02 – 1.32
6-row	-1.01 – 0.70	0.13 – 0.64	-1.33 – 0.70	0.02 – 0.85
7-row	-0.90 – 0.53	0.07 – 0.40	-1.15 – 0.61	0.02 – 0.62

predictable device performance is shown in Table 15. The footprint area of the 3D-IC stack for ckt5 increases almost  $4\times$  if TSV cell size is 6-row. The increased area is primarily consumed by TSV cells. In an extreme case, almost half of silicon area is consumed by TSV cells when TSV cell size is 7-row. Increasing footprint area inevitably results in  $2\times$  wirelength increase because of large KOZ choice.

**Table 15:** Impact of KOZ on area and wirelength for ckt5.

TSV Cell	Footprint (mm <sup>2</sup> )	TSV Cell Area (mm <sup>2</sup> )	Wirelength (m)	
			Regular TSV Position	Irregular TSV Position
2-row	0.176 (1.00)	0.130 (18.47 %)	3.415 (1.00)	2.970 (1.00)
3-row	0.250 (1.42)	0.293 (29.33 %)	3.970 (1.16)	3.475 (1.17)
4-row	0.360 (2.04)	0.521 (36.21 %)	4.726 (1.38)	4.196 (1.41)
5-row	0.504 (2.86)	0.815 (40.40 %)	5.526 (1.62)	4.654 (1.57)
6-row	0.672 (3.81)	1.173 (43.61 %)	6.331 (1.85)	5.328 (1.79)
7-row	0.884 (5.01)	1.597 (45.17 %)	7.179 (2.10)	6.036 (2.03)

#### 4.8.6 Impact of KOZ on TSV-stress-aware Timing

SA 3D STA is performed after obtaining placement results from wirelength-driven (WLD), timing-driven (TD), and TSV-stress-driven (SD) placers. The results are shown in Table 16. First, under TSV stress, the timing results from timing-driven placement can be unpredictable and worse than the results from even wirelength-driven placement in many cases. Traditional timing-driven placer is oblivious to the change in carrier mobility of devices, and only tries to reduce the capacitive load on timing-critical gates. Second, the TSV-stress-driven placer outperforms timing-driven placer consistently. The improvements over wirelength-driven placement on worst negative slack (WNS) and total negative slack

(TNS) are up to 39 % and 42 %, respectively. Third, using 2-row TSV cells, the TSV-stress-driven placer provides better result for design with irregular TSV position than the result for design with regular TSV position. Design with irregular TSV position has shorter wirelength and higher carrier mobility variation, which the placement algorithm can intelligently exploit, than design with regular TSV position. Finally, as TSV cell size increases, the benefit from TSV-stress-driven placement decreases. Large KOZ leaves not much mobility variation for the TSV-stress-driven placer to exploit.

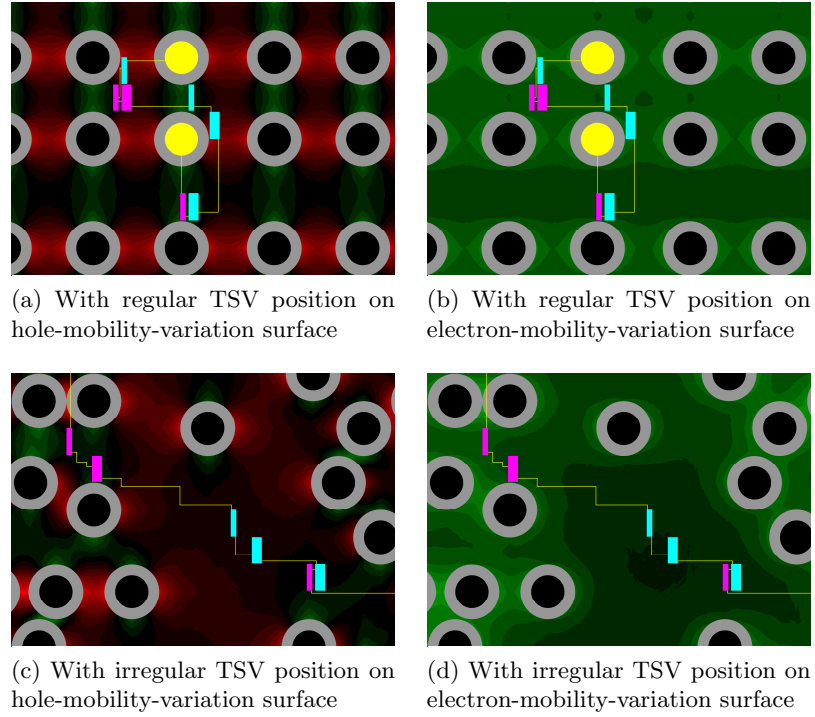
**Table 16:** Impact of KOZ on TSV-stress-aware timing for ckt5.

TSV Cell	Regular TSV Position						Irregular TSV Position					
	WLD		TD		SD		WLD		TD		SD	
	WNS	TNS	WNS	TNS	WNS	TNS	WNS	TNS	WNS	TNS	WNS	TNS
	(ps)	(ps)					(ps)	(ps)				
	100 %	100 %	(%)	(%)	(%)	(%)	100 %	100 %	(%)	(%)	(%)	(%)
2-row	-92.72	-143	113.66	126.57	77.01	69.93	-79.26	-120	127.26	143.33	60.62	57.50
3-row	-96.62	-156	70.14	65.38	70.60	62.82	-77.89	-118	133.11	147.46	94.12	92.37
4-row	-102.86	-170	85.06	82.94	78.20	74.71	-85.42	-134	111.53	114.18	92.95	90.30
5-row	-99.28	-157	88.43	87.90	88.48	87.90	-88.32	-139	100.83	100.72	99.91	99.28
6-row	-88.45	-139	99.27	99.28	99.31	99.28	-88.33	-139	99.43	99.28	99.54	99.28
7-row	-88.55	-139	99.02	99.28	99.09	99.28	-88.28	-139	99.63	99.28	99.43	99.28

#### 4.8.7 TSV-stress-driven Placement Results

Placement results are obtained from the TSV-stress-driven placer. The snapshots of ckt3 are shown in Figure 41. In the figures, gray band surrounding TSVs is KOZ. Logic gates in magenta are hole-mobility critical. Their timing arcs are rising on the critical paths. The placer positions them (if possible) in green area of Figure 41(a) where they receive hole mobility enhancement, or, at least, in black area where they do not experience hole mobility degradation. On the other hand, logic gates in sky blue are electron-mobility critical. Their timing arcs are falling on the critical paths. The placer positions them (if possible) in bright green area of Figure 41(b) where they receive high electron-mobility enhancement.

The results from different placement algorithms using 2-row TSV cells are shown in Table 17. On average, timing-driven placer does not provide performance improvement over wirelength-driven placer when evaluated by SA 3D STA. The gates on critical paths may be placed in position that their carrier mobility is degraded by TSV stress. On the



**Figure 41:** Zoom-up snapshots of TSV-stress-driven placement results for ckt3 using 2-row TSV cells.

other hand, the TSV-stress-driven placer consistently provides better performance than the other placers. On average, the performance improvement over wirelength-driven placement on WNS and TNS are 21.6 % and 28.0 %, respectively. It is observed here again that the results for design with irregular TSV position are better than those for design with regular TSV position in all cases.

**Table 17:** Timing comparison for regular and irregular TSV position with 2-row TSVs.

Ckt.	Regular TSV Position						Irregular TSV Position					
	WLD		TD		SD		WLD		TD		SD	
	WNS (ps)	TNS (ps)	WNS (ps)	TNS (ps)	WNS (ps)	TNS (ps)	WNS (ps)	TNS (ps)	WNS (ps)	TNS (ps)	WNS (ps)	TNS (ps)
ckt1	-163.50	-1,167	-156.67	-1,034	-156.39	-1,034	-157.04	-1,063	-161.21	-1,107	-155.15	-1,004
ckt2	-159.35	-5,104	-180.86	-6,076	-129.35	-4,105	-127.28	-4,005	-134.70	-4,327	-120.45	-3,888
ckt3	-79.35	-605	-65.85	-428	-53.72	-321	-73.40	-482	-56.65	-348	-51.88	-307
ckt4	-55.39	-131	-49.25	-106	-38.01	-72	-50.38	-109	-40.75	-80	-34.95	-66
ckt5	-92.72	-143	-105.39	-181	-71.40	-100	-79.26	-120	-100.87	-172	-48.05	-69
Ave (%)	100.00	100.00	98.98	97.18	78.03	69.40	100.00	100.00	98.76	100.22	78.82	74.65

## 4.9 *Summary*

In this chapter, a first-order compact model for TSV-stress-induced mobility variation and an STI-stress-induced mobility variation are developed. A design methodology is proposed to analyze the systematic variation and optimize layout by locating critical cells in a mobility enhanced region of TSVs or changing STIs surrounding the cells. The proposed TSV-STI-stress-aware timing analysis framework for 3D ICs also opens the opportunity for stress-aware layout optimizations, such as placement and TSV-STI optimizations. The mobility variation models and timing analysis framework allow the study of the impact of TSV and STI stresses and their interaction on full-chip timing. Manual placement optimizations are performed as examples of the design methodology. The mobility variation models and timing analysis framework also allow the study of the impact of KOZ dimension around TSVs on the mechanical stress, carrier mobility variation, area, wirelength, and performance of 3D ICs. Large KOZs practically nullify the impact of TSV-induced stress on carrier mobility at the cost of increase in chip stack footprint area and wirelength. Finally, a TSV-stress-aware placement algorithm is proposed to regain footprint area. Instead of avoiding hole and electron mobility variation caused by TSV stress, the placement algorithm exploits the variation to reduce KOZ dimension.



## CHAPTER V

### EXPLOITING DIE-TO-DIE THERMAL COUPLING IN 3D-IC PLACEMENT

Increasing functionality while miniaturizing footprint of integrated circuits (ICs) is today's trend of electronic industry. Moving to smaller technology node is a traditional approach toward that goal; however, investing in new production lines needs to be economically justified. Three-dimensional (3D) stacking of thinned dies provides feasibility to keep the trend while staying at current technology node. Polymer adhesive is a popular material used to bond thinned dies together [23]. Interleaving layers of thinned dies and polymer adhesive are, therefore, commonly found in 3D ICs.

Stacking thinned dies in 3D ICs results in increasing power density, thus rising temperature, which leads to other reliability problems, such as electromigration [21] and negative-bias-temperature instability [22]. Because of low thermal conductivity, polymer adhesive exacerbates the problem. Moreover, if the thinned dies are silicon on insulator, an extremely high temperature can be expected. Heat must be removed from the die quickly; otherwise, reliability problems may arise.

A few recent works on temperature-aware placement for 3D ICs have been published. In [2], a force-directed approach was proposed for 3D thermal placement; however, it did not include through-silicon vias (TSVs), which are commonly found in 3D ICs. In [4], a partitioning-based approach was proposed for 3D thermal placement. The work considered the impact of parasitic resistance and capacitance of signal TSVs on power, but failed to include thermal properties of TSVs. Failing again to acknowledge TSV area, it also reported unreasonably large numbers of TSVs even for small circuits. The work in [5] considered TSV thermal properties; however, it assumed that adhesive is an ideal insulator. In reality, heat can still flow through (silicon and) adhesive because of its thinness. Based on the assumption, the work balanced only the number of TSVs in a bin to heat dissipated from

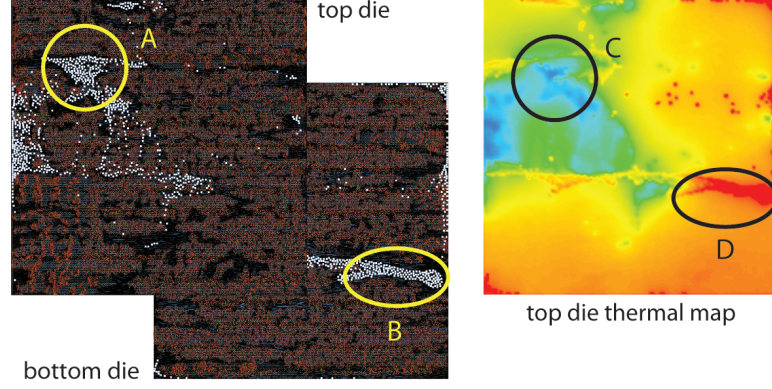
cells in the same bin and bins vertically below.

In this work, two effective heuristics, namely TSV spread and alignment method (TSA) and thermal coupling-aware placement (CA), that exploit the die-to-die thermal coupling in 3D ICs in force-directed temperature-aware placement are proposed. New placement forces are presented, and how to manage them to obtain high quality placements is discussed. The framework used to evaluate the impact of TSVs on temperature of 3D ICs is presented. The main components of the framework are power analysis and GDSII-level thermal analysis for 3D ICs. Extensive experiments are performed to show the trade-off among wirelength, delay, power, and temperature results obtained from GDSII layouts. The proposed placers outperform several state-of-the-art placers published in recent literature [2, 3, 4, 5, 1].

### **5.1 Motivation**

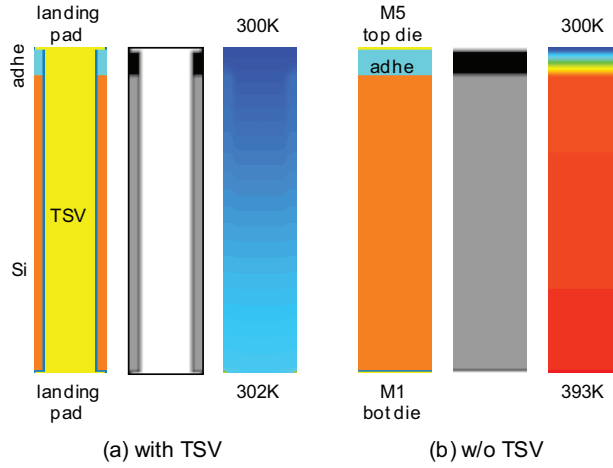
Because of their occupied area and high thermal conductivity of copper, widely used fill material, TSVs have significant impact on temperature. In a 3D-IC layout, logic gates cannot overlap with TSVs. Area occupied by TSVs becomes “power whitespace” because no power is consumed, and thus no heat is generated. In addition, TSVs conduct majority of heat through polymer adhesive between dies toward the heatsink as shown in Figure 42. In the figure, the hotspot D on the top metal layer of the top die is caused by the TSVs in spot B from the bottom die. Heat flows through TSVs so intensely that its effect still remains on the top die. Thus, the temperature distribution of the top die results from the combination of power profile of the top die and heat flowing from the bottom die through TSVs. The TSV spread and alignment method presented in this work exploits these thermal properties of TSVs by distributing TSVs evenly to reduce power density in local power hotspots and vertically aligning TSVs of adjacent dies to establish direct paths to the heatsink.

Using ANSYS FLUENT [55], a part of bulk silicon with and without TSVs (and their related structures, e.g., landing pad and liner) is simulated as shown in Figure 43. To obtain the temperature distribution, the temperature is fixed on the top side of the models, and constant power density is applied on the bottom side. The simulation results indicate that heat flowing through a TSV increases temperature far less than the same amount of heat



**Figure 42:** Die-to-die heat coupling from TSVs. TSVs are shown in white. The top die is closer to heatsink. The cold spot C is caused by the TSVs in spot A on the same die. The hot spot D is caused by the TSVs in spot B from the bottom die.

flowing through bulk silicon and adhesive. The temperature slowly increases in bulk silicon with TSVs. On the other hand, in bulk silicon without any TSV, low thermal conductivity of bonding adhesive results in steep temperature rise at first, but temperature does not rise as much inside the silicon. The average thermal conductivity of bulk silicon with and without TSVs is computed, and used to guide the thermal coupling-aware placer presented in this work.



**Figure 43:** Structure, thermal conductivity, and thermal profile of bulk silicon with and without TSVs. Dark shade represents low thermal conductivity.

## 5.2 *Global Placement Algorithms*

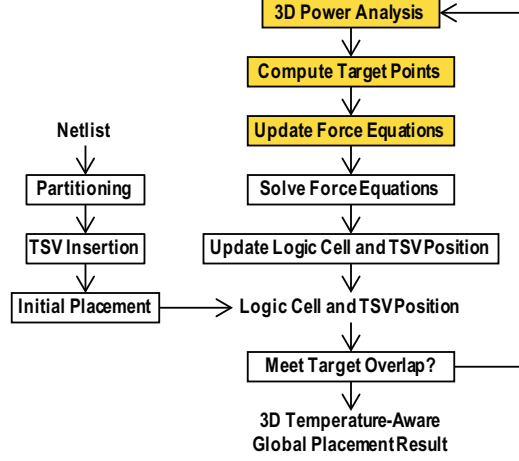
In this section, the two 3D temperature-aware global placement algorithms that are based on force-directed methodology [9] are described. This placer is extended in two ways to perform thermal optimization in 3D ICs. In the first algorithm, TSVs are laterally spread in each die to form even thermal conductivity while perturbing TSV position to increase vertical overlap among TSVs across the dies in 3D stack. In the second algorithm, the logic cells on each die are positioned by using thermal conductivity-based force while TSVs are positioned by using power density-based force.<sup>1</sup>

### 5.2.1 **Design Flow**

The overall flow of the placement algorithms, where the position of cells and TSVs is determined simultaneously, is shown in Figure 44. Given a netlist, cells are partitioned into dies if the partition is not also given. Then, the minimum number of TSVs required to connect cells on different dies is inserted. Once this die partitioning is fixed, cells are not moved across dies during placement. The reason is that changing cell partition results in change in the number of TSVs, and this change causes the complexity of problem to become unmanageable. Next, wirelength is minimized to obtain initial placement, which may contain high overlap among cells and TSVs. In the main loop to resolve the overlap, TSV density and TSV position are used to compute target point for TSVs in the first algorithm. In the second algorithm, 3D power analysis (explained in Section 5.3.1) is periodically performed based on current cell and TSV position. Then, the cell power, TSV density, and average thermal conductivity of bulk silicon (obtained from the simulation results in Section 5.1) are used to compute target points for cells and TSVs to move towards. After updating force equations and solving them, the position of cells and TSVs are updated. This loop continues until the overlap is sufficiently reduced.

---

<sup>1</sup>An attempt to combine these two methods was made, but the results were not consistent.



**Figure 44:** Design flow for the 3D IC global placement.

### 5.2.2 Force-directed 3D Placement

In a quadratic placement [9], quadratic wirelength  $\Gamma_x$  and  $\Gamma_y$  along x- and y-axis are separately minimized to obtain the placement result. Treated  $\Gamma_x$  as spring energy, its derivative can be regarded as net force  $\mathbf{f}_x^{\text{net}}$ . By setting  $\mathbf{f}_x^{\text{net}}$  to zero, the minimum  $\Gamma_x$  and the corresponding placement are found; however, cells may overlap in few small areas. Hold force  $\mathbf{f}_x^{\text{hold}}$  prevents  $\mathbf{f}_x^{\text{net}}$  from pulling cells back to the initial placement. In addition, density-based force  $\mathbf{f}_x^{\text{den}}$  reduces the overlap by spreading cells in high density region.

To extend [9] for 3D ICs, cells are not moved across dies during placement in [1] because they are already assigned into dies by the partitioner. In addition,  $\mathbf{f}_x^{\text{den}}$  is computed die-by-die based on the placement density  $D_d$  of each die  $d$ , which is defined as

$$D_d(x, y) = D_d^{\text{cell}}(x, y) - D_d^{\text{die}}(x, y), \quad (25)$$

where  $D_d^{\text{cell}}$  is the cell density on die  $d$ , and  $D_d^{\text{die}}$  is the die capacity scaled to match the total cell area on the die. Then, the placement potential  $\Phi_d$  is computed by solving Poisson's equation

$$\Delta \Phi_d(x, y) = -D_d(x, y). \quad (26)$$

The target point  $\hat{x}_i^d$  to connect density-based spring of cell  $i$  is computed by

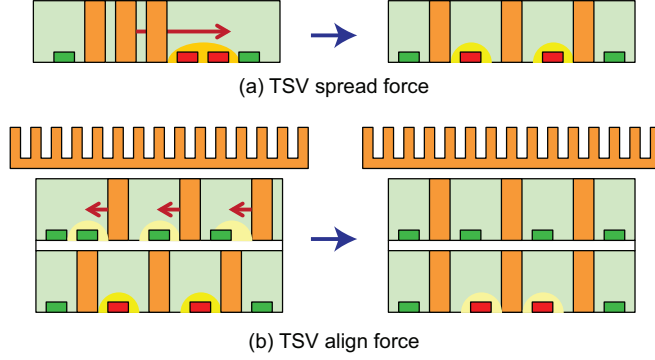
$$\hat{x}_i^d = x'_i - \frac{\partial}{\partial x} \Phi_d(x, y) \Big|_{(x'_i, y'_i)}, \quad (27)$$

where  $x'_i$  is the x-position of cell  $i$  on die  $d$  from the last iteration. Lastly, for each placement iteration, the placement result can be obtained by setting total force  $\mathbf{f}_x$  to zero, and solve

$$\mathbf{f}_x = \mathbf{f}_x^{\text{net}} + \mathbf{f}_x^{\text{hold}} + \mathbf{f}_x^{\text{den}} = \mathbf{0}. \quad (28)$$

### 5.2.3 TSV Spread and Alignment

In this algorithm, one of thermal properties of TSVs is exploited to help alleviate thermal problems as shown in Figure 45(a). TSVs occupy placement area, but do not dissipate power. The existence of TSVs among cells with high power dissipation reduces local dissipated power density, which in-turn helps reduce local temperature. Therefore, spreading TSVs evenly on each die should help reduce intra-die thermal variation in 3D ICs. This algorithm is proposed because it is simple yet effective. It can be viewed as a method to mimic uniform TSV position. Instead of moving TSVs based on the placement density computed from both TSV and cell area, TSVs are moved based on TSV density only. In other words,  $D_d^{\text{cell}}$  in Equation (25) is computed from TSV area only, and  $D_d^{\text{die}}$  is scaled to match the total TSV area on the die.



**Figure 45:** TSV spread and TSV align forces.

In addition to TSV spread, another thermal property of TSVs is exploited to help alleviate thermal problems as shown in Figure 45(b). TSVs conduct majority of heat through polymer adhesive between dies, causing local hot spots on the adjacent die between the TSVs and heatsink. Therefore, aligning TSVs on each die to TSVs on the adjacent die should help prevent this kind of hot spots, and direct the heat toward the heatsink quickly, resulting in overall temperature decrease. To align TSVs during global placement, an

additional force for TSVs, alignment force denoted  $\mathbf{f}_x^{\text{align}}$ , is introduced into Equation (28). This force can be represented by alignment springs connected to TSVs, and defined as

$$\mathbf{f}_x^{\text{align}} = \mathring{\mathbf{C}}_x^a(\mathbf{x} - \mathring{\mathbf{x}}^a), \quad (29)$$

where vector  $\mathring{\mathbf{x}}^a$  represents the x-position of target points to connect alignment springs to TSVs, and diagonal matrix  $\mathring{\mathbf{C}}_x^a$  collects spring constants  $\mathring{w}_{x,i}^a$  of the alignment spring connected to TSV  $i$ .

Alignment force is applied to TSV  $i$  only when its closest TSV  $j$  on the adjacent die farther from the heatsink is within a certain range so that wirelength does not excessively increase. The range is set to the size of TSV because of the high probability of aligning the TSVs in few iterations. To balance  $\mathbf{f}_x^{\text{align}}$  against other forces,  $\mathring{w}_{x,i}^a$  is set to density-based spring constant  $\mathring{w}_{x,i}^d$  of  $\mathbf{f}_x^{\text{den}}$ , and alignment target point  $\mathring{x}_i^a$  is set to  $x'_j$ , the x-position from last iteration of TSV  $j$  (on the adjacent die farther from heatsink) closest to TSV  $i$ . This method naturally balances  $\mathbf{f}_x^{\text{align}}$  against  $\mathbf{f}_x^{\text{den}}$ .

The intuition is that because of the high cell overlap in the early placement iterations, the target point  $\mathring{x}_i^d$  is farther away from TSV  $i$  than the alignment target point  $\mathring{x}_i^a$ . Thus,  $\mathbf{f}_x^{\text{den}}$  dominates. When cells are evenly distributed in the late iterations of placement,  $\mathring{x}_i^d$  is closer to TSV  $i$ . Then,  $\mathbf{f}_x^{\text{den}}$  becomes weaker, and  $\mathbf{f}_x^{\text{align}}$  affects the TSV position more.

#### 5.2.4 Thermal Coupling-aware Placement

In this algorithm, the die-to-die thermal coupling is considered during placement. The basic approach is to introduce two new forces, the first that moves cells and the second that moves TSVs, both in an attempt to place high-power cells close to the TSV-to-heatsink path. Since the heat dissipated by a cell must flow toward heatsink, cells are placed based on their power density and the effective thermal conductivity computed using the same die *and the dies above*. In addition, since TSV conducts heat without raising temperature too much, TSVs are placed based on the total power density of the same die *and the dies below*.

The basic approach is that the area with high power density and low thermal conductivity leads to high temperature. Thus, the temperature at a certain position depends on the difference (or imbalance) between power density and thermal conductivity. The force

that moves cells (TSVs) on a die also changes the power density (thermal conductivity) distribution of the die. The goal is to use these forces to balance the power density and the thermal conductivity at each position on the die. The force in an area with high difference should be stronger than the force in an area with low difference. The strength of a spring force depends on the distance to the connection point, so the strength is set based on this difference. Based on this concept, a map of the difference is built first, and then smoothed in an iterative fashion. The notations used in this section are shown in Table 18.

**Table 18:** Notations used for thermal coupling-aware placement.

$P_d^{\text{cell}}$	cell power density of each die $d$
$K_d^{\text{sink}}$	effective thermal conductivity from die $d$ to heatsink
$K_d^{\text{die}}$	thermal conductivity across the opposite sides of die $d$
$p_i$	power of cell $i$
$N_d^{\text{TSV}}$	total number of TSVs on die $d$
$N_{\text{die}}$	number of dies
$B_d^{\text{cond}}$	balance factor for the thermal conductivity-based force on die $d$
$s_d^{\text{cond}}$	scaling factor to match the effective thermal conductance to heatsink to cell power on die $d$
$B_d^{\text{pow}}$	balance factor for the power density-based force on die $d$
$s_d^{\text{pow}}$	scaling factor to match the cell power of die $d$ and below to the thermal conductance of die $d$
$s_d^{\text{PD}}$	scaling factor to normalize the cell power to the cell area on die $d$
$s_d^{\text{KD}}$	scaling factor to normalize the thermal conductance of die $d$ to the cell area on die $d$
$\alpha$	weighting constant for thermal coupling forces

#### 5.2.4.1 For Cell Movement

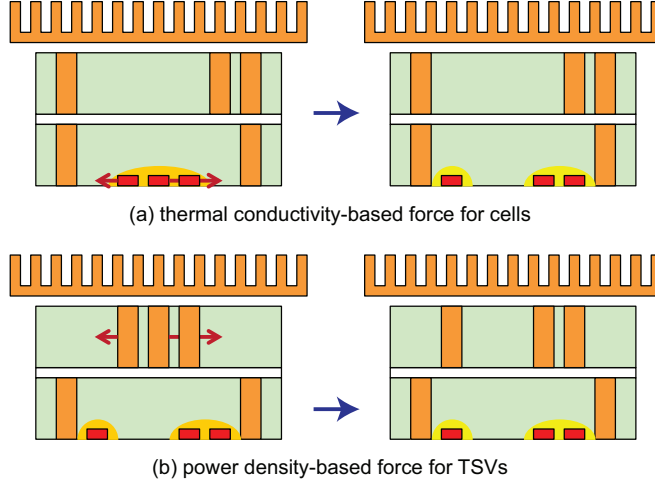
The thermal conductivity-based force  $\mathbf{f}_x^{\text{cond}}$  is introduced as illustrated in Figure 46(a). It moves high-power cells toward the position with high thermal conductivity to heatsink, and



is defined as

$$\mathbf{f}_x^{\text{cond}} = \mathring{\mathbf{C}}_x^c(\mathbf{x} - \mathring{\mathbf{x}}^c), \quad (30)$$

where the vector  $\mathring{\mathbf{x}}^c$  represents the x-position of target points to connect thermal conductivity-based springs to cells, and the diagonal matrix  $\mathring{\mathbf{C}}_x^c$  contains spring constants  $\mathring{w}_{x,i}^c$  of the spring connected to cell  $i$ .



**Figure 46:** Thermal conductivity-based vs power density-based forces.

Here,  $\mathbf{f}_x^{\text{cond}}$  is computed die-by-die by balancing the cell power density  $P_d^{\text{cell}}$  of each die  $d$  against its effective thermal conductivity to heatsink, denoted  $K_d^{\text{sink}}$ . Under the demand-supply system of the force-directed framework in [9],  $P_d^{\text{cell}}$  and  $K_d^{\text{sink}}$  represent the demand and supply to remove the heat from die  $d$  in the 3D stack. The thermal conductivity-based balance factor  $B_d^{\text{cond}}$  for die  $d$  is defined as (see Figure 47)

$$B_d^{\text{cond}}(x, y) = P_d^{\text{cell}}(x, y) - s_d^{\text{cond}} \cdot K_d^{\text{sink}}(x, y), \quad (31)$$

where  $s_d^{\text{cond}}$  is a scaling factor to match  $K_d^{\text{sink}}$  to  $P_d^{\text{cell}}$  across the die. Here,  $s_d^{\text{cond}}$  is used to balance the total supply ( $K_d^{\text{sink}}$ ) and the total demand ( $P_d^{\text{cell}}$ ), and computed by

$$s_d^{\text{cond}} = \frac{\int \int P_d^{\text{cell}}(x, y) dx dy}{\int \int K_d^{\text{sink}}(x, y) dx dy}. \quad (32)$$

Here,  $K_d^{\text{sink}}$  is computed as

$$K_d^{\text{sink}}(x, y) = \frac{1}{\sum_{j=d}^{N_{\text{die}}} \frac{1}{K_j^{\text{die}}(x, y)}}, \quad (33)$$

where  $K_j^{\text{die}}$  is the thermal conductivity of die  $j$ , and die  $N_{\text{die}}$  is the die closest to the heatsink (see Figure 48). Here,  $K_{N_{\text{die}}}^{\text{die}}$  includes the thermal conductivity of the thick substrate and heatsink, and  $K_j^{\text{die}}$  is computed based on the TSV density at each position on the die and the average thermal conductivity of bulk silicon with and without TSVs, obtained from the simulation results in Section 5.1.

The potential  $\Phi_d^{\text{cond}}$  for  $B_d^{\text{cond}}$  is computed by solving Poisson's equation

$$\Delta \Phi_d^{\text{cond}}(x, y) = -B_d^{\text{cond}}(x, y). \quad (34)$$

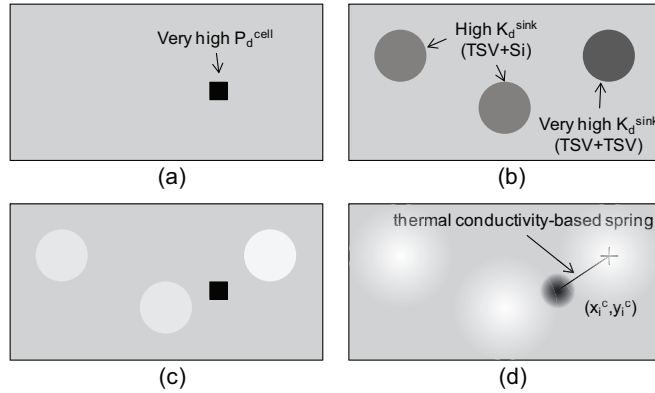
The target point  $\hat{x}_i^c$  of cell  $i$  is computed by

$$\hat{x}_i^c = x'_i - \frac{\partial}{\partial x} \Phi_d^{\text{cond}}(x, y) \Big|_{(x'_i, y'_i)}, \quad (35)$$

where  $x'_i$  is the x-position of cell  $i$  on die  $d$  from the last iteration. Spring constant  $\hat{w}_{x,i}^c$  for cell  $i$  is set based on cell power and the total cell power by

$$\hat{w}_{x,i}^c = p_i / \sum_{\forall j} p_j, \quad (36)$$

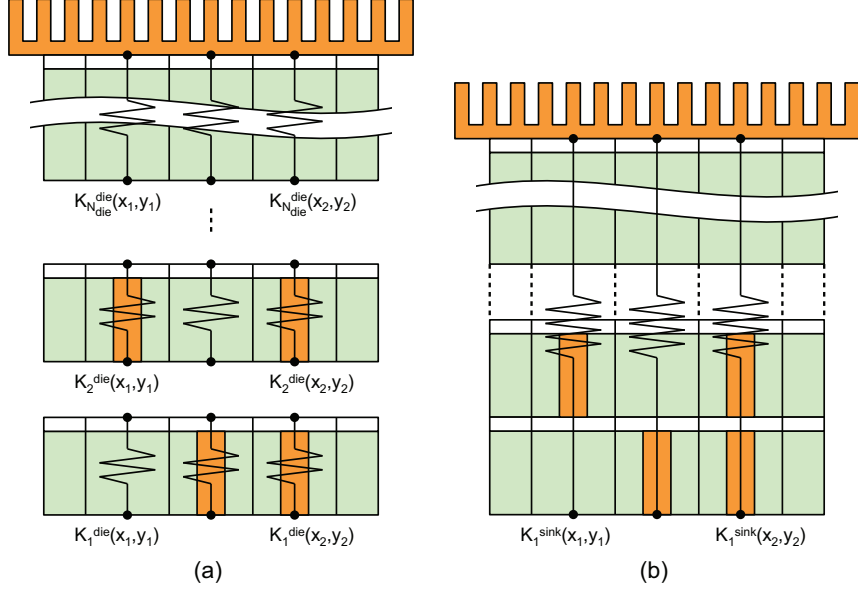
where  $p_i$  is the power of cell  $i$ , and  $j$  is a cell on die  $d$ . Therefore, a high-power cell is connected to a strong thermal conductivity-based spring.



**Figure 47:** Illustration of  $B_d^{\text{cond}}$ . (a)  $P_d^{\text{cell}}$ , (b)  $s_d^{\text{cond}} \cdot K_d^{\text{sink}}$ , (c)  $B_d^{\text{cond}}$ , (d) potential for  $B_d^{\text{cond}}$  after solving Poisson's equation.

#### 5.2.4.2 For TSV Movement

Power density-based force  $\mathbf{f}_x^{\text{pow}}$  is introduced as illustrated in Figure 46(b). It moves TSVs toward the position with high cell power density on the same die and the dies below, and



**Figure 48:** Computation of  $K_d^{\text{sink}}$ . (a)  $K_j^{\text{die}}$ , (b)  $K_1^{\text{sink}}$ .

defined as

$$\mathbf{f}_x^{\text{pow}} = \mathring{\mathbf{C}}_x^{\text{p}}(\mathbf{x} - \hat{\mathbf{x}}^{\text{p}}), \quad (37)$$

where the vector  $\hat{\mathbf{x}}^{\text{p}}$  represents the x-position of target points to connect power density-based springs to TSVs, and the diagonal matrix  $\mathring{\mathbf{C}}_x^{\text{p}}$  contains spring constants  $\mathring{w}_{x,i}^{\text{p}}$  of the spring connected to TSV  $i$ .

Here,  $\mathbf{f}_x^{\text{pow}}$  is computed die-by-die by balancing the thermal conductivity  $K_d^{\text{die}}$  of each die  $d$  against the total power density  $\sum P_j^{\text{cell}}$  that flows through the die toward heatsink. Under the demand-supply system of the force-directed framework in [9],  $K_d^{\text{die}}$  and  $\sum P_j^{\text{cell}}$  represent the demand and supply to conduct heat from the same die and dies below to heatsink. The power density-based balance factor  $B_d^{\text{pow}}$  for die  $d$  is defined as

$$B_d^{\text{pow}}(x, y) = K_d^{\text{die}}(x, y) - s_d^{\text{pow}} \cdot \sum_{j=1}^d P_j^{\text{cell}}(x, y), \quad (38)$$

where  $s_d^{\text{pow}}$  is a scaling factor to match  $\sum P_j^{\text{cell}}$  to  $K_d^{\text{die}}$  across the die. Here,  $s_d^{\text{pow}}$  is used to balance the total supply ( $\sum P_j^{\text{cell}}$ ) and the total demand ( $K_d^{\text{die}}$ ), and computed by

$$s_d^{\text{pow}} = \frac{\int \int K_d^{\text{die}}(x, y) dx dy}{\int \int \sum_{j=1}^d P_j^{\text{cell}}(x, y) dx dy}. \quad (39)$$

The potential  $\Phi_d^{\text{pow}}$  for  $B_d^{\text{pow}}$  is computed by solving Poisson's equation

$$\Delta \Phi_d^{\text{pow}}(x, y) = -B_d^{\text{pow}}(x, y). \quad (40)$$

The target point  $\hat{x}_i^{\text{p}}$  of TSV  $i$  is computed by

$$\hat{x}_i^{\text{p}} = x'_i - \frac{\partial}{\partial x} \Phi_d^{\text{pow}}(x, y) \Big|_{(x'_i, y'_i)}, \quad (41)$$

where  $x'_i$  is the x-position of TSV  $i$  on die  $d$  from the last iteration. Spring constant  $\hat{w}_{x,i}^{\text{p}}$  is set to  $1/N_d^{\text{TSV}}$ , where  $N_d^{\text{TSV}}$  is the total number of TSVs on die  $d$ . Therefore, the power density-based spring for each TSVs has the same strength.

#### 5.2.4.3 Balancing the Forces

The new forces are balanced against  $\mathbf{f}_x^{\text{den}}$  because  $\mathbf{f}_x^{\text{den}}$  is the main force that moves cells and TSVs. First, the new forces are scaled so that they have the same magnitude as  $\mathbf{f}_x^{\text{den}}$ . Then, weighting constants are applied to  $\mathbf{f}_x^{\text{den}}$ ,  $\mathbf{f}_x^{\text{cond}}$ , and  $\mathbf{f}_x^{\text{pow}}$  so that their contribution to the total force can be controlled.

First, to scale  $\mathbf{f}_x^{\text{cond}}$  to  $\mathbf{f}_x^{\text{den}}$ ,  $P_d^{\text{cell}}$ , the demand for  $B_d^{\text{cond}}$  in Equation (31), is normalized to  $D_d^{\text{cell}}$  by a scaling factor  $s_d^{\text{PD}}$  defined as

$$s_d^{\text{PD}} = \frac{\int \int D_d^{\text{cell}}(x, y) dx dy}{\int \int P_d^{\text{cell}}(x, y) dx dy}. \quad (42)$$

Then,  $P_d^{\text{cell}}$  in Equation (31) and Equation (32) is replaced by  $s_d^{\text{PD}} \cdot P_d^{\text{cell}}$ . Second, to scale  $\mathbf{f}_x^{\text{pow}}$  to  $\mathbf{f}_x^{\text{den}}$ ,  $K_d^{\text{die}}$ , the demand for  $B_d^{\text{pow}}$  in Equation (38), is normalized to  $D_d^{\text{cell}}$  by a scaling factor  $s_d^{\text{KD}}$  defined as

$$s_d^{\text{KD}} = \frac{\int \int D_d^{\text{cell}}(x, y) dx dy}{\int \int K_d^{\text{die}}(x, y) dx dy}. \quad (43)$$

Then,  $K_d^{\text{die}}$  in Equation (38) and Equation (39) is replaced by  $s_d^{\text{KD}} \cdot K_d^{\text{die}}$ .

Both  $\mathbf{f}_x^{\text{cond}}$  and  $\mathbf{f}_x^{\text{pow}}$  are scaled to  $\mathbf{f}_x^{\text{den}}$  based on  $D_d^{\text{cell}}$ , not on the gradient of  $\Phi_d$  because of the stability issue. After normalizing  $P_d^{\text{cell}}$  and  $K_d^{\text{die}}$  to  $D_d^{\text{cell}}$  as shown in Equation (42) and Equation (43), the magnitude of  $B_d^{\text{cond}}$  and  $B_d^{\text{pow}}$  and gradient of their potential are properly normalized. At an equilibrium, a small magnitude of the gradients results in a small magnitude of  $\mathbf{f}_x^{\text{cond}}$  and  $\mathbf{f}_x^{\text{pow}}$ . If  $\mathbf{f}_x^{\text{cond}}$  and  $\mathbf{f}_x^{\text{pow}}$  were scaled to  $\mathbf{f}_x^{\text{den}}$  based on the

gradient of  $\Phi_d$  instead, the magnitude of the gradient of potential of  $B_d^{\text{cond}}$  and  $B_d^{\text{pow}}$  would be exaggerated after the normalization, which in turn cause instability.

In summary,  $\mathbf{f}_x^{\text{cond}}$  moves cells in such a way that high power density flows through the paths with high thermal conductivity to heatsink. In addition,  $\mathbf{f}_x^{\text{pow}}$  moves TSVs in such a way that each TSV establishes a heat path for the high-power cells in the same die and the dies below. The overall force equation is expressed as follows:

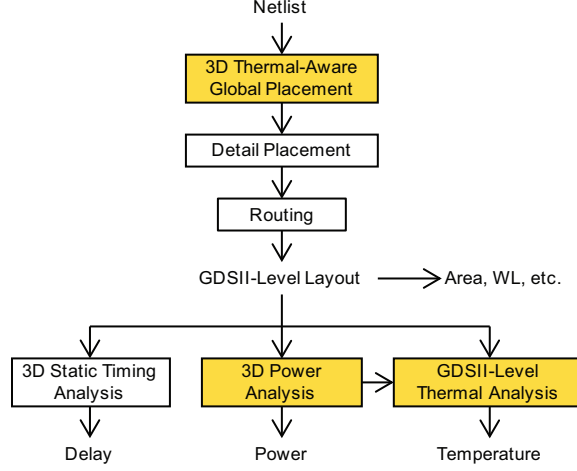
$$\mathbf{f}_x = \mathbf{f}_x^{\text{net}} + \mathbf{f}_x^{\text{hold}} + (1 - \alpha)\mathbf{f}_x^{\text{den}} + \alpha(\mathbf{f}_x^{\text{cond}} + \mathbf{f}_x^{\text{pow}}) = \mathbf{0}. \quad (44)$$

By increasing  $\alpha$ , the forces  $\mathbf{f}_x^{\text{cond}}$  and  $\mathbf{f}_x^{\text{pow}}$  dominate the movement of cells and TSVs for additional thermal optimization. The impact of  $\alpha$  is studied in Section 5.4.2.

### 5.3 Evaluation Flow

In this section, the framework to evaluate the impact of TSVs on temperature of 3D ICs is presented. The main components of the framework are power analysis and GDSII-level thermal analysis for 3D ICs. The presented evaluation flow allows evaluation of the effectiveness of the proposed 3D temperature-aware global placement algorithms in reducing temperature. The result of the study is analyzed, and reported in detail in Section 5.4.

The evaluation flow for temperature-aware 3D-IC global placement is shown in Figure 49. After obtaining 3D temperature-aware global placement result, detail placement and detail routing are performed. Traditional metrics, e.g., area and routed wirelength, of the final GDSII-level layout are reported. Then, 3D static timing analysis, power analysis, and GDSII-level thermal analysis are performed to report delay, power, and temperature, respectively. How Cadence SoC Encounter and Synopsys PrimeTime PX are used for accurate power analysis for 3D ICs is explained in Section 5.3.1. How thermal analysis is performed from GDSII-level layouts of 3D ICs by using ANSYS FLUENT together with a layout analyzer is explained in Section 5.3.2. Note that the result from power analysis needs to be presented to GDSII-level thermal analysis because logic cell power is the heat source in 3D ICs during thermal analysis.



**Figure 49:** Evaluation flow for temperature-aware 3D-IC global placement.

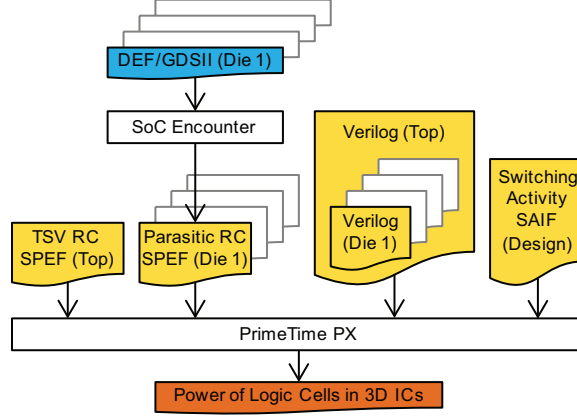
### 5.3.1 Power Analysis for 3D ICs

The power analysis flow for 3D ICs developed in this work starts by obtaining the layout of all dies in a 3D IC in DEF or GDSII format (see Figure 50). Next, they are presented to Cadence SoC Encounter to extract parasitic resistance and capacitance in SPEF format. A separate SPEF file for parasitic resistance and capacitance of TSVs is generated. The top-level verilog connects the verilog of all dies together, and the connection of all dies inside this top-level verilog represents TSVs. The switching activity of all logic cells in the whole design can be obtained by propagating switching probability, as well as static state probability, from all primary inputs into all nets of the design. Additional accuracy can be gained by performing functional simulation of the whole design. Finally, PrimeTime PX is used to perform static power analysis, and reports power dissipation of each logic cell. By stitching all the dies in this method, the parasitic resistance and capacitance of TSVs and wires running across dies also account for the total power of the 3D IC.

### 5.3.2 GDSII-Level Thermal Analysis

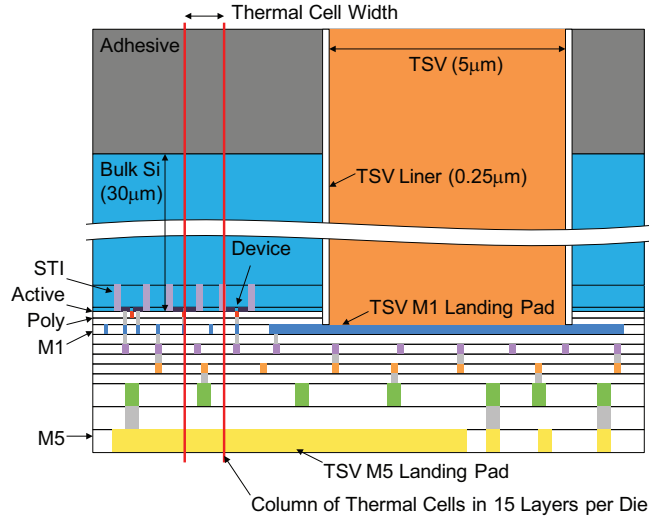
Steady-state temperature of a point  $\mathbf{p} = (x, y, z)$  inside a 3D structure can be obtained by solving the heat equation

$$\nabla \cdot (k(\mathbf{p})\nabla T(\mathbf{p})) + S_h(\mathbf{p}) = 0, \quad (45)$$



**Figure 50:** Power analysis flow for 3D ICs.

where  $k$  is thermal conductivity in  $\text{W/m}\cdot\text{K}$ ,  $T$  is temperature in  $\text{K}$ , and  $S_h$  is volumetric heat source in  $\text{W/m}^3$ . This model can be implemented by meshing analyzed structure of a 3D IC into elements as shown in Figure 51. Each element, called a thermal cell, is a volume of specific width and height, and its thickness is the same as each physical layer inside the 3D IC.

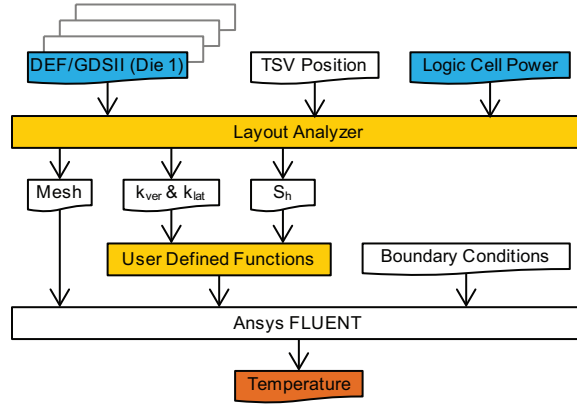


**Figure 51:** Analyzed structure of a TSV-based 3D IC. Each die is modeled with 15 layers of different materials. The entire 4-die structure contains 62 layers.

To solve Equation (45), boundary conditions must be given on the six surfaces of a 3D chip stack. Generally, a 3D chip stack is very thin and flat, and packaged inside molding materials, which are not good thermal conductor. The majority of heat flows from the stack toward the heatsink. Therefore, adiabatic boundary condition is applied on bottom and

four sides of the stack, and convective boundary condition is applied on the top side, which is the heatsink.

The thermal analysis flow developed in this work is shown in Figure 52. It starts by presenting the layout of all dies in a 3D IC in GDSII format and power dissipation of each logic cell to the layout analyzer that is developed for this work. The position of all TSVs is also presented to the layout analyzer so that all TSV related elements, e.g., landing pad and liner, are included into consideration. The layout analyzer automatically generates meshed structure of the 3D IC along with thermal conductivity and volumetric heat source of each thermal cell.



**Figure 52:** GDSII layout-level thermal analysis flow.

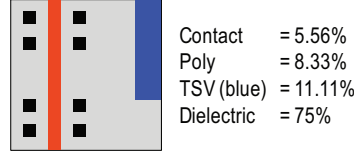
A thermal cell can be composed of several different materials, for example, polysilicon, tungsten in vias, copper in TSVs, and dielectric (see Figure 53). With sufficiently fine thermal cell size, equivalent thermal conductivity based on thermal resistive model can be used [56]. In theory, if a thermal cell size is very small, material inside it is homogeneous, and its thermal conductivity is isotropic. However, using small cell size requires high computing resource and long run time. For practical purpose, large thermal cell size can be used. Because of typical structural geometries found in GDSII layouts, thermal conductivity of each thermal cell is anisotropic. Vertical thermal conductivity  $k_{\text{ver}}$  and lateral thermal conductivity  $k_{\text{lat}}$  of a thermal cell consisting of  $N$  materials can be computed from

$$k_{\text{ver}} = r_1 \cdot k_1 + r_2 \cdot k_2 + \cdots + r_N \cdot k_N, \quad (46)$$

$$1/k_{\text{lat}} = r_1/k_1 + r_2/k_2 + \cdots + r_N/k_N, \quad (47)$$



where  $r_i$  is the ratio of material  $i$  volume to thermal cell volume, and  $k_i$  is the thermal conductivity of material  $i$ . The layout analyzer computes  $r_i$  directly from GDSII layout of all dies in the 3D chip stack.



**Figure 53:** Material composition inside a thermal cell.

From the power dissipation and position of each logic cell, total power dissipated inside a thermal cell  $P_{\text{cell}}$  can be computed. Then, volumetric heat source  $S_h$  can be computed from

$$S_h = \frac{P_{\text{cell}}}{W_{\text{cell}} \cdot H_{\text{cell}} \cdot T_{\text{cell}}}, \quad (48)$$

where  $W_{\text{cell}}$ ,  $H_{\text{cell}}$ , and  $T_{\text{cell}}$  are width, height, and thickness of the thermal cell, respectively.

Equation (45) is solved by using ANSYS FLUENT, a commercial tool. Meshed structure generated from the layout analyzer can be presented to FLUENT directly. However,  $k_{\text{ver}}$ ,  $k_{\text{lat}}$ , and  $S_h$  need to be presented to FLUENT through user defined functions because they vary with thermal cell position. Finally, with the boundary conditions described earlier, FLUENT can be run to obtain steady-state temperature of all positions inside a 3D chip stack.

#### 5.4 Experimental Results

In this work, 45-nm technology from FreePDK45 is used for experiments. TSV diameter is  $5 \mu\text{m}$ , and the landing pad width is  $7 \mu\text{m}$ . TSV liner thickness is 250 nm [57]. Copper TSVs with  $\text{SiO}_2$  liner [57] and 2.6- $\mu\text{m}$ -thick benzocyclobutene bonding adhesive [23] is used for experiments. Each die in the 3D chip stack is thinned to  $30 \mu\text{m}$  except that the topmost die, which is attached to heatsink, retains its thickness at  $530 \mu\text{m}$ . The ambient temperature on top of the heatsink is 300 K. The TSV parasitic resistance and capacitance are  $0.1 \Omega$  and  $125 \text{ fF}$ , respectively. All experiments are based on 4-die chip stacks.

IWLS 2005 benchmarks and several industrial circuits from OpenCores are used. The circuits are synthesized using Synopsys Design Compiler to obtain gate-level netlist, and

the target clock period of each circuit is used when performing all analyses. The benchmark characteristics are listed in Table 19. The numbers of TSVs are based on partitioning results from an implementation of [4]. The same die partitioning results are used for all algorithms for fair comparison in Section 5.4.3. Because [4] does not consider TSV area, it inserts high number of TSVs, resulting in low placement utilization.

**Table 19:** Benchmark circuits.

Ckt.	#Gates	#TSVs	Util.	Footprt (mm <sup>2</sup> )	Profile
ckt1	119,040	5,725	0.66	$0.50 \times 0.50$	Data encryption
ckt2	191,420	24,540	0.63	$0.90 \times 0.90$	Graphic accelerator
ckt3	280,933	17,362	0.49	$0.98 \times 0.98$	Video compression
ckt4	383,329	17,436	0.53	$1.04 \times 1.04$	Signal processing
ckt5	644,357	15,024	0.53	$1.16 \times 1.16$	Image encoder

The circuits are not optimized after placement because buffers and sized gates can change power profile, thus affecting temperature. The results reported in this work are from commercial tools. Cadence Encounter is used to route the layouts, Synopsys PrimeTime is used to analyze timing and power, and ANSYS FLUENT is used to analyze temperature. All temperature results are reported in terms of the increase from the ambient temperature measured at the top of the heatsink.

#### 5.4.1 Impact of TSV Density Uniformity

In this experiment, how TSV density uniformity impacts thermal profile is shown. Two baseline 3D placements are wirelength-driven placement with uniform TSV position [1] and wirelength-driven placement with non-uniform TSV position [1]. First, both baseline placements are obtained using an implementation of [1]. Then, power and thermal analyses are performed on both placement results. The routed wirelength, longest path delay, and power are shown in Table 20, and temperatures are shown in Table 21. Although the placement with non-uniform TSV position has shorter wirelength, better timing, and lower power than the placement with uniform TSV position, its temperature, especially the thermal variation, is worse. Both the non-uniform power density and the non-uniform thermal conductivity, caused by the non-uniform distribution of TSVs in the 3D chip stack, contribute to the

problem. In the placement with non-uniform TSV position, the area with high TSV density has low power density and low temperature, vice versa. These two opposite trends are responsible for high thermal variation.

**Table 20:** Routed wirelength, longest path delay, and power of placements with uniform [1] and non-uniform [1] TSV position.

	Uniform			Non-uniform		
	rWL	D <sub>max</sub>	P	rWL	D <sub>max</sub>	P
Ckt.	(m)	(ns)	(W)	(m)	(ns)	(W)
ckt1	3.897	5.320	0.752	3.014	4.836	0.728
ckt2	11.718	16.510	2.661	7.744	13.694	2.463
ckt3	13.532	8.814	2.353	9.326	6.535	2.288
ckt4	19.355	20.788	2.710	12.457	12.515	2.640
ckt5	22.708	19.772	3.209	18.711	13.798	3.122
ratio	1.405	1.350	1.039	1.000	1.000	1.000

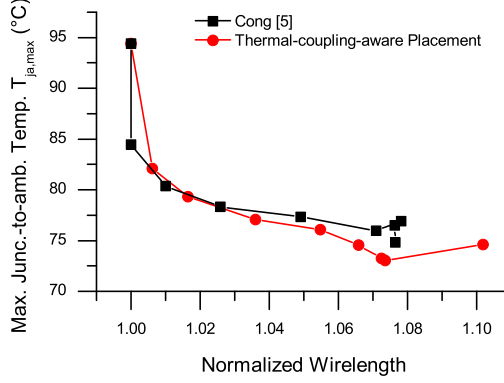
**Table 21:** Temperature (°C) of placements with uniform [1] and non-uniform [1] TSV position. ( $\Delta T_{ja} = T_{ja,max} - T_{ja,min}$ )

	Uniform			Non-uniform		
Ckt.	T <sub>ja,max</sub>	$\Delta T_{ja}$	T <sub>ja,ave</sub>	T <sub>ja,max</sub>	$\Delta T_{ja}$	T <sub>ja,ave</sub>
ckt1	71.55	17.60	64.50	74.13	18.33	63.98
ckt2	101.14	47.14	69.41	94.41	50.19	64.78
ckt3	70.38	31.01	55.06	80.09	42.81	55.48
ckt4	64.91	18.76	54.32	75.98	38.01	55.16
ckt5	66.77	35.40	53.13	75.24	39.32	54.50
ratio	1.000	1.000	1.000	1.081	1.325	0.995

#### 5.4.2 Temperature-Wirelength Trade-off

The proposed thermal-coupling-aware placement algorithm provides an efficient way to explore temperature-wirelength trade-off. The temperature-wirelength trade-off is studied in this experiment. By increasing the weighting constant  $\alpha$  in Equation (44), the placer increases the magnitude of  $\mathbf{f}_x^{\text{cond}}$  and  $\mathbf{f}_x^{\text{pow}}$  while decreasing  $\mathbf{f}_x^{\text{den}}$ , i.e., trading wirelength for temperature. The temperature-wirelength trade-off for ckt2 is shown in Figure 54. The placer from [5] is also implemented for comparison, and its trade-off curve is shown in Figure 54. A weighting constant  $\beta$  is used in [5].

With  $\alpha = \beta = 0$ , both placers perform as a wirelength-driven placer (= left-most points). As  $\alpha$  and  $\beta$  increase, temperature decreases while wirelength increases. It is observed that,



**Figure 54:** Temperature-Wirelength trade-off.

as  $\alpha$  and  $\beta$  increases, thermal coupling-aware placer outperforms [5]: the proposed placer provides shorter routed wirelength at the same temperature, and has lower temperature at the same wirelength. Note that [5] shows convergence problem with large  $\beta$  values. When [5] moves a high-power cell into a bin, it moves cells out of other bins in the dies above or below, resulting in potential wirelength increase and convergence problem as discussed in [5]. In addition, [5] does not consider vertical alignment of TSVs so that even if it moves high-power cells into a bin with many TSVs, the heat captured in the bin may not be easily dissipated vertically to the heatsink. The proposed algorithms overcome these limitations.

### 5.4.3 Comparison with State-of-the-Art

The proposed temperature-aware global placement algorithms are compared with the following recent state-of-the-art temperature-aware placers:<sup>2</sup>

[2] (force-directed placer): In this work, thermal analysis is performed at the beginning of every global placement iteration. The thermal gradient obtained from the analysis is used to compute repulsive force, which moves logic cells from high-temperature area toward low-temperature area. A version of this work is implemented by calling ANSYS FLUENT from inside the placer and combining scaled thermal gradient into density-based force  $\mathbf{f}_x^{\text{den}}$ .

[3] (force-directed placer): Instead of moving logic cells based on placement area density,

---

<sup>2</sup>This task is challenging due to the discrepancy among the settings and assumptions made in each work. However, best effort was made to provide fair and meaningful comparison, including in-depth discussions with the authors.

it moves logic cells based on placement power density. Therefore, logic cells are spread according to their power dissipation, and logic cells with high power dissipation occupy more space than logic cells with low power dissipation, leading to uniform power density and thermal profile across the die. A version of this work is implement.

[4] (partitioning-based placer): In this work, logic cells are partitioned into placement area and different dies based on the switching activity and parasitic capacitance of connecting wires and TSVs. Global routing is performed to determine the position of TSVs as proposed in [26] after performing global placement using an implementation of [4].

[5] (analytical placer): This method is implemented by balancing the power density combined across dies in vertical direction against the TSV density and solving the density for potential function. The gradient of the potential is used to compute a force to move cells and TSVs to maintain the balance. The force is added to  $\mathbf{f}_x^{\text{den}}$  with a user-defined parameter  $\beta$  to provide temperature-wirelength trade-off similar to the proposed work.

In Table 22, the routed wirelength, delay, power, and temperature comparison based on the GDSII layouts built by using these placers are shown. The wirelength, delay, and power values are normalized to the wirelength-driven non-uniform TSV placement [1] shown in Table 20. The temperature values are normalized to the wirelength-driven uniform TSV placement [1] shown in Table 21. Recall that non-uniform placer achieves high-quality wirelength, delay, and power results, whereas uniform placer leads to high-quality temperature values.

First, it is observed that [2] produces wirelength, delay, and power results comparable to non-uniform TSV placer [1]. In case of temperature, [2] obtains worse result compared with uniform TSV placer [1]. An attempt to increase the magnitude of thermal-gradient-based force was made, but large increase in wirelength without much additional temperature improvement was found. Moving cells out of a high-temperature area on a die may not reduce temperature if the high temperature is a result from thermal coupling with other dies. Also, without considering TSV thermal properties during thermal analysis, the thermal gradient does not capture the impact of TSVs on temperature accurately, thereby misleading the placement. Second, [3] obtains wirelength and delay results that are significantly worse than

**Table 22:** Comparison with state-of-the-art temperature-aware placers [2, 3, 4, 5, 1]. The proposed placers are TSA (TSV spread and alignment) and CA (Coupling-aware placement). The routed wirelength, delay, and power values are normalized to the non-uniform TSV placement [1] shown in Table 20. The temperature values are normalized to the uniform TSV placement [1] shown in Table 21.

Ckt.	routed wirelength (m)						max junc.-to-amb. temp, $T_{ja,max}$ ( $^{\circ}C$ )					
	[2]	[3]	[4]	[5]	TSA	CA	[2]	[3]	[4]	[5]	TSA	CA
ckt1	3.046	3.109	3.784	3.240	3.250	3.133	72.48	73.12	82.86	70.69	70.85	70.41
ckt2	7.740	8.780	14.924	8.349	7.892	8.314	91.70	74.21	101.00	76.89	100.19	73.05
ckt3	9.347	10.544	16.028	10.706	10.355	10.261	77.74	64.39	69.80	66.34	72.41	65.60
ckt4	12.480	13.902	19.871	15.234	14.901	14.545	73.79	62.43	80.11	60.14	65.50	59.31
ckt5	18.869	21.482	27.649	20.125	19.845	19.994	74.86	79.22	76.25	61.95	64.45	61.60
ratio	<b>1.005</b>	1.112	1.595	1.120	1.093	1.090	1.056	0.964	1.105	0.909	0.997	<b>0.895</b>
Ckt.	longest path delay (ns)						temp difference, $T_{ja,max} - T_{ja,min}$ ( $^{\circ}C$ )					
	[2]	[3]	[4]	[5]	TSA	CA	[2]	[3]	[4]	[5]	TSA	CA
ckt1	4.935	4.796	5.128	5.067	4.786	4.871	16.29	14.94	28.12	14.69	15.55	14.16
ckt2	13.679	15.004	15.231	14.416	13.588	14.785	46.96	15.15	51.16	22.39	53.87	17.15
ckt3	6.567	6.797	7.865	7.276	6.530	6.906	39.89	19.68	28.69	23.82	33.65	22.97
ckt4	12.518	12.695	16.158	13.609	13.695	13.113	35.46	16.69	39.76	15.87	21.83	14.27
ckt5	13.931	16.427	15.649	13.674	13.799	14.664	38.08	36.39	38.02	23.77	33.07	24.53
ratio	<b>1.007</b>	1.066	1.160	1.058	1.015	1.051	1.235	0.744	1.360	0.719	1.042	<b>0.673</b>
Ckt.	power consumption (W)						average temp, $T_{ja,ave}$ ( $^{\circ}C$ )					
	[2]	[3]	[4]	[5]	TSA	CA	[2]	[3]	[4]	[5]	TSA	CA
ckt1	0.729	0.734	0.776	0.736	0.736	0.732	63.80	63.70	69.52	63.32	63.27	63.35
ckt2	2.463	2.548	2.564	2.521	2.487	2.523	64.81	66.84	69.36	66.07	65.14	66.14
ckt3	2.290	2.331	2.351	2.318	2.306	2.321	55.41	55.49	55.97	54.53	54.14	55.08
ckt4	2.640	2.671	2.737	2.682	2.672	2.675	55.07	54.35	60.42	53.91	53.63	53.85
ckt5	3.127	3.194	3.255	3.166	3.130	3.156	54.51	55.08	57.97	53.22	51.91	52.90
ratio	<b>1.001</b>	1.019	1.043	1.015	1.009	1.014	0.994	0.999	1.059	0.984	<b>0.973</b>	0.984

non-uniform TSV placer. This is mainly because it moves logic cells based only on power density. However, this move helps reduce maximum temperature and thermal variation inside the 3D chip stack significantly. Although it attempts to spread power over placement area, TSVs obstruct this effort frequently.

Third, the routed wirelength and delay of results from [4] are worse than all other placers. The main reason is that [4] does not consider TSV area during placement. Thus, the TSVs inserted during routing affects the placement quality significantly. The maximum temperature, thermal variation, and average temperature are also worse than uniform TSV placer. The router tends to insert TSVs in the middle of the die to minimize wirelength, leaving low thermal conductivity at chip corners, thus high temperature. Fourth, although

the wirelength of result from [5] is worse than other placers, temperature improvement is among the best. Because the algorithm consider the impact of TSV on chip area and temperature, it utilizes TSVs effectively to help improve temperature results.

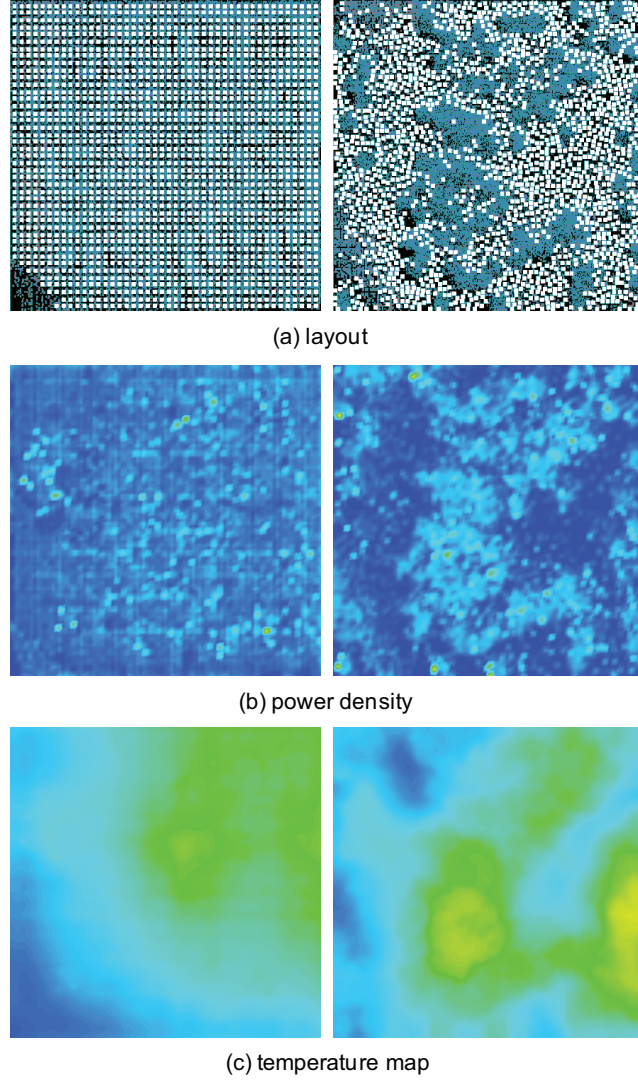
Fifth, the proposed TSV spread and alignment method (TSA) achieves comparable delay and power results at the cost of wirelength degradation compared with non-uniform placer. In case of temperature, TSA obtains better average temperature than uniform TSV and comparable maximum temperature and temperature difference. But, the wirelength of TSA method is significantly better than that of uniform TSV placer. These results show that TSA method is better in reducing wirelength while optimizing temperature compared with uniform TSV placer.

Lastly, the proposed thermal coupling-aware placement (CA) achieves the best temperature results among all placers [2, 3, 4, 5], including uniform TSV placer [1]. In particular, CA method outperforms uniform TSV placer by 10 % and 33 % in terms of maximum temperature and temperature difference. CA obtains 9 % worse wirelength and 5 % worse delay results compared with non-uniform TSV placer, but CA is among the best in terms of wirelength and delay among other placers [2, 3, 4, 5]. The power overhead is negligible. The TSVs in the placement by CA method are not spread as evenly as TSA placer and uniform TSV placer, but they are spread only sufficiently to help remove heat from the dies in the stack while maintaining high-quality wirelength. In addition, high-power logic cells are also placed effectively to dissipate heat using the nearby TSVs that are vertically aligned all the way to the heatsink.

#### **5.4.4 Power and Thermal Maps**

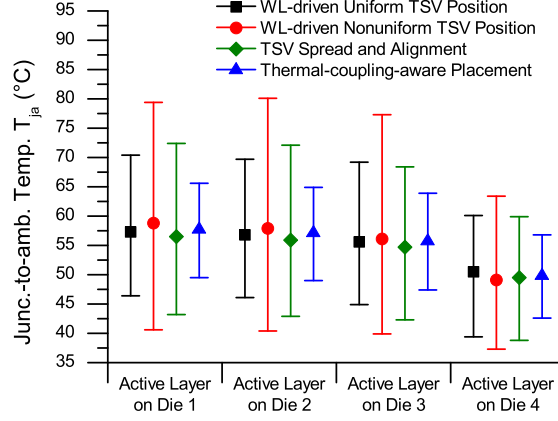
The snapshots of layout, power density, and temperature of placement results from previous experiments are shown in this section. The thermal maps of uniform TSV position [1] and wirelength-driven placement with non-uniform TSV position [1] are shown in Figure 55. In placement with uniform TSV position, power white space (TSVs) is evenly distributed, resulting in lower local power density and thus temperature than placement with non-uniform TSV position.

The temperature of all active layers in various placements is shown in Figure 56. By simply spreading and aligning TSVs across dies, the thermal variation becomes comparable to that of the uniform TSV placement. The temperature of thermal coupling-aware placement result is also shown in Figure 56. By considering thermal coupling, temperature results even better than the uniform TSV placement can be obtained.



**Figure 55:** Power and thermal profile of designs with uniform [1] (left) and non-uniform [1] (right) TSV position. (TSVs are in white in the layout. Area with low power density or temperature is in blue.)





**Figure 56:** Temperature of ckt3 placed by different placement algorithms. Die 1 is close to PCB, and Die 4 is close to heatsink.

#### 5.4.5 Runtime Results

The runtime of wirelength-driven placer with uniform TSV position [1], wirelength-driven placer with non-uniform TSV position [1], state-of-the-art temperature-aware placers [2, 3, 4, 5], and the proposed placers are shown in Table 23. The runtime for [2] includes running power analysis and thermal analysis between iterations. The runtime for [3, 5] and the proposed thermal coupling-aware placer includes running power analysis between iterations.

The runtime of all temperature-aware placement algorithms is roughly in the same magnitude. Except for TSA method, all other placement algorithms require power simulation (and thermal simulation in the case of [2]), resulting in larger runtime than [1].

**Table 23:** Runtime comparison of uniform TSV placement [1], non-uniform TSV placement [1], state-of-the-art temperature-aware placers [2, 3, 4, 5] and the proposed placers. The proposed placers are TSA (TSV spread and alignment) and CA (Coupling-aware placement).

Ckt.	runtime (min.)						TSA	CA
	[1]	[1]	[2]	[3]	[4]	[5]		
ckt1	13.04	11.07	19.89	25.78	31.15	21.00	9.06	24.04
ckt2	62.96	52.38	75.50	96.81	49.49	67.73	52.68	99.56
ckt3	45.05	42.46	78.09	127.29	65.10	95.70	53.35	102.40
ckt4	74.88	58.25	102.87	231.77	88.01	262.18	90.59	244.48
ckt5	169.10	229.08	293.14	388.36	165.47	652.01	168.51	423.03
total	365.04	393.24	569.50	870.01	399.22	1,098.62	374.18	893.52

## 5.5 *Summary*

In this chapter, a demonstration that temperature-aware placers must consider TSV thermal properties and die-to-die thermal coupling during placement is shown. Two temperature-aware placement algorithms for 3D ICs are presented. The methods effectively exploit the die-to-die thermal coupling in the 3D-IC stack. In the first algorithm, TSVs are spread on each die to reduce the local power density, and vertically aligned across dies simultaneously to increase thermal conductivity to the heatsink. In the second algorithm, high-power logic cells are moved to the location that has high thermal conductivity to the heatsink, and TSVs in the upper dies are moved so that high-power cells are vertically overlapping below the TSVs. These methods are employed in a force-directed 3D placement successfully. Experimental results show that the placers achieve the best temperature results among all placers used in the comparison.

## CHAPTER VI

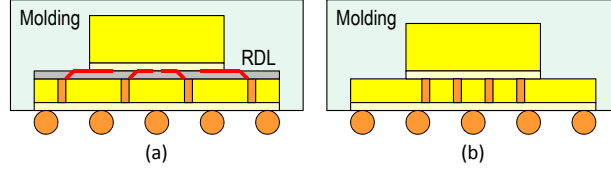
### BLOCK-LEVEL 3D-IC DESIGN QUALITY TRADE-OFFS STUDY FOR UNBALANCED DIE STACKING

Three-dimensional integrated circuits (3D ICs) are built in various design styles and at different levels. Wide-I/O memory technology provides extremely high memory bandwidth and very short core-to-memory connections [58]. Homogeneous 3D integration technology provides compact and high-degree of logic integration [59]. On the other hand, heterogeneous integration technology allows various electronic components such as analog circuits, memory elements, logic, and sensors in the same 3D-IC stack [60].

In many cases, 3D die stacking is done for the dies that have the identical footprint. The main motivation behind this stacking is to allow wafer-to-wafer bonding, which in turn reduces the cost compared with other alternatives such as die-to-wafer and die-to-die bonding. However, there are several cases where die-to-wafer bonding is more practical and costs less than wafer-to-wafer bonding. For example, in case of a two-die memory and logic stacking [61], it is very possible for these two dies to be built by different companies and thus have different footprint. In this case, wafer-to-wafer bonding is simply not possible, which makes die-to-wafer bonding the only practical choice. The wafer will have larger dies, and the smaller dies will be aligned and bonded to the wafer individually. The same argument applies to logic-to-logic stacking, where the two dies are from different companies. It is also possible that designers use IP blocks that may enforce them to use different footprint between the two dies.

In this work, how to design block-level 3D ICs, where the footprint of the dies in the stack are different, is studied. Among several possible configurations, a two-die 3D IC, where the bottom die has larger footprint than the top die, is focused. Both dies are facing down so that the heat sink is located above the back-side (= bulk) of the top die, and C4

bumps are below the front-side (= top metal layer) of the bottom die.<sup>1</sup> The design of the top die is further assumed to be fixed so that the block-level design of the bottom die is the main focus. Depending on the position of through-silicon-vias (TSVs) in the bottom die, a redistribution layer (RDL) is necessary on the back-side of the bottom die to connect the two dies as shown in Figure 57.



**Figure 57:** Side view of a 3D IC. (a) with RDLs and (b) without RDLs.

In this work, three different ways to place TSVs in the bottom die, namely TSV-farm, TSV-distributed, and TSV-whitespace, are investigated. Since each design style has its own advantages and disadvantages, these three design styles are compared in terms of area, wirelength, timing, power, temperature, and mechanical stress. In addition, the TSV size and pitch are varied to investigate their impact on each design style. These issues in practical 3D-IC designs are addressed in this work.

## 6.1 Background

### 6.1.1 Die Bonding and Redistribution Layers

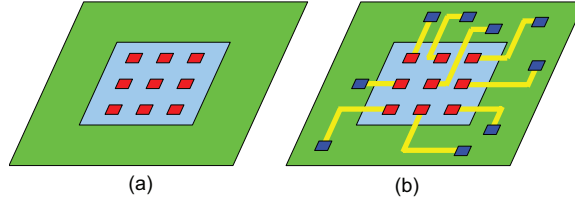
For die-to-wafer bonding in 3D ICs, three methods have been proposed: back-to-back, face-to-back, and face-to-face. If back-to-back bonding is utilized, a signal should go through two TSVs when it is transmitted from one die to its adjacent die. Since TSVs have nonnegligible capacitance, transferring a signal through two TSVs might degrade the delay and the signal integrity of the net. On the other hand, both the face-to-back and the face-to-face bonding methods enable less TSV capacitance overhead than the back-to-back bonding. In addition, since metal layers can be deposited on the back side of silicon dies, they also enable the flip chip packaging for 3D ICs. The face-to-face and the face-to-back bonding methods also help reduce temperature because a heatsink can be mounted on the back side of the top

<sup>1</sup>This stacking allows good power delivery and potentially good cooling if the top die consumes much power.

die.

When the face-to-face bonding is utilized between two dies with different die sizes, I/Os must be positioned on exposed top layer of the metal on the large die. In this case TSVs would not be needed. Wire-bonded packaging is usually utilized to connect the I/Os to the package pins. This kind of stacking, however, is not compatible with popular flip-chip packaging.

When the face-to-back bonding is utilized between two dies with different die sizes, redistribution-layer (RDL) routing on the back side of the bottom die is required in some cases. If all TSVs inserted in the bottom die are inside the footprint area of the top die as shown in Figure 58(a), the TSVs in the bottom die can be directly bonded to the bonding pads in the top die. However, if some TSVs in the bottom die are outside the footprint area of the top die, RDL routing is necessary to connect the TSVs to the bonding pads of the top die as illustrated in Figure 58(b).



**Figure 58:** RDL wires connecting TSVs on bottom die to bonding pads on top die.

Although the RDL allows connections between TSV landing pads on the back side of the bottom die and the bonding pads in the top die, it causes several negative effects. First of all, typical wires on the RDL are wide, possibly as wide as wires on the topmost metal layers. Thus, their parasitic capacitance is much higher than local metal wires, and causes timing degradation and dynamic power overhead. In addition, the large minimum pitch between adjacent wires in the RDL limits the minimum TSV pitch in a TSV array. For example, if four TSVs are placed in a  $2 \times 2$  array, they can be placed as close to each other as possible. However, if 25 TSVs are placed in a  $5 \times 5$  array, the TSV in the center cannot be routed by an escape routing unless the TSV pitch is several times greater than the minimum pitch.

### 6.1.2 The Goal of This Work

According to the above discussion, two options are available for the design of 3D ICs with different die sizes: insert all TSVs inside the footprint area of the top die so that RDL routing is not required, or insert TSVs wherever they are needed and perform RDL routing to connect them in the bottom die to the bonding pads in the top die. The former limits TSV positions, but it does not require RDL wires. The latter provides higher degree of freedom on TSV positions than the former option, but it requires RDL wires and routing. In addition, different TSV insertion styles lead to very different layout qualities. In this work, therefore, three different design styles: TSV-farm (without RDLs), TSV-distributed (with RDLs and regularly placed TSVs), and TSV-whitespace (with RDLs and irregularly placed TSVs), are compared. Since each design style has its own advantages and disadvantages, as many design metrics as possible are used so that the three design styles can be investigated in many different points of view. Two important design parameters, i.e., TSV diameter and pitch, impact the quality of each design. Therefore, these parameters are also varied, and their impact on various performance metrics is studied.

Since the 3D-IC design space is too large, the scope of this work is restricted to the following assumptions. Two dies are stacked, and face-to-back bonding is used between the two dies. Since placing the larger die of the two in the bottom of the 3D stack provides more benefits than the opposite case<sup>2</sup>, the bottom die is also assumed to be larger than the top die. A heatsink is mounted on the back side of the top die.

## 6.2 Block-Level 3D-IC Design

In this section, the block-level design styles (TSV-farm, TSV-distributed, and TSV-whitespace) are explained in detail. These three design styles are distinguished by how TSVs are distributed in the layout. In the TSV-farm style, TSVs are placed inside the footprint area of the top die. In the TSV-distributed style, TSVs are evenly distributed over the layout. In the TSV-whitespace style, TSVs are irregularly inserted.

---

<sup>2</sup>For example, the bottom die is connected to the package, so large bottom die area allows the chip to have many I/Os for power and ground.

### 6.2.1 Partitioning

In the first stage of all design styles, blocks are partitioned into two dies by a partitioner. During partitioning, various factors should be taken into account to control the quality of the 3D IC. The cut size, which is the number of cut 3D nets between the two dies, directly determines the number of TSVs. In addition, assigning low-power blocks to the bottom die and high-power blocks to the top die reduces temperature because the top die is closer to the heatsink than the bottom die. On the other hand, assigning thermally-sensitive blocks such as memory blocks to the top die and thermally-insensitive blocks to the bottom die increases predictability and reliability of the 3D IC. Because of these design factors, blocks are not moved across dies after partitioning. Partitioning are manually performed with all these factors considered.

### 6.2.2 TSV Insertion and Floorplanning

In the TSV-farm and the TSV-distributed styles, TSVs are preplaced in arrays and treated as obstacles during floorplanning. In the TSV-farm style, an array of TSVs (and bonding pads) are placed in the middle of the bottom die (and top die). In the TSV-distributed style, on the other hand, TSVs are placed all over the bottom die. Therefore, some of the TSVs exist outside the footprint area of the top die.

After preplacing TSVs, floorplanning of the blocks in the bottom die is manually performed. Since functional blocks and TSVs should not overlap, blocks are placed around the TSV farm in the TSV-farm style. Since the TSV farm area is usually large, if all blocks are highly connected, the TSV-farm design style causes significant wirelength overhead. On the other hand, if the interblock connectivity is not high, the farm in the center of the layout does not cause wirelength overhead. The TSV-distributed style might not cause significant wirelength overhead because, unlike one large TSV array in TSV-farm style, TSVs are grouped in small arrays in TSV-distributed style. However, some large blocks have very limited locations for their position because they cannot be placed in the space between adjacent TSV arrays. This design constraint might degrade wirelength, timing, and power. However, the TSV-distributed style is expected to show low temperature and small TSV

stress.

On the other hand, a 3D floorplanner is used to obtain 3D-IC layouts of the TSV-whitespace style. After floorplanning, TSVs are manually inserted into whitespace existing between blocks close to a pin of each 3D net. Therefore, TSVs are irregularly placed. When TSVs are inserted, the current floorplan is perturbed by moving blocks to create or expand whitespace. Since a 3D floorplanner is used, the TSV-whitespace design style is expected to optimize the wirelength better than all other design styles.

### **6.2.3 Bonding Pad Assignment and RDL Routing**

In the TSV-farm style, all TSVs exist inside the footprint of the top die. Therefore, the position of the bonding pads in the top die are duplicated from the position of the TSVs in the bottom die. In the TSV-distributed and the TSV-whitespace styles, the position of the bonding pads in the top die are determined by recursive bipartitioning before floorplanning of the top die. The recursive bipartitioning increases the routability of the RDL routing. After the bonding pad assignment in the top die, the blocks in the top die are floorplanned. The primary objective at this point is to minimize the wirelength. After floorplanning of the top die, RDL routing is performed in the TSV-distribute and the TSV-whitespace styles.

## **6.3 Design Evaluation**

In this section, a brief overview of the methodology to evaluate 3D-IC layouts is presented. Traditional metrics and reliability metrics are reported in this work. The traditional metrics are area, wirelength, timing, and power, and the reliability metrics are temperature and mechanical stress.

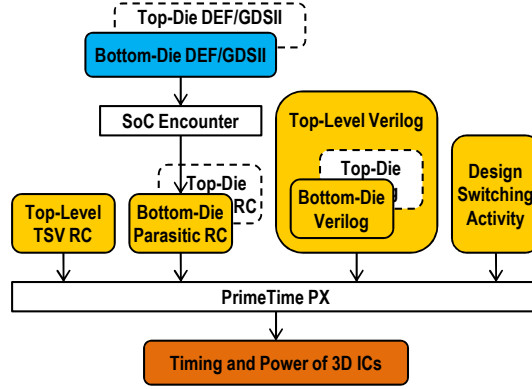
### **6.3.1 Traditional Metrics**

Traditional metrics such as area and wirelength are important for both 2D ICs and 3D ICs. These traditional metrics are obtained directly from layouts of both bottom and top dies.

Timing (longest path delay) and power analysis flow is shown in Figure 59. It is performed as follows. First, the parasitic resistance and capacitance of each die are extracted using Cadence QRC Extraction. Since the face-to-back die bonding style and 30- $\mu\text{m}$  thick



bottom die are assumed, the capacitive coupling between the bottom and the top dies are not included in resistance and capacitance extraction. Therefore, parasitic resistance and capacitance extraction of each die is performed separately. Parasitic resistance and capacitance of the RDL is also extracted. For 3D timing analysis, the top and the bottom dies are represented as modules in a top-level verilog file. A top-level SPEF file is also created. It includes not only the parasitic resistance and capacitance of both dies, but also resistance and capacitance of TSVs and the RDL wires. For accurate power analysis, switching activities of all logic cells is obtained by functional simulation of the whole chip. Synopsys PrimeTime is used to perform static timing and power analysis.

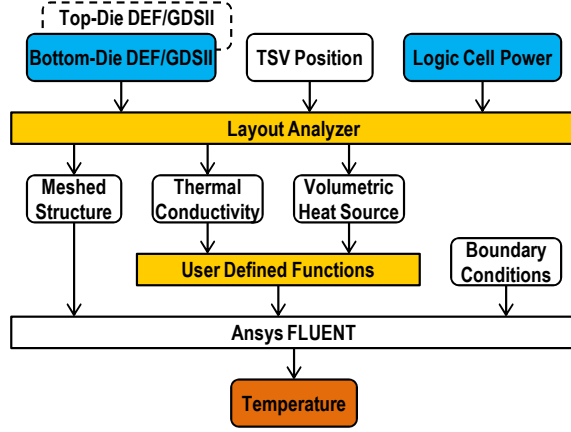


**Figure 59:** Timing and power analysis flow for die-to-wafer stacked 3D ICs.

### 6.3.2 Thermal Analysis

The thermal analysis flow used in this work is shown in Figure 60. For thermal analysis, ANSYS FLUENT is used. To perform thermal analysis, first, a meshed structure is created. Each grid, which is called a thermal cell, in the meshed structure contains material composition information such as copper density in the cell. This information is extracted from GDSII layout files which include logic cells in the blocks as well as TSVs. These files together with power dissipation of each logic cell in the blocks are presented to the layout analyzer. The layout analyzer automatically generates a meshed structure and layout information of each thermal cell from its inputs. The layout information of a thermal cell consists of total power dissipated in the cell and thermal conductivity computed from the components inside the cell such as polysilicon used for transistor gates, tungsten used for

vias, copper used for TSVs, and dielectric material. With a sufficiently fine thermal cell size, equivalent thermal conductivity can be computed based on thermal resistive model [56].



**Figure 60:** GDSII layout-level thermal analysis flow.

### 6.3.3 Mechanical Stress Analysis

The mechanical stress of a layout is analyzed using the stress analyzer obtained from [49]. Inputs to the analyzer are die size, TSV diameter, TSV locations, simulation grid density, and precomputed data of TSV stress tensor. The analyzer outputs a von Mises stress map, which is a widely used mechanical reliability metric. Computation of stress at a point affected by multiple TSVs is based on the principle of linear superposition of stress tensors. With stress tensors obtained from finite element analysis (FEA) using an FEA tool ABAQUS, a full-chip stress analysis can be performed.

## 6.4 Experimental Results

For the experiments, 45-nm technology [62] is used. An open-source hardware IP core [63] is synthesized using an open cell library [64]. The thickness of bottom die and top die is  $30\ \mu\text{m}$  and  $530\ \mu\text{m}$ , respectively. High-thermal-conductivity molding compound [65] is assumed.

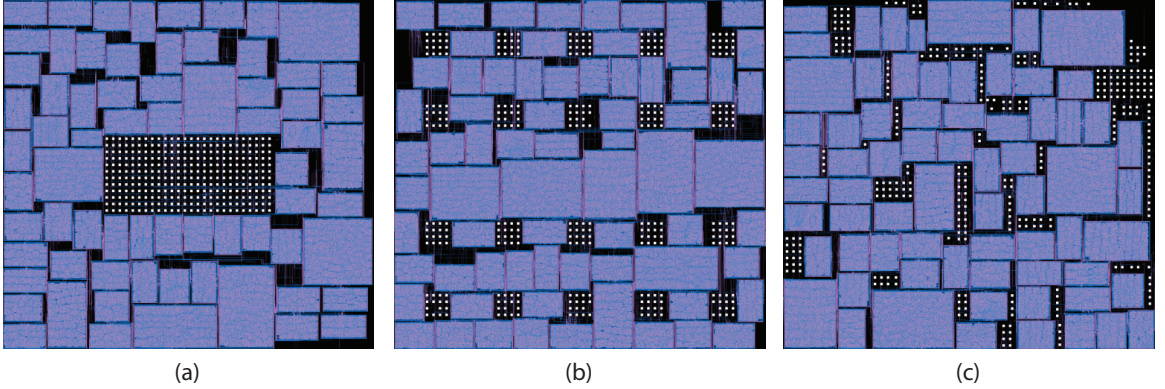
### 6.4.1 Baseline Designs

Layouts of the circuit in the three different styles are first designed and compared. The characteristics of the test circuit and its baseline designs are listed in Table 24. The number of TSVs depends on the partitioning of blocks and area ratio of the two dies. In this work,

the same partitioning, thus same number of TSVs, is used in all the three styles for fair comparison. The TSV size is  $10\ \mu\text{m}$ , and TSV pitch is  $30\ \mu\text{m}$ . The parasitic capacitance and resistance are  $50\ \text{fF}$  and  $50\ \text{m}\Omega$ , respectively. RDL wire width and spacing of  $0.4\ \mu\text{m}$  is used in the experiments. The layout of bottom die of the circuit in TSV-farm, TSV-distributed, and TSV-whitespace styles is shown in Figure 61. The RDL routing of the circuit in TSV-distributed and TSV-whitespace styles is shown in Figure 62.

**Table 24:** Characteristics of the test circuit (reconfigurable computing array) and baseline design.

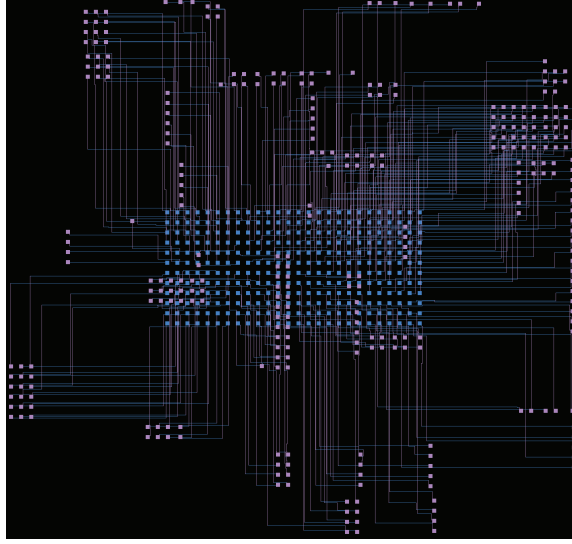
Total #gates	1,363,536	Total #blocks	95
#Interblock nets	1,853	#Blocks on top die	26
#TSVs	312	#Blocks on bottom die	69



**Figure 61:** Layout of bottom die of the circuit in (a) TSV-farm, (b) TSV-distributed, and (c) TSV-whitespace styles. TSVs are in white.

#### 6.4.1.1 Area, Footprint, and Wirelength

The area, footprint, and block-to-block (B2B) and RDL wirelength of layout in the three styles are shown in Table 25. The same area and footprint are used for all the three styles. Design in TSV-farm style has the shortest wirelength because all the TSVs occupy only one area in the middle of the bottom die, confining the obstruction of optimal block placement in small area. Design in TSV-distributed style has the longest block-to-block wirelength (27% longer than TSV-farm) because the TSV arrays distributed all over the bottom die obstruct optimal block placement. Design in TSV-whitespace style has little longer wirelength (2%) than design in TSV-farm style because blocks are moved from optimal block placement



**Figure 62:** RDL routing in TSV-whitespace style.

only when it is necessary to insert TSVs in some position without whitespace nearby. Most TSVs are inserted in the original whitespace, and do not interfere with placement of the blocks much. In addition, the design in TSV-distributed and TSV-whitespace styles require RDL routing as shown in Figure 62.

**Table 25:** Comparison of area, footprint, and wirelength in different layouts. TSV-f, TSV-d, and TSV-w are TSV-farm, TSV-distributed, and TSV-whitespace, respectively. The numbers in parenthesis after design style are TSV size and pitch in  $\mu\text{m}$ .

Design style	Area ( $\text{mm}^2$ )		Footprint ( $\text{mm}^2$ )		Wirelength (m)	
					B2B	RDL
TSV-f (10,30)	3.979	(100.00 %)	2.766	(100.00 %)	1.447	(100.00 %)
TSV-d (10,30)	3.979	(+0.00 %)	2.766	(+0.00 %)	1.842	(+27.30 %)
TSV-w (10,30)	3.979	(+0.00 %)	2.766	(+0.00 %)	1.483	(+2.46 %)
TSV-f (5,30)	3.979	(+0.00 %)	2.766	(+0.00 %)	1.450	(+0.23 %)
TSV-d (5,30)	3.979	(+0.00 %)	2.766	(+0.00 %)	1.814	(+25.32 %)
TSV-w (5,30)	3.979	(+0.00 %)	2.766	(+0.00 %)	1.475	(+1.91 %)
TSV-f (5,15)	3.699	(-7.03 %)	2.487	(-10.11 %)	1.482	(+2.39 %)
TSV-d (5,15)	3.699	(-7.03 %)	2.487	(-10.11 %)	1.741	(+20.31 %)
TSV-w (5,15)	3.699	(-7.03 %)	2.487	(-10.11 %)	1.471	(+1.62 %)

#### 6.4.1.2 Longest Path Delay and Buffers

The longest path delay (LPD) with and without timing optimization are also shown in Table 26. The timing optimization proposed in [66] is used with the target delay of 1.25 ns.

Without timing optimization, none of the designs meets the target delay; however, the design in TSV-farm style has the shortest delay. With timing optimization, all designs are closer to meet the target delay, and the delay of the design in TSV-farm style is still the shortest. The delay of the design in TSV-distributed and TSV-whitespace styles is longer than the delay of the design in TSV-farm style 10 % and 15 %, respectively. Because of long wirelength, it is hard to optimize the design in both TSV-distributed and TSV-whitespace styles to meet timing. In addition, no buffer can be added along the RDL routing because the routing is on the backside of the bottom die. The number of buffers inserted during timing optimization is also shown in Table 26. The design in TSV-farm style uses the smallest number of buffers. Because of long wirelength, the design in TSV-distributed and TSV-whitespace styles uses 27 % and 10 % more buffers than the design in TSV-farm style.

**Table 26:** Comparison of longest path delay (LPD), with and without optimization, and number of buffers in different layouts. TSV-f, TSV-d, and TSV-w are TSV-farm, TSV-distributed, and TSV-whitespace, respectively. The numbers in parenthesis after design style are TSV size and pitch in  $\mu\text{m}$ .

Design style	LPD (ns)			# Buffers	
	w/o	w/ opt.			
TSV-f (10,30)	3.136	1.293	(100.00 %)	3,459	(100.00 %)
TSV-d (10,30)	4.252	1.425	(+10.20 %)	4,386	(+26.80 %)
TSV-w (10,30)	4.568	1.492	(+15.38 %)	3,798	(+9.80 %)
TSV-f (5,30)	2.920	1.269	(-1.87 %)	3,462	(+0.09 %)
TSV-d (5,30)	4.045	1.411	(+9.16 %)	4,288	(+23.97 %)
TSV-w (5,30)	4.636	1.420	(+9.79 %)	3,709	(+7.23 %)
TSV-f (5,15)	2.792	1.256	(-2.86 %)	3,545	(+2.49 %)
TSV-d (5,15)	3.762	1.404	(+8.61 %)	4,081	(+17.98 %)
TSV-w (5,15)	3.872	1.422	(+10.01 %)	3,887	(+12.37 %)

#### 6.4.1.3 Power and Temperature

Power analysis is performed. The total power at maximum speed of each design is shown in Table 27. The design in TSV-distributed and TSV-whitespace styles consumes 6 % and 10 % less power than the design in TSV-farm style not because they are efficient design, but because they can only operate at slower speed than the design in TSV-farm style as discussed earlier.

Thermal analysis is performed at the maximum speed of each design. The maximum,

**Table 27:** Comparison of power and temperature of different layouts. TSV-f, TSV-d, TSV-w, and TSV-p are TSV-farm, TSV-distributed, and TSV-whitespace, respectively. The numbers in parenthesis after design style are TSV size and pitch in  $\mu\text{m}$ .

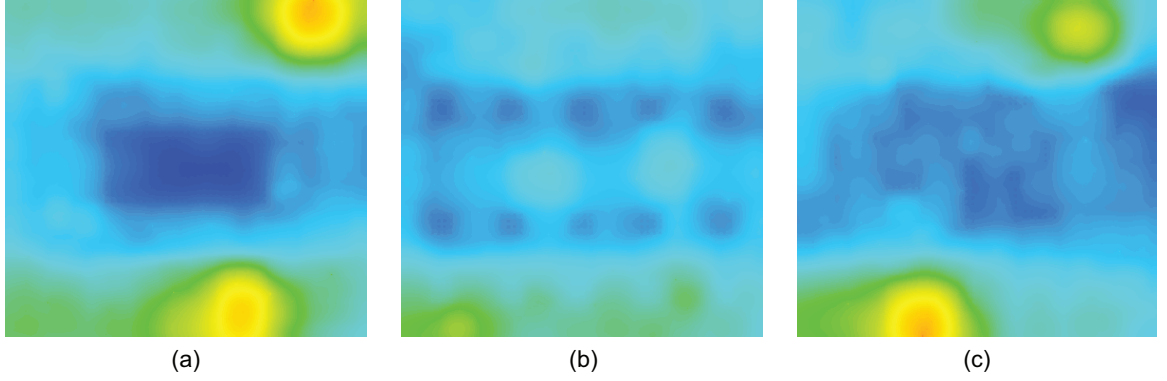
Design style	$P_{\text{total}}$ (mW)	$T_{\text{max}}$ ( $^{\circ}\text{C}$ )	$T_{\text{min}}$ ( $^{\circ}\text{C}$ )	$T_{\text{ave}}$ ( $^{\circ}\text{C}$ )
TSV-f (10,30)	1,183 (100.00 %)	76.87	38.04	47.56
TSV-d (10,30)	1,107 (-6.40 %)	62.43	39.15	46.28
TSV-w (10,30)	1,065 (-9.99 %)	77.04	38.65	46.19
TSV-f (5,30)	1,199 (+1.41 %)	77.55	38.29	47.88
TSV-d (5,30)	1,114 (-5.86 %)	62.53	39.22	46.45
TSV-w (5,30)	1,104 (-6.67 %)	78.71	39.20	46.94
TSV-f (5,15)	1,210 (+2.28 %)	74.85	41.33	48.47
TSV-d (5,15)	1,117 (-5.60 %)	59.17	39.81	47.46
TSV-w (5,15)	1,103 (-6.75 %)	79.26	39.76	48.10

minimum, and average temperatures are shown in Table 27. Although the minimum and average temperature across all the three designs are about the same, the maximum temperature of the three designs is different. The design in TSV-distributed style has the lowest maximum temperature not actually because it consumes low power, resulted from relatively low speed, but primarily because TSVs distributed all over the bottom die help conduct heat to heatsink. Design in TSV-farm style has high maximum temperature because TSVs in the center of the bottom die cannot help conduct heat from high-power blocks far from them. The design in TSV-whitespace styles also has high maximum temperature although it consumes the least power because of the same reason.

The thermal profile of bottom dies of the circuit in TSV-farm, TSV-distributed, and TSV-whitespace styles at the maximum speed of each design is shown in Figure 63. TSVs help reduce temperature. Local cool spots on bottom dies correspond to TSV arrays. Among 3D designs, design in TSV-distributed style has the lowest maximum temperature because TSVs are distributed across the bottom die. Design in TSV-whitespace style has the highest temperature because high-power blocks can be far from TSVs.

#### 6.4.1.4 Mechanical Stress

Mechanical stress analysis is performed. The maximum and average stresses are shown in Table 28. The area with stress higher than 10 MPa is also shown in the table. Despite high

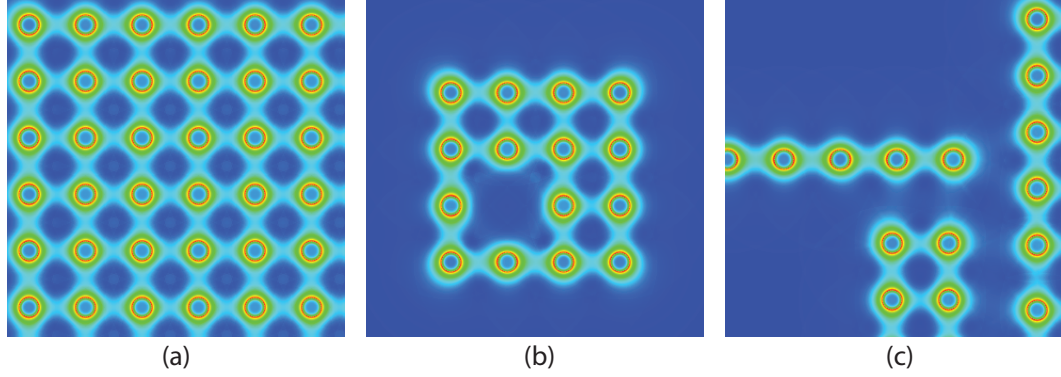


**Figure 63:** Temperature of bottom die of the circuit in (a) TSV-farm, (b) TSV-distributed, and (c) TSV-whitespace styles.

TSV density, TSV-farm style has the lowest maximum stress among the designs. TSVs in this style are in complete rows and columns. Because stress interference on a TSV coming from horizontal and vertical directions has opposite impacts, their effect partially cancel each other [67]. The average stresses above the 10-MPa threshold on the bottom die shows the opposite trend. The design in TSV-farm style has the highest average stress. When TSVs are close to each other (relative to TSV size), the impact of interference from neighboring TSVs becomes noticeable. The trend of the area above the threshold is completely opposite to the trend of average stress. The design in TSV-whitespace has the widest area of stress above the threshold, and the design in TSV-farm style has the narrowest area of stress above the threshold. When TSVs are grouped in small area, stress increases because of interference from neighboring TSVs, but the impacted area decreases. The stress profile of the design in three different styles is shown in Figure 64.

**Table 28:** Comparison of stress of different layouts.

Design style	$\sigma_{\max}$ (MPa)	$\sigma_{\text{ave}, \sigma > 10}$ (MPa)	$\text{Area}_{\sigma > 10}$ (mm <sup>2</sup> )
TSV-f (10,30)	676.78 (100.0 %)	150.20 (100.0 %)	0.353 (100.00 %)
TSV-d (10,30)	691.29 (+2.1 %)	97.73 (-34.9 %)	0.598 (+69.7 %)
TSV-w (10,30)	688.99 (+1.8 %)	88.95 (-40.8 %)	0.695 (+97.2 %)
TSV-f (5,30)	629.97 (-6.9 %)	72.08 (-52.0 %)	0.276 (-21.8 %)
TSV-d (5,30)	629.97 (-6.9 %)	67.89 (-54.8 %)	0.294 (-16.6 %)
TSV-w (5,30)	629.97 (-6.9 %)	64.65 (-57.0 %)	0.310 (-12.0 %)
TSV-f (5,15)	643.76 (-4.9 %)	150.94 (+0.5 %)	0.087 (-75.3 %)
TSV-d (5,15)	654.71 (-3.3 %)	101.12 (-32.7 %)	0.143 (-59.4 %)
TSV-w (5,15)	657.18 (-2.9 %)	85.21 (-43.3 %)	0.183 (-48.2 %)



**Figure 64:** Stress of bottom die of the circuit with 10- $\mu\text{m}$  TSVs in (a) TSV-farm, (b) TSV-distributed, and (c) TSV-whitespace styles.

#### 6.4.2 Impact of TSV Size

In this experiment, the size of TSVs is reduced to 5  $\mu\text{m}$ . The parasitic capacitance and resistance are updated to 25 fF and 200 m $\Omega$ , respectively. Die area and footprint are kept the same as shown in Table 25. The block-to-block wirelength and RDL wirelength are not much different from the baseline designs as shown in the table. The decrease in TSV capacitance improves timing slightly. As shown in Table 26, the longest path delay (LPD) with and without timing optimization decrease a little from the baseline designs. The design in TSV-farm style still has the shortest delay both with and without timing optimization. The longest path delay of the designs with 5- $\mu\text{m}$  TSV is 2 % to 6 % shorter than the longest path delay of the designs with 10- $\mu\text{m}$  TSV. As also shown in Table 26, the number of buffers inserted during timing optimization decreases a little because of the decrease in TSV capacitance.

As shown in Table 27, the total power at maximum speed of each design increases a little from the baseline designs because of the decrease in longest path delay. The maximum, minimum, and average temperatures at the maximum speed of each design are shown in Table 27. The temperature increases a little because of the decrease of TSV size, thus decrease in thermal conductivity to heatsink, and the increase in power, resulted from the decrease in longest path delay. The design in TSV-distributed style still has the lowest maximum temperature.

The maximum and average stresses are different from the baseline designs as shown in



Table 28. With decrease in TSV size, the stress decreases. The region affected by stress from each TSV shrinks, and stops overlapping each other, decreasing stress further. Because the TSV size, compared to TSV spacing, is now small, stress from TSVs are almost isolated from each other, resulting in the same maximum stress across all design styles. The average stresses above 10-MPa threshold is noticeably lower than that of the baseline designs, and are not much different across all the designs. Because of the decrease in TSV size, the area with stress higher than the threshold on the bottom die decreases dramatically from the baseline designs.

### 6.4.3 Impact of TSV Pitch

After reducing TSV size, the mechanical stress decreases, leading to an opportunity to place TSVs closer to each other. In this experiment, the pitch of TSVs is changed to  $15\mu\text{m}$ . The area and footprint decrease by 7 % and 10 %, respectively. As a result, wirelength decreases a little. However, the RDL wirelength for the design in TSV-distributed style increases because the bonding pad on the top die are moved closer to the center of the die.

The decrease in wirelength also affects timing noticeably. As shown in Table 26, the longest path delay (LPD) with and without timing optimization decrease further. The design in TSV-farm style still has the shortest delay both with and without timing optimization. It almost meets the target delay now after optimization.

As shown in Table 27, the total power at maximum speed of each design increases because of the decrease in longest path delay. The maximum, minimum, and average temperatures at the maximum speed of each design are shown in Table 27. The minimum and average temperature increase because of the decrease in footprint area, thus increase in power density, and the increase in power, resulted from the increase in speed. The design in TSV-distributed style still has the lowest maximum temperature.

The maximum and average stresses increase as shown in Table 28. Because of decreasing TSV pitch, stress from each TSV starts overlapping each other again. The maximum stress of each design styles is still lower than that of the baseline designs with  $10\text{-}\mu\text{m}$  TSV. The average stress above 10-MPa threshold on the bottom die increase to the same level of the

baseline designs because of the same relative size of TSV to the pitch. The area with stress above the threshold on the bottom die decreases dramatically from the designs with 30- $\mu\text{m}$  TSV pitch because TSVs are placed closer to each other.

## **6.5 Summary**

In this chapter, 3D-IC designs with different die sizes in the IC stack are investigated. Layouts of a circuit are designed at block level in various styles, and the trade-offs of the design styles on the area, footprint, timing, power, temperature, and mechanical stress are studied. Because of the lack of RDL wiring, TSV-farm style has the best timing. The design in this style has the highest average stress, but the area impacted by stress is smallest. TSV-distributed style has the worst wirelength because TSV arrays interfere with block placement. However, it has the lowest temperature because TSVs distributed across the die help reduce temperature.

## CHAPTER VII

### CONCLUSIONS

Three-dimensional integrated circuits (3D ICs) are expected to deliver promising benefits, such as performance improvement and power reduction. Stacking thinned dies and connecting them by TSVs are a widely used method for 3D-IC fabrication; however, the position of TSVs impacts the performance and reliability of 3D ICs tremendously. The thesis of this dissertation is that the major performance and reliability problems found in 3D ICs that use TSVs can be addressed at placement stage. To support this thesis, the following algorithms and studies are presented in this dissertation:

- A wirelength-driven placement algorithm for gate-level design of 3D ICs
- A study of the physical impact of TSVs on the layout of 3D ICs
- A study of the impact of mechanical stress on the timing of 3D ICs
- A TSV-stress-driven placement algorithm for gate-level design of 3D ICs
- Two temperature-aware placement algorithms for gate-level design of 3D ICs
- A study of the impact of TSVs on the temperature of 3D ICs
- A study of design quality trade-offs of block-level placements of die-to-wafer bonded 3D ICs

The simulation results from the studies provide the insight into how the position of TSVs impacts the performance and reliability of 3D ICs. They help outline the direction in which the placement algorithms for 3D ICs are developed to achieve high performance and reliability.

Despite a number of contributions, the limitation of researches presented in this dissertation should be noted. For the wirelength-driven placement algorithm, die partitioning is

assumed to be given. Partitioning gates into dies differently can result in different number of TSVs, and thus affect the layout of 3D ICs. Also, the placement algorithm uses only one TSV to connect each 3D net between adjacent dies, which may result in suboptimal placements. The stress-driven placement algorithm considers only TSV stress. The impact of STI stress on the timing should also be included in the placement algorithm. Unlike TSV stress, STI stress changes with gate position. Therefore, the placement algorithm may consider STI stress differently from TSV stress. For the temperature-aware placement algorithm, the gate power is assumed to be static. During operation, gate power can change depending on work load, resulting in temperature change. The placement algorithm should consider such change so that the impact of gate power change on temperature is minimal. Finally, besides traditional metrics, the design quality trade-off study is limited to only well-studied reliability metrics such as temperature and stress.

Future researches pertaining to the researches presented in this dissertation can be pursued in the following directions. The wirelength-driven placement algorithm may consider both die partitioning and position of gates and TSVs simultaneously to obtain optimal wirelength. The impact of stress caused by molding material as well as package on timing of 3D ICs may be studied, and the stress-driven placement algorithm may be extended to consider the impact. The impact of other thermal structures besides TSV, for example microfluidic channel, may be studied, and the temperature-aware placement algorithm may consider such structures in addition to TSV. Finally, the block-level placement algorithm for the die-to-wafer bonded 3D ICs may be developed based on the results from the design quality trade-off study.

## REFERENCES

- [1] KIM, D. H., ATHIKULWONGSE, K., and LIM, S. K., “A study of through-silicon-via impact on the 3D stacked IC layout,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 674–680, Nov. 2–5 2009.
- [2] GOPLEN, B. and SAPATNEKAR, S., “Efficient thermal placement of standard cells in 3D ICs using a force directed approach,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 86–89, Nov. 9–13 2003.
- [3] OBERMEIER, B. and JOHANNES, F. M., “Temperature-aware global placement,” in *Proc. Asia and South Pacific Design Automation Conf.*, (Yokohama, Japan), pp. 143–148, Jan. 27–30 2004.
- [4] GOPLEN, B. and SAPATNEKAR, S., “Placement of 3D ICs with thermal and inter-layer via considerations,” in *Proc. ACM Design Automation Conf.*, (San Diego, CA), pp. 626–631, June 4–8 2007.
- [5] CONG, J., LUO, G., and SHI, Y., “Thermal-aware cell and through-silicon-via co-placement for 3D ICs,” in *Proc. ACM Design Automation Conf.*, (San Diego, CA), pp. 670–675, June 5–9 2011.
- [6] CHAN, T., CONG, J., and SZE, K., “Multilevel generalized force-directed method for circuit placement,” in *Proc. Int. Symp. Physical Design*, (San Francisco, CA), pp. 185–192, Apr. 2–5 2005.
- [7] CHEN, T.-C., JIANG, Z.-W., HSU, T.-C., CHEN, H.-C., and CHANG, Y.-W., “A high-quality mixed-size analytical placer considering preplaced blocks and density constraints,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 187–192, Nov. 5–9 2006.
- [8] EISENMANN, H. and JOHANNES, F. M., “Generic global placement and floorplanning,” in *Proc. ACM Design Automation Conf.*, (San Francisco, CA), pp. 269–274, June 15–19 1998.
- [9] SPINDLER, P., SCHLICHTMANN, U., and JOHANNES, F. M., “Kraftwerk2—a fast force-directed quadratic placement approach using an accurate net model,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, pp. 1398–1411, Aug. 2008.
- [10] JOYNER, J. W., ZARKESH-HA, P., DAVIS, J. A., and MEINDL, J. D., “A three-dimensional stochastic wire-length distribution for variable separation of strata,” in *Proc. IEEE Int. Interconnect Technology Conference*, (Burlingame, CA), pp. 126–128, June 5–7 2000.
- [11] BEYNE, E. *et al.*, “Through-silicon via and die stacking technologies for microsystems-integration,” in *IEEE Int. Electron Devices Meeting Tech. Dig.*, (San Francisco, CA), pp. 495–498, Dec. 15–17 2008.

- [12] KIM, D. H., MUKHOPADHYAY, S., and LIM, S. K., "Through-silicon-via aware interconnect prediction and optimization for 3D stacked ICs," in *Proc. ACM/IEEE Int. Workshop on System Level Interconnect Prediction*, (San Francisco, CA), pp. 85–92, July 26–27 2009.
- [13] CONG, J., LUO, G., WEI, J., and ZHANG, Y., "Thermal-aware 3D IC placement via transformation," in *Proc. Asia and South Pacific Design Automation Conf.*, (Yokohama, Japan), pp. 780–785, Jan. 23–26 2007.
- [14] CAUGHEY, D. M. and THOMAS, R. E., "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, pp. 2192–2193, Dec. 1967.
- [15] ARORA, N. D., HAUSER, J. R., and ROULSTON, D. J., "Electron and hole mobilities in silicon as a function of concentration and temperature," *IEEE Trans. on Electron Devices*, vol. 29, pp. 292–295, Feb. 1982.
- [16] SMITH, C. S., "Piezoresistance effect in germanium and silicon," *Physical Review*, vol. 94, pp. 42–49, Apr. 1954.
- [17] THOMPSON, S. E. *et al.*, "A 90-nm logic technology featuring strained-silicon," *IEEE Trans. on Electron Devices*, vol. 51, pp. 1790–1797, Nov. 2004.
- [18] LU, K. H. *et al.*, "Thermo-mechanical reliability of 3-D ICs containing through silicon vias," in *IEEE Electronic Components and Technology Conf.*, (San Diego, CA), pp. 630–634, May 26–29 2009.
- [19] KAHNG, A. B., SHARMA, P., and TOPALOGLU, R. O., "Chip optimization through STI-stress-aware placement perturbations and fill insertion," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, pp. 1241–1252, July 2008.
- [20] CHAKRABORTY, A., SHI, S. X., and PAN, D. Z., "Stress aware layout optimization leveraging active area dependent mobility enhancement," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, pp. 1533–1545, Oct. 2010.
- [21] BLACK, J. R., "Electromigration—a brief survey and some recent results," *IEEE Trans. on Electron Devices*, vol. 16, pp. 338–347, Apr. 1969.
- [22] BLAT, C. E., NICOLLIAN, E. H., and POINDEXTER, E. H., "Mechanism of negative-bias-temperature instability," *J. Applied Physics*, vol. 69, pp. 1712–1720, Feb. 1991.
- [23] LEDUC, P. *et al.*, "Challenges for 3D IC integration: Bonding quality and thermal management," in *Proc. IEEE Int. Interconnect Technology Conference*, (Burlingame, CA), pp. 210–212, June 4–6 2007.
- [24] SHAHIDI, G., "SOI technology for the GHz era," *IBM J. Res. Develop.*, vol. 46, pp. 121–131, Mar. 2002.
- [25] SU, L. T., CHUNG, J. E., ANTONIADIS, D. A., GOODSON, K. E., and FLIK, M. I., "Measurement and modeling of self-heating in SOI NMOSFET's," *IEEE Trans. on Electron Devices*, vol. 41, pp. 69–75, Jan. 1994.
- [26] PATHAK, M., LEE, Y.-J., MOON, T., and LIM, S. K., "Through-silicon-via management during 3D physical design: When to add and how many?," in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 387–394, Nov. 7–11 2010.

- [27] HSU, M.-K., CHANG, Y.-W., and BALABANOV, V., "TSV-aware analytical placement for 3D IC designs," in *Proc. ACM Design Automation Conf.*, (San Diego, CA), pp. 664–669, June 5–10 2011.
- [28] KNECHTEL, J., MARKOV, I. L., and LIENIG, J., "Assembling 2-D blocks into 3-D chips," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, pp. 228–241, Feb. 2012.
- [29] KIM, D. H., TOPALOGLU, R. O., and LIM, S. K., "Block-level 3D IC design with through-silicon-via planning," in *Proc. Asia and South Pacific Design Automation Conf.*, (Sydney, Australia), pp. 335–340, Jan. 30–Feb. 2 2012.
- [30] TSAI, M.-C., WANG, T.-C., and HWANG, T., "Through-silicon via planning in 3-D floorplanning," *IEEE Trans. on Very Large Scale Systems*, vol. 19, pp. 1448–1457, Aug. 2011.
- [31] CADENCE, "SoC Encounter RTL-to-GDSII system." [http://www.cadence.com/products/di/soc\\_encounter/pages/default.aspx](http://www.cadence.com/products/di/soc_encounter/pages/default.aspx), May 2009.
- [32] IWLS, "IWLS 2005 benchmarks." <http://www.iwls.org/iwls2005/benchmarks.html>, June 2005.
- [33] YAN, H., LI, Z., ZHOU, Q., and HONG, X., "Via assignment algorithm for hierarchical 3-D placement," in *Proc. IEEE Int. Conf. on Communications, Circuits and Systems*, (Hong Kong, China), pp. 1225–1229, May 27–30 2005.
- [34] DAO, T., TRIYOSO, D. H., PETRAS, M., and CANONICO, M., "Through silicon via stress characterization," in *Proc. IEEE Int. Conf. on Integrated Circuit Design and Technology*, (Austin, TX), pp. 39–41, May 18–20 2009.
- [35] SELVANAYAGAM, C. S. *et al.*, "Nonlinear thermal stress/strain analyses of copper filled TSV (through silicon via) and their flip-chip microbumps," in *IEEE Electronic Components and Technology Conf.*, (Lake Buena Vista, FL), pp. 1073–1081, May 27–30 2008.
- [36] YANG, J.-S., ATHIKULWONGSE, K., LEE, Y.-J., LIM, S. K., and PAN, D. Z., "TSV stress aware timing analysis with applications to 3D-IC layout optimization," in *Proc. ACM Design Automation Conf.*, (Anaheim, CA), pp. 803–806, June 13–18 2010.
- [37] KAHNG, A. B., SHARMA, P., and TOPALOGLU, R. O., "Exploiting STI stress for performance," in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 83–90, Nov. 5–8 2007.
- [38] OKORO, C. *et al.*, "Analysis of the induced stresses in silicon during thermocompression Cu-Cu bonding of Cu-through-vias in 3D-SIC architecture," in *IEEE Electronic Components and Technology Conf.*, (Reno, NV), pp. 249–255, May 29–June 1 2007.
- [39] CHAKRABORTY, A., SHI, S. X., and PAN, D. Z., "Layout level timing optimization by leveraging active area dependent mobility of strained-silicon devices," in *Proc. Design, Automation and Test in Europe*, (Munich, Germany), pp. 849–855, Mar. 10–14 2008.

- [40] HOPCROFT, M. A., NIX, W. D., and KENNY, T. W., “What is the Young’s modulus of silicon?,” *Journal of Microelectromechanical Systems*, vol. 19, pp. 229–238, Apr. 2010.
- [41] SERIN, N., SERIN, T., Ş. HORZUM, and ÇELİK, Y., “Annealing effects on the properties of copper oxide thin films prepared by chemical deposition,” *Semiconductor Science and Technology*, vol. 20, pp. 398–401, May 2005.
- [42] THOMPSON, S. E., SUN, G., CHOI, Y. S., and NISHIDA, T., “Uniaxial-process-induced strained-Si: Extending the CMOS roadmap,” *IEEE Trans. on Electron Devices*, vol. 53, pp. 1010–1020, May 2006.
- [43] MOROZ, V., SMITH, L., LIN, X.-W., PRAMANIK, D., and ROLLINS, G., “Stress-aware design methodology,” in *Proc. Int. Symp. on Quality Electronic Design*, (San Jose, CA), pp. 807–812, Mar. 27–29 2006.
- [44] MIYAMOTO, M. and OTHER, “Impact of reducing STI-induced stress on layout dependence of MOSFET characteristics,” *IEEE Trans. on Electron Devices*, vol. 51, pp. 440–443, Mar. 2004.
- [45] IRIE, H., KITA, K., KYUNO, K., and TORIUMI, A., “In-plane mobility anisotropy and universality under uni-axial strains in n- and p-MOS inversion layers on (100), (110), and (111) Si,” in *IEEE Int. Electron Devices Meeting Tech. Dig.*, (San Francisco, CA), pp. 225–228, Dec. 13–15 2004.
- [46] SUTHRAM, S., ZIEGERT, J. C., NISHIDA, T., and THOMPSON, S. E., “Piezoresistance coefficients of (100) silicon nMOSFETs measured at low and high channel stress,” *IEEE Electron Device Letters*, vol. 28, pp. 58–60, Jan. 2007.
- [47] LUNDSTROM, M. S., “On the mobility versus drain current relation for a nanoscale MOSFET,” *IEEE Electron Device Letters*, vol. 22, pp. 293–295, June 2001.
- [48] UCHIDA, K., KRISHNAMOHAN, T., SARASWAT, K. C., and NISHI, Y., “Physical mechanisms of electron mobility enhancement in uniaxial stressed MOSFETs and impact of uniaxial stress engineering in ballistic regime,” in *IEEE Int. Electron Devices Meeting Tech. Dig.*, (Washington, DC), pp. 129–132, Dec. 5–7 2005.
- [49] JUNG, M., MITRA, J., PAN, D., and LIM, S. K., “TSV stress-aware full-chip mechanical reliability analysis and optimization for 3D IC,” in *Proc. ACM Design Automation Conf.*, (San Diego, CA), pp. 188–193, June 5–10 2011.
- [50] ANSYS, “Structural mechanics solutions.” <http://www.ansys.com/Products/Simulation+Technology/Structural+Mechanics>, July 2011.
- [51] SYNOPSYS, “PrimeTime.” <http://www.synopsys.com/Tools/Implementation/SignOff/Pages/PrimeTime.aspx>, Nov. 2009.
- [52] ZHAO, W. and CAO, Y., “New generation of predictive technology model for sub-45 nm early design exploration,” *IEEE Trans. on Electron Devices*, vol. 53, pp. 2816–2823, Nov. 2006.



- [53] KAHNG, A. B., SHARMA, P., and ZELIKOVSKY, A., “Fill for shallow trench isolation CMP,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 661–668, Nov. 5–9 2006.
- [54] TIAN, R., TANG, X., and WONG, M. D. F., “Dummy-feature placement for chemical-mechanical polishing uniformity in a shallow-trench isolation process,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, pp. 63–71, Jan. 2002.
- [55] ANSYS, “ANSYS FLUENT flow modeling simulation software.” <http://www.ansys.com/Products/Simulation+Technology/Fluid+Dynamics/ANSYS+FLUENT>, Nov. 2010.
- [56] XU, C. *et al.*, “Fast 3-D thermal analysis of complex interconnect structures using electrical modeling and simulation methodologies,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 658–665, Nov. 2–5 2009.
- [57] VAN DER PLAS, G. *et al.*, “Design issues and considerations for low-cost 3D TSV IC technology,” in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Paper*, (San Francisco, CA), pp. 148–149, Feb. 7–11 2010.
- [58] KIM, J.-S. *et al.*, “A 1.2 V 12.8 GB/s 2 Gb mobile wide-I/O DRAM with  $4 \times 128$  I/Os using TSV based stacking,” *J. Solid-State Circuits*, vol. 47, pp. 107–116, Jan. 2012.
- [59] THOROLFSSON, T., GONSALVES, K., and FRANZON, P. D., “Design automation for a 3DIC FFT processor for synthetic aperture radar: A case study,” in *Proc. ACM Design Automation Conf.*, (San Francisco, CA), pp. 51–56, July 26–31 2009.
- [60] WEERASEKERA, R., PAMUNUWA, D., ZHENG, L.-R., and TENHUNEN, H., “Two-dimensional and three-dimensional integration of heterogeneous electronic systems under cost, performance, and technological constraints,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 28, pp. 1237–1250, Aug. 2009.
- [61] ZHANG, T. *et al.*, “A customized design of DRAM controller for on-chip 3D DRAM stacking,” in *Proc. IEEE Custom Integrated Circuits Conf.*, (San Jose, CA), Sept. 19–22 2010.
- [62] NORTH CAROLINA STATE UNIVERSITY, “FreePDK45.” <http://www.eda.ncsu.edu/wiki/FreePDK>, Mar. 2009.
- [63] OPENCORES.ORG, “OpenCores.” <http://opencores.org>, Dec. 2009.
- [64] NANGATE INC., “NanGate 45nm Open Cell Library.” <http://www.nangate.com>, July 2009.
- [65] HU, X. *et al.*, “High thermal conductivity molding compound for flip-chip packages.” US Patent 2009/0004317 A1, Jan. 1 2009.
- [66] LEE, Y.-J. and LIM, S. K., “Timing analysis and optimization for 3D stacked multi-core microprocessors,” in *Proc. IEEE Int. 3D Systems Integration Conf.*, (Munich, Germany), Nov. 16–18 2010.

- [67] JUNG, M., LIU, X., SITARAMAN, S. K., PAN, D. Z., and LIM, S. K., “Full-chip through-silicon-via interfacial crack analysis and optimization for 3D IC,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 563–570, Nov. 7–10 2011.

## PUBLICATIONS

This dissertation is based on and/or related to the works and results presented in the following publications in print:

- [1] Kim, D. H., **Athikulwongse, K.**, and Lim, S. K., “A study of through-silicon-via impact on the 3D stacked IC layout,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 674–680, Nov. 2–5 2009.
- [2] Yang, J.-S., **Athikulwongse, K.**, Lee, Y.-J., Lim, S. K., and Pan, D. Z., “TSV stress aware timing analysis with applications to 3D-IC layout optimization,” in *Proc. ACM Design Automation Conf.*, (Anaheim, CA), pp. 803–806, June 13–18 2010.
- [3] **Athikulwongse, K.**, Chakraborty, A., Yang, J.-S., Pan, D. Z., and Lim, S. K., “Stress-driven 3D-IC placement with TSV keep-out zone and regularity study,” in *Proc. IEEE Int. Conf. Computer-Aided Design*, (San Jose, CA), pp. 669–674, Nov. 7–11 2010.
- [4] Pan, D. Z., Lim, S. K., **Athikulwongse, K.**, Jung, M., Mitra, J., Pak, J., Pathak, M., and Yang, J.-S., “Design for manufacturability and reliability for TSV-based 3D ICs,” in *Proc. Asia and South Pacific Design Automation Conf.*, (Sydney, Australia), pp. 750–755, Jan. 30–Feb. 2 2012.
- [5] **Athikulwongse, K.**, Pathak, M., and Lim, S. K., “Exploiting die-to-die thermal coupling in 3D IC placement,” in *Proc. ACM Design Automation Conf.*, (San Francisco, CA), pp. 741–746, June 3–7 2012.

In addition, this dissertation is based on and/or related to the works and results presented in the following papers accepted/in submission:

- [6] Kim, D. H., **Athikulwongse, K.**, and Lim, S. K., “A study of through-silicon-via impact on the 3D stacked IC layout,” accepted by *IEEE Trans. on Very Large Scale Systems*.
- [7] **Athikulwongse, K.**, Pathak, M., and Lim, S. K., “Exploiting die-to-die thermal coupling in 3D IC placement,” submitted to *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*.
- [8] **Athikulwongse, K.**, Yang, J.-S., Pan, D. Z., and Lim, S. K., “Impact of mechanical stress on the full chip timing for TSV-based 3D ICs,” submitted to *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*.
- [9] **Athikulwongse, K.**, Kim, D. H., Jung, M., and Lim, S. K., “Block-level designs of die-to-wafer bonded 3D ICs and their design quality tradeoffs,” submitted to *Proc. IEEE Int. Conf. Computer-Aided Design*, 2012.

The author has also completed works unrelated to this dissertation presented in the following publications in print:

- [10] **Athikulwongse, K.**, Zhao, X., and Lim, S. K., “Buffered clock tree sizing for skew minimization under power and thermal budgets,” in *Proc. Asia and South Pacific Design Automation Conf.*, (Taipei, Taiwan), pp. 474–479, Jan. 18–21 2010.
- [11] Healy, M. B., **Athikulwongse, K.**, Goel, R., Hossain, M. M., Kim, D. H., Lee, Y.-J., Lewis, D. L., Lin, T.-W., Liu, C., Jung, M., Ouellette, B., Pathak, M., Sane, H., Shen, G., Woo, D. H., Zhao, X., Loh, G. H., Lee, H.-H. S., and Lim, S. K., “Design and analysis of 3D-MAPS: A many-core 3D processor with stacked memory,” in *Proc. IEEE Custom Integrated Circuits Conf.*, (San Jose, CA), Sep. 19–22 2010.
- [12] Bashir, M., Kim, D. H., **Athikulwongse, K.**, Lim, S. K., and Milor, L., “Backend low-k TDDDB chip reliability simulator,” in *Proc. IEEE Int. Reliability Physics Symp.*, (Monterey, CA), pp. 2C.2.1–2C.2.10, Apr. 10–14 2011.
- [13] Kim, D. H., **Athikulwongse, K.**, Healy, M., Hossain, M., Jung, M., Khorosh, I., Kumar, G., Lee, Y.-J., Lewis, D., Lin, T.-W., Liu, C., Panth, S., Pathak, M., Ren, M., Shen, G., Song, T., Woo, D. H., Zhao, X., Kim, J., Choi, H., Loh, G., Lee, H.-H., and Lim, S. K., “3D-MAPS: 3D massively parallel processor with stacked memory,” in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Paper*, (San Francisco, CA), pp. 188–190, Feb. 19–23 2012.

## VITA

Krit Athikulwongse received the B.Eng and M.Eng degrees from the Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand, in 1995 and 1997, respectively. He was awarded a scholarship from the Royal Thai Government to pursue a Ph.D. degree abroad in 2000. He received the M.S. degree from the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, in 2005. He joined the Georgia Tech Computer Aided Design Laboratory in 2008, and earned his Ph.D. degree in 2012. His research focus is on physical design, especially placement, for 3D integrated circuits. His other research interests include VLSI design, high-performance computer architectures, and embedded systems.