POWER DELIVERY AND THERMAL CONSIDERATIONS FOR 2.5-D AND 3-D INTEGRATION TECHNOLOGIES

A Dissertation
Presented to
The Academic Faculty

By

Md Obaidul Hossen

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 2019

Copyright 2019 © Md Obaidul Hossen

POWER DELIVERY AND THERMAL CONSIDERATIONS FOR 2.5-D AND 3-D INTEGRATION TECHNOLOGIES

Approved by:

Dr. Muhannad S. Bakir, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Azad Naeemi
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Arijit Raychowdhury
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Tushar Krishna
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Yogendra Joshi
George W. Woodruff School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: November 05, 2019

Everything will be okay in the end. If it’s not okay, it’s not the end.

John Lennon

To my family and my friends

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Muhannad S. Bakir. I am grateful to Dr. Bakir for always believing in me. His encouraging words helped me pass by the rough edges during my PhD journey. His constant drive for new ideas and active participation in research always amazed me. I would like to thank Dr. Bakir for trusting me with different collaborations at different times. Regardless of the nature of the results I presented to our collaborators, Dr. Bakir always had my back and took the blame in numerous occasions. I would like to thank him for making this ride as smooth as possible.

I would like to thank Dr. Azad Naeemi and Dr. Arijit Raychowdhury for serving as the thesis reading committee members. Also, I would like to thank Dr. Yogendra Joshi serving in my thesis committee. I also like to thank Dr. Tushar Krishna for his time to serve in my thesis committee and also for the collaborative work that I have been doing with his group.

I wish to thank all the current and previous members of the I3DS group. First of all, I would like express my gratitude to my academic mentor Dr. Yang Zhang for his relentless support. I know mentor-ship can be hard at times. I am grateful to him for being patient with me. We worked together for a couple of years and his thought process always amazed me. I also would like to thank Dr. Paul Jo and Joe Gonzalez for their collaborative work. We started our PhD around the same time and we have hung out more as friends rather than as colleagues. Joe's flexible interconnect optimization work enhanced the quality of my interposer-to-motherboard work. Also, I wish to thank Dr. Tom Sarvey and Dr. William Wahby for their constant support. I don't think I submitted any paper without one of these guys proof reading it at least once. I would like to thank Ankit Kaul for his support while taking over my research. I could not help him with all his queries at the right moment. However, he was always patient with me when he wanted to learn something.

I was very fortunate to have worked with a lot of industry partners. I would like to thank Hesam Fathi Moghadam, Michael Dayringer, and Dr. Yue Zhang for their valuable feed-

back on my power delivery work and my flexible interconnect optimization work. I would like to thank Arvind Dasu, Pant Mandira, Ren Jihong, Wendem Beyene, Ilya Ganusov, and Ravi Gutala from Intel Corporation for their feedback on my thermal-power delivery co-analysis research. I would like to thank Bharani Chava, Geert Van der Plas, and Eric Beyne for the collaboration on our backside power delivery research.

I would like to thank Kent Kalpakjian from Micron Technology Inc. for giving me the internship opportunity. It was a pleasant experience. I would like to thank my mentors Jake Anderson, Adam El-Mansouri, and Fuad Badrieh for teaching me a lot during that short period of time.

Lastly, I would like to thank my family and friends for their relentless support in this journey. My true inspiration are my parents, especially my mother who paved the way in so many ways for me to be where I am now. I would like to thank my wife, Rehnuma Afrin, for her support throughout my PhD life. Friends always played an important role in my life. I am blessed to have a lot of amazing friends who are always available when I need them. Graduate school can be pretty rough at times. However, my friends and family somehow managed to say the right words and do the right things during my moments of crisis.

TABLE OF CONTENTS

Acknowledgments v

List of Tables xiii

List of Figures xv

Summary xxii

Chapter 1: Introduction 1

1.1 Current Relevant Research 4

1.1.1 Heterogeneous Integration Technologies 4

1.1.2 Thermo-Mechanical Reliability Analysis for Advanced Packaging Technologies 8

1.1.3 Power Delivery Network Modeling for 2.5-D and 3-D ICs 11

1.1.4 Thermal-PDN Co-Anaysis Modeling 13

1.2 Organization of This Thesis 16

Chapter 2: Thermomechanical Analysis and Package Level Optimization of Mechanically Flexible Interconnects (MFIs) for Interposer-on-Motherboard Assembly 20

2.1 MFI Orientation and Package Level Optimization for Reduced Stress and Warpage 21

2.1.1 Simulation Specifications 21

2.1.2 Meshing Profile . . . 23
2.1.3 Thermally Induced Warpage and Stress Results . . . 24
2.1.4 Radially Oriented MFI Distribution . . . 25
2.1.5 System Level Optimization Methodology . . . 26
2.2 Impact of Interposer Size on Thermo-Mechanical Reliability . . . 29
2.3 MFI Orientation (Radial) Along the Thermal Expansion/ Contraction Contour 30
2.4 Impact of MFI Pitch on Warpage and Stress . . . 34
2.5 Conclusion . . . 35

Chapter 3: Power Delivery Network Modeling for Emerging Heterogeneous Integration technologies and Design Space Exploration of Power Delivery Including Voltage Regulator Modules . . . 37

3.1 Modeling Methodology . . . 38
3.1.1 Board-Level PDN . . . 38
3.1.2 Package-Level PDN . . . 38
3.1.3 On-Die PDN . . . 40
3.1.4 PDN Analysis Formulation . . . 42
3.2 Validation . . . 43
3.2.1 Steady-State Results . . . 44
3.2.2 Transient-State Results . . . 45
3.3 PDN Evaluation of Emerging Heterogeneous Integration Platforms . . . 45
3.3.1 2.5-D/3-D Integration Scenarios . . . 46
3.3.2 Design Parameters and Specifications . . . 47
3.3.3 Benchmarking . . . 49

3.4 Impact of PDN in the Bridge-Chip 53

3.4.1 PDN Schematics with Bridge-Chip PDN 54

3.4.2 Bridge-Chip PDN Analysis for 2.5-D of CPU-FPGA Integration . . 56

3.4.3 PDN Analysis for a 2.5-D Integration of Stacked Memory-FPGA Configuration 62

3.5 Design Space Exploration of Power Delivery Including Voltage Regulator Modules 65

3.5.1 Benchmark Architectures 66

3.5.2 PDN Topology and Specifications 67

3.5.3 DC IR-Drop Comparison of Different Benchmark Configurations . 68

3.5.4 Comparison of Transient Noise for different configurations 72

3.5.5 Thermal Implications of Different Architectures 76

3.5.6 Power Delivery Capabilities of Different Architectures 78

3.6 Conclusion 80

Chapter 4: Benchmarking power delivery networks for Fan-out Wafer Level Packaging (FOWLP) technologies 82

4.1 Modeling Framework 83

4.1.1 Simulation Configurations 83

4.1.2 PDN with Multiple Voltage Domain 84

4.1.3 Analysis Type 86

4.2 FOWLP Benchmarking 87

4.2.1 Specification 87

4.2.2 PDN Analysis Results 88

4.3 Design Space Exploration of Fan-out Wafer Level Technology 92

4.3.1 Impact of Solder Bump Distribution 92
4.3.2 Impact of RDL Density 93
4.3.3 Impact of Copper Pillar Pitch 94
4.3.4 Double-sided RDL in 3-D FOWLP Technology 95
4.3.5 Through Mold Via Distribution 97
4.3.6 Comparison Between 3-D FOWLP, FC POP, and 3-D IC with FOWLP 98
4.4 Conclusion 99

Chapter 5: Power Delivery Network (PDN) Modeling for Backside-PDN Configurations with Buried Power Rails and μTSVs 100

5.1 Modeling Framework and Specifications 100
5.1.1 Simulation Configurations 100
5.1.2 Specifications 102
5.1.3 Adaptive Meshing 104
5.2 Power Delivery Network Benchmarking 104
5.2.1 Uniform Power Density Maps 104
5.2.2 Hotspot Power Densities 105
5.2.3 Physical Design Results 107
5.3 Sensitivity Analysis 109
5.3.1 Chip-to-Package Interconnection 109
5.3.2 Impact of Input Pulse 110
5.3.3 MIM Decoupling Cap Density 111
5.3.4 Thermal Implications of a Backside PDN Configuration 112
5.4 Conclusion 113

Chapter 6: Thermal- Power Delivery Network (PDN) Co-Analysis of 2.5-D Integration Technologies . . . 115

6.1 Thermal Evaluation of Bridge-Chip Based 2.5-D Configurations . . . 116

6.1.1 Bridge-Chip Based 2.5-D Configuration . . . 116

6.1.2 Thermal Modeling Specifications . . . 116

6.1.3 Thermal Results . . . 117

6.2 Thermal PDN Co-Analysis for Bridge-Chip Based 2.5-D Configuration . . . 119

6.2.1 Steady-state IR-drop Modeling Framework . . . 119

6.2.2 Steady-State Thermal-PDN Co-Analysis Results . . . 121

6.2.3 Impact of Different Interaction Models and Number of Bridge-Chips 123

6.3 Transient-State Thermal-PDN Co-Analysis . . . 125

6.3.1 Transient-state IR-drop Co-Analysis Modeling Framework . . . 125

6.3.2 Transient-State Co-Analysis Results . . . 126

6.4 Conclusions . . . 128

Chapter 7: Summary and Future Work . . . 129

7.1 Summary of the Presented Work . . . 129

7.2 Future Research Extensions . . . 132

7.2.1 Thermo-Mechanical Analysis for Emerging Technologies . . . 132

7.2.2 Power Delivery Network and Thermal-PDN Co-Analysis . . . 132

7.2.3 PDN-Signaling Co-Analysis with Backside PDN . . . 133

7.2.4 Impact of Emerging Heterogeneous Integration Technologies on Network-on-Chip for Applied Machine Learning Algorithms . . . 133

References . . . 135

Vita 149

LIST OF TABLES

1.1 Key heterogeneous integration technologies 6

1.2 Relevant PSN work in the literature 14

1.3 Relevant thermal-PSN co-simulation work in the literature 15

2.1 Simulation setup 21

2.2 Material properties 24

2.3 Interposer warpage and worst-case MFI stress for different MFI orientations 33

3.1 Validation results (modeling vs. open source benchmarks) 43

3.2 Parameters of the PDN model 49

3.3 Transient state analysis results 53

3.4 PDN parameters 68

3.5 Thermal simulation parameters 76

3.6 Thermal results 76

4.1 Chip specifications 84

4.2 General parameters for PDN model 85

4.3 Specifications for conventional multi-die FC-BGA and FOWLP PDN modeling 87

4.4 Specifications for 3-D FC-POP and FOWLP-POP PDN modeling 88

5.1 PDN specifications for different configurations 103

5.2 Summary physical design results for a RISC-V architecture 108

5.3 Summary PSN results from physical design and PDN modeling 109

6.1 Thermal simulation parameters 117

6.2 Comparison of different interaction models 124

6.3 Impact of bridge-chip splitting 124

LIST OF FIGURES

1.1 (a) Annual data usage by 2025, (b) Power consumption by data centers . . . 3

1.2 (a) Microsoft FPGA accelerator, (b) Xilinx project EVEREST, (c) Cerebras wafer scale engine, and (d) Intel Foveros 3-D integration 4

1.3 (a) ITRS projection for server current, (b) Thermal coupling between dice placed on the same package 5

1.4 (a) Xilinx FPGA with interposer, (b) AMD GPU with HBM, and (c) NVIDIA GP100 6

1.5 (a) Intel EMIB technology, (b) Georgia Tech HIST platform, and (c) imec bridge technology 7

1.6 (a) Solder joint crack, (b) Interconnect and via delamination, (c) TSV crack, and (d) Grating coupler misalignment 9

1.7 (a) Mechanically flexible interconnect (MFI) and (b) Compliance measurement of MFI 10

1.8 Target impedance of a PDN and capacitor requirement for different frequency ranges 11

1.9 Lateral thermal coupling between dice in an interposer based 2.5-D configuration 15

1.10 Vertical thermal coupling between dice in an 3-D IC configuration 15

1.11 Thermal-PDN interaction models 16

2.1 MFI-enabled large integrated system with interposer-on-motherboard 20

2.2 (a) top view, (b) side view of an MFI. 21

2.3 MFI distribution with (a) baseline orientation (b) radial orientation 22

2.4 Boundary condition for the simulations. Only the interposer is shown. The MFIs (not shown in the figure) are on top of the interposer 23

2.5 Worst case MFI stress distribution with adaptive meshing (a) Initial meshing and (b) denser meshing 24

2.6 Thermally induced warpage and stress results for different configurations . 25

2.7 MFI distribution along the diagonal of the interposer 26

2.8 Flow diagram for optimization methodology 27

2.9 (a) Initial design and, (b) Optimized design of MFI 28

2.10 Stress distribution in worst case MFI for optimized design 29

2.11 For different interposer sizes, (a) warpage comparison between different configurations and, (b) worst case MFI stress 29

2.12 MFI distribution following the deformation contours 31

2.13 (a) Optimized MFI structure and, (b) worst case MFI stress distribution . . . 32

2.14 Interposer maximum warpage and MFI maximum stress tradeoff 32

2.15 Optimized MFI shape for different pitches 34

2.16 Optimized MFI warpage and stress for different pitches 35

3.1 (a) The PDN modeling hierarchy. From left to right: lumped model of the board-level PDN, distributed model of the package-level PDN, and the distributed model of the on-chip PDN. (b) Flow diagram of the PDN analysis showing different steps of the framework. 39

3.2 The two-layer (two power and two ground) package PDN model of power/ground planes 40

3.3 The on-die PDN model. Only one current source and one C4 bump is shown. 40

3.4 Re-organization of a non-uniform PDN layout 41

3.5 Map fine-grained power PDN layout to coarse meshing grids (a) vias (b) wires. 41

3.6 The noise profile of IBM3 benchmark (a) results from open-source *IBM* PG benchmarks and (b) our modeling results. 44

3.7 Bump current comparison for IBM3. 45

3.8 The transient noise of the node with the maximum error for IBM2. 46

3.9 Various heterogeneous integration platforms (a) interposer, (b) bridge-chip within fan-out, (c) EMIB, and (d) HIST 46

3.10 Illustration of bridge-chip placement (a) a single large bridge-chip (b) five small bridge-chips. 47

3.11 The current density of each die. (a) die #1 (b) die #2 48

3.12 The IR-drop profiles of both dice for (a) standalone, (b) interposer case with one TSV per C4 bump, (c) interposer case with 25 TSVs per C4 bump, (d) single bridge-chip, and (e) five bridge-chips. 50

3.13 IR-drop comparison of interposer and bridge-chip technologies as a function of key parameters for each case 51

3.14 (a) Impedance analysis of a single on-die PDN node and illustration of the switching current activity (a) waveform #1 1 GHz frequency (c) waveform #2, 4 GHz frequency 52

3.15 Transient analysis results of the point with largest droop (a) waveform #1 (b) waveform #2 54

3.16 PDN schematic diagram (a) excluding bridge-chip PDN and (b) including bridge-chip PDN 55

3.17 (a) Ground net in the bridge-chip, (b) power and ground nets in the bridge-chip, and (c) metal-insulator-metal capacitors in the bridge-chip 55

3.18 (a) CPU-FPGA configuration with re-routed PDN for the peripheral circuits, (b) die-to-package bump map with no PDN in the bridge, and (c) die-to-package bump map with ground net in the bridge-chip 56

3.19 DC IR-drop results for (a) no PDN in the bridge-chip, (b) ground network in the bridge-chip, and (c) both power and ground network in the bridge-chip 57

3.20 Impact of bridge-chip PDN resistance on DC IR drop 58

3.21 Transient analysis results for a 1 GHz pulse on-die excitation for (a) CPU die excluding bridge-chip PDN, (b) CPU die including bridge-chip PDN, (c) FPGA die excluding bridge-chip PDN, and (d) FPGA die including bridge-chip PDN 59

3.22 Transient analysis results including metal-insulator-metal capacitors in the bridge-chip for (a) CPU die and (b) FPGA die 60

3.23 Impact of MIM capacitor density on (a) PSN and (b) high frequency ripple 61

3.24 Impact of MIM capacitor density on maximum noise location for (a) no MIM capacitors and (b) 10 nF/mm^2 MIM capacitor density 62

3.25 HBM-FPGA configuration with bridge-chip 62

3.26 (a) FPGA-stacked memory power specifications and power map for memory (b) core die 0, (c) core die 1, (d) core die 2, (e) core die 3 63

3.27 DC IR-drop results for the FPGA and memory dice (a) excluding PDN in the bridge-chip and (b) including PDN in the bridge-chip 64

3.28 DC IR-drop results for different memory dice in the stacked memory; (a) core die 0, (b) core die 1, (c) core die 2, and (d) core die 3 65

3.29 Impact of bridge-chip overlap for a die with varying power 66

3.30 Benchmark architectures 67

3.31 The non-uniform current density map used for the analysis 67

3.32 a) Single on-package VRM configuration, b) four on-package VRM configuration; DC IR-drop for c) single on-package VRM case, d) four on-package VRMs case with uniform current density map; e) single on-package VRM case, and f) four on-package VRMs case with non-uniform current density map 69

3.33 Comparison of DC IR-drop for different configurations 70

3.34 Comparison of maximum IR-drop for different VRM-chip gaps in the on-package VRM configurations 71

3.35 Comparison of IR-drop for different bump pitches in the 3-D IC chip-on-VRM configuration 72

3.36 Comparison of transient noise for different on-package VRM configurations 73

3.37 PSN comparison for different key benchmarks 74

3.38 Maximum PSN of some key configurations for different on-chip decap density 75

3.39 Processor and VRM maximum temperature for different configurations with respect to different VRM power density 78

3.40 Power delivery limit of different configurations for (a) uniform and (b) non-uniform current density map 79

4.1 Packaging trend including FOWLP technology 82

4.2 (a) Conventional multi-die flip-chip configuration and (b) Conventional multi-die FOWLP configuration 83

4.3 (a) 3-D flip-chip POP configuration and (b) 3-D FOWLP POP configuration 83

4.4 PDN structure for FOWLP technology 84

4.5 Loop inductance structure for FOWLP PDN 84

4.6 Voltage domains for each die in (a) Multi-die FOWLP and (b) 3-D FOWLP POP 85

4.7 DC IR-drop results for (a) Die-1 and (b) Die-2 for baseline, multi-die FOWLP, multi-die FC, 3-D FOWLP, FC POP configurations 89

4.8 Impedance analysis results (a) Die-1 and (b) Die-2 for FC, FOWLP, FC POP, and 3-D FOWLP configurations 90

4.9 Simultaneous switching noise based transient analysis results for (a) multi-die package and (b) 3-D package-on-package configurations. Each figure shows PSN results for both Die-1 and Die-2. The baseline configuration is a single die in a single package case. 91

4.10 Impact of solder bump pitch on (a) multi-die FOWLP and (b) 3-D FOWLP configurations. In both cases, we report the worst-case scenario; we report Die-1 results for multi-die FOWLP and Die-2 results for 3-D FOWLP configurations. 92

4.11 Impact of RDL resistivity on PSN for (a) multi-die FOWLP and (b) 3-D FOWLP configurations. 93

4.12 Impact of on-die PDN resistance on FOWLP supply noise 94

4.13 Impact of RDL to on-die PDN connectivity both FOWLP supply noise . . . 95

4.14 Double sided RDL in FOWLP POP structure 95

4.15 (a) IR-drop analysis and (b) transient analysis results for Double sided RDL 3-D FOWLP structures. The figures show a comparison between Die-2 PSNs for 3-D FOWLP and 3-D double sided RDL configurations, respectively. 96

4.16 Different TMV-BGA distribution for the top die in 3-D FOWLP configurations: (a) Single line BGA+TMVs and (b) Dual-line BGA+TMVs 97

4.17 Power supply noise comparison for different integration technologies 98

5.1 Die placement and metal configurations for a conventional front-side PDN configuration and backside PDN configuration 101

5.2 On-die PDN structure for (a) conventional interleaved BEOL PDN configuration and (b) meshed backside PDN configuration 101

5.3 Power supply noise results for uniform power map 105

5.4 (a) Power map with five adjacent hotspots, (b) PSN results for the hotspot power map with zero background power, and (c) PSN results for the hotspot power map with 17.1 W uniform background power 106

5.5 Physical design results for different PDN configurations 108

5.6 Peak IR-drop comparison for different package-to-chip bump pitches 110

5.7 Impact of rise time variation on step response for backside PDN configuration. The red line shows the step response result for conventional BEOL PDN with 1 ns rise time 111

5.8 (a) Step response results for different MIM densities and (b) supply noise for 1 GHz pulse input 111

5.9 Temperature distribution for (a) conventional front-side PDN configuration and (b) backside PDN configuration 113

6.1 Thermal-PDN interaction models 115

6.2 Bridge-chip based simulation configuration 116

6.3 Top view of thermal profile of each die (processor-FPGA) 117

6.4 Maximum temperature of different layers with respect to the variation in low power die power 118

6.5 Thermal profile of (a) 74 W processor – 44.8 W FPGA integration and (b) 74 W processor – 74 W FPGA integration 119

6.6 The flow chart for the thermal-PDN co-analysis 120

6.7 The temperature distribution for (a) standalone model, and (b) co-analysis model in multi-chip packages 121

6.8 The temperature distribution for (a) standalone model, and (b) co-analysis model in bridge-based 2.5-D packages 122

6.9 Steady-state IR-drop comparison for different configurations 122

6.10 Different interaction models (a) standalone model, (b) thermal-leakage model, and (c) full model 123

6.11 The flow chart for transient thermal-PDN co-analysis 125

6.12 PDN step response results for (a) standalone model, and (b) co-analysis model in bridge-based 2.5-D packages 126

6.13 Average temperature profile for a 1 GHz on-die excitation 127

6.14 PDN results for a pulse excitation for (a) standalone model, and (b) co-analysis model in bridge-based 2.5-D packages 127

7.1 3-D stacking with backside PDN for (a) face-to-face bonding, and (b) fan-out wafer level packaging based package-on-package 133

SUMMARY

Interconnect technologies are undergoing a revolution to meet the rapid growth in system interconnection requirements. A number of 2.5-D and 3-D integration technologies are being explored to integrate high-performance dice such as, CPU, FPGA, GPU, etc. with memory. The mobile computing space is expanding its opportunities as well. In all these configurations, while there are a number of benefits in communication bandwidth, power efficiency, footprint reduction, there are important thermal, mechanical, and electrical considerations that need to be addressed. To enable the design space exploration of these systems from the perspective of temperature and power supply noise, a thermal and a PDN modeling framework is presented. Various 2.5-D and 3-D heterogeneous integration technologies are investigated and benchmarked for thermal and electrical performance and inter-dependencies.

First, the use of flexible interconnects for thermo-mechanical reliability improvement in interposer assembly is analyzed. The goal of this work is to explore the opportunity to remove the secondary organic substrate from an assembled subsystem. Hence, a thermally-induced warpage comparison between solder bumps and mechanically flexible interconnects (MFIs) in an interposer-to-motherboard assembly is reported. Impact of chip size on thermo-mechanical warpage and stress is also evaluated. A comprehensive MFI distribution technique utilized for improved thermo-mechanical reliability and a genetic algorithm based structural optimization of MFIs are presented.

Second, power delivery network (PDN) modeling including advanced packaging of voltage regulator modules is evaluated. Different 2.5-D integration technologies are benchmarked. Specifically, a bridge-chip based 2.5-D integration technology is benchmarked with and without a PDN in the bridg-chip. Both steady-state IR-drop and transient Ldi/dt noises are reported.

Third, PDN modeling and benchmarking of fan-out wafer level packages (FOWLP) is

evaluated. Both multi-die FOWLP and 3D FOWLP technologies are benchmarked with respect to corresponding flip-chip Ball Grid Array (FC-BGA) configurations. Power supply noise results for both steady-state and transient-state simulations are presented. A comprehensive design space exploration of FOWLP technologies is performed.

Fourth, a new PDN architecture named 'backside PDN' is benchmarked. The differences between backside and conventional front-side PDN configurations are introduced. The power delivery performance of a backside PDN configuration is evaluated. Results for different power maps are presented. Moreover, the modeling results are validated with physical implementation results. A design space exploration is performed to analyze the impact of package-to-die interconnection pitch, input pulse, capacitor density on PDN performance. Additionally, thermal implications of dielectric bonding for a backside PDN configuration are evaluated.

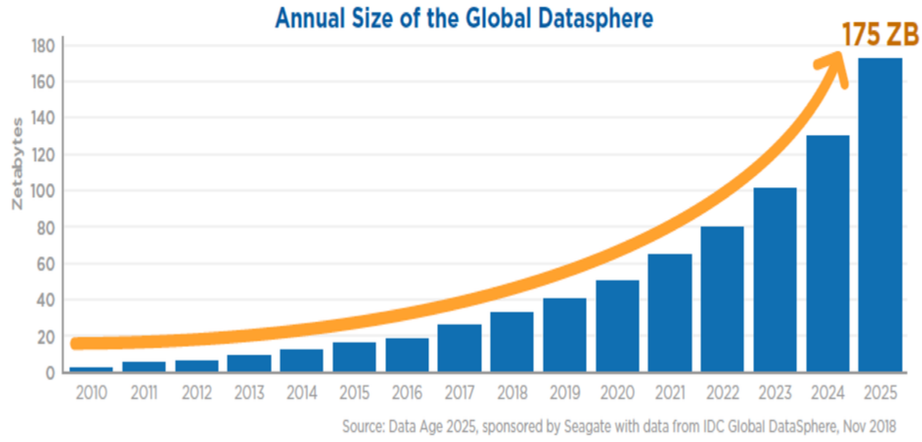
Lastly, a framework for thermal-PDN co-analysis is extended to evaluate bridge-based 2.5D integration technologies. Inter-dependencies between temperature distribution of the dice in a package and the supply voltage noise are captured. Some thermal aspects of a bridge-chip based 2.5-D integration are highlighted. Thermal-PDN frameworks for both steady-state and transient-state PDN are presented. Impact of different interaction models is characterized.

CHAPTER 1

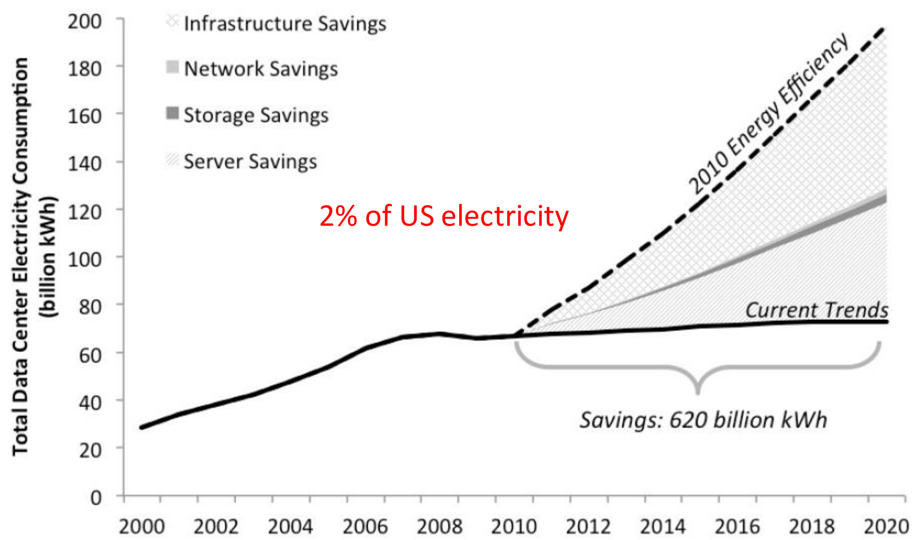
INTRODUCTION

The demand for data generated by different applications in machine learning, artificial intelligence, internet-of-things, etc. is exponentially increasing, driving the need high-performance computing systems. Fig 1.1(a) shows this trend and explosion of data [1]. This significant volume of data is driving the growth of data centers around the world. Fig. 1.1 shows the electical energy consumed by data centers in the US[2]. Although the growth in the number of data centers has decreased over the years, the increase is scale-out and scale-up growth of large data centers is significant. In 2010, electricity used in data centers globally was $\sim$1.5% of total electricity use. This consumption is $\sim$2% of total electricity consumption in the US, as shown in Fig. 1.1. To this ends, there have been innovations in infrastructure, network, storage, and server platforms, which help reduce the overall power consumption of data centers. While power savings is an important factor for these data centers, heat removal is an additional challenge which adds significant overhead cost. Cooling accounts for $\sim$30% of the overall power consumption of a data center [3]. For example, Microsoft reported that its under-water cooling project supports a 240 kW data center [4]. The key components of these data centers are computing blocks or processing units and storage units or memory. In order to keep up with the >TB bandwidth requirements of rapidly evolving computing fabrics such as FPGAs integrated with server class CPUs, several emerging integration technologies have been studied. Some of the key heterogeneous integration technologies include interposer/bridge 2.5-D ICs [5, 6], 3-D ICs [7], and fan-out wafer-level based packages including package-on-package (PoP) technology [8]. For example, the Stratix 10 FPGA currently integrates a large programmable fabric and daughter dice, such as transceivers and High Bandwidth Memory (HBM), with high-bandwidth EMIB links [9]. Further integration of Xeon CPU dice with FPGA dice

into a single package could greatly enhance computing performance and efficiency for many applications. Since this is still an ongoing research field, more innovative advances are anticipated to be proposed in the near future. As such, the computing platforms are migrating towards a modular based package design with compact heterogeneous integration of CPU, GPU, FPGA, memory, etc. Fig. 1.2(a) shows a system used for Microsoft Bing search where FPGAs are used to accelerate the computations [10]. Fig. 1.2(b) shows a heterogeneous interconnect fabric developed by Xilinx [11]. Fig. 1.2(c) shows a wafer scale engine (WSE) where the whole wafer is used to make one single complete system [12]. This WSE has 1.2 trillion transistors, 400k linear algebra cores, 18 GB of on-die memory, 9 PB/sec of memory bandwidth across the chip, and separate fabric bandwidth of up to 100 Pb/sec. Fig. 1.2(d) shows a 3-D package-on-package configuration developed by Intel [13]. Similar to these approaches, a semiconductor package may contain a number of functionalities including but not limited to stacked memories, RF devices, application processors, MEMS, power management ICs, etc. In these configurations, there are important thermal, mechanical, and electrical considerations that need to be addressed. As functionally-diverse dice are packed into a smaller space, the corresponding increased thermal load is an added concern to the thermo-mechanical reliability of the solder joints. Moreover, owing to these advanced technologies, the total power density is expected to increase beyond 100 W/cm^2 [14]; power delivery becomes a critical challenge, and advanced cooling solutions (for example, microfluidic cooling) are turning into a necessity [15]. Fig. 1.3(a) shows the increase in per socket current requirement for the server chips. Reduced noise margin determined by the scaling trend of the technology is making the power delivery to the chip ever more challenging. Placing dice side-by-side, as shown in Fig. 1.3, poses thermal coupling issues where heat flows from the high power die to the low power die. Moreover, temperature, supply voltage, and power dissipation are dependent on each other. The temperature impacts the leakage power and the power/ground grid resistivity. Power dissipation determines the source current of the chip and is also the ex-



(a)



(b)

Figure 1.1: (a) Annual data usage by 2025, (b) Power consumption by data centers

citation of the power delivery network (PDN) noise. However, the power supply voltage impacts both leakage and dynamic power. Without considering the interactions between each of the individual interaction models, for emerging architectures with increased power density, the results from the standalone or partially integrated models could be overestimated by as much as 30% [16]. Therefore, in this research effort, thermal-mechanical and thermal-power interactions are investigated.

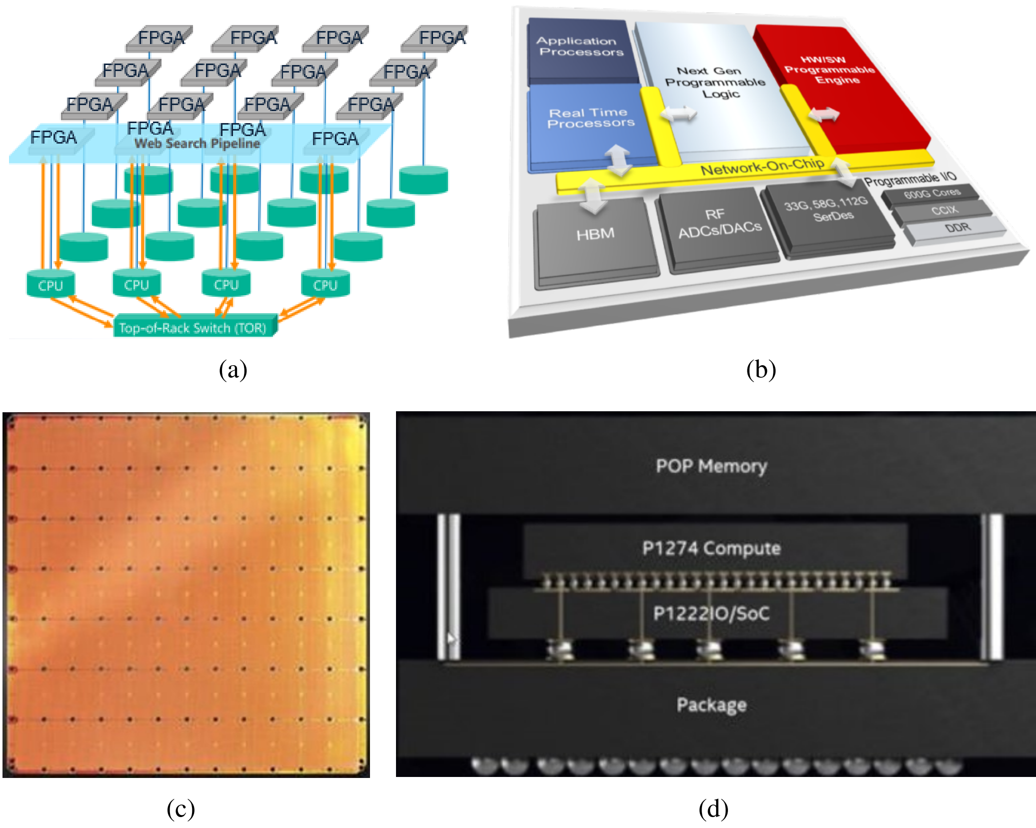


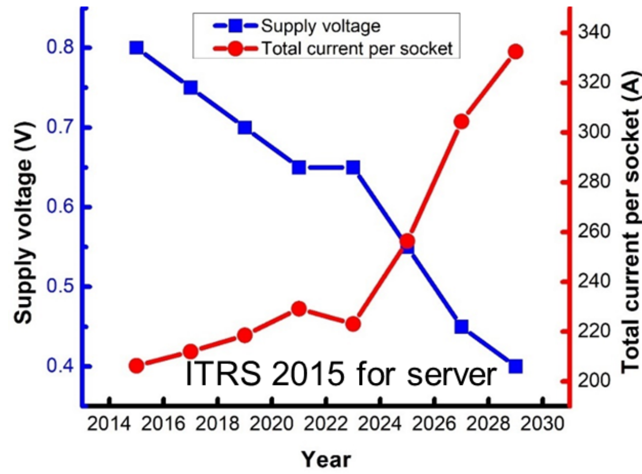
Figure 1.2: (a) Microsoft FPGA accelerator, (b) Xilinx project EVEREST, (c) Cerebras wafer scale engine, and (d) Intel Foveros 3-D integration

1.1 Current Relevant Research

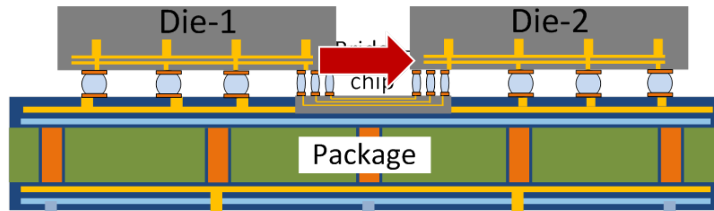
The evolution of different 2.5-D/3-D integration technologies brings about a number of challenges. Thermo-mechanical reliability, thermal integrity, power integrity, etc. are a few of them. Significant research effort has been put to address these challenges. Some of the noteworthy efforts are delineated in the section below.

1.1.1 Heterogeneous Integration Technologies

Heterogeneous integration technologies (2.5-D/3-D ICs) provide high-bandwidth density and low-energy connectivity as well as ultra-small form factors. Table 1.1 summarizes some key 2.5-D/3-D heterogeneous integration technologies. Silicon interposer has shown



(a)



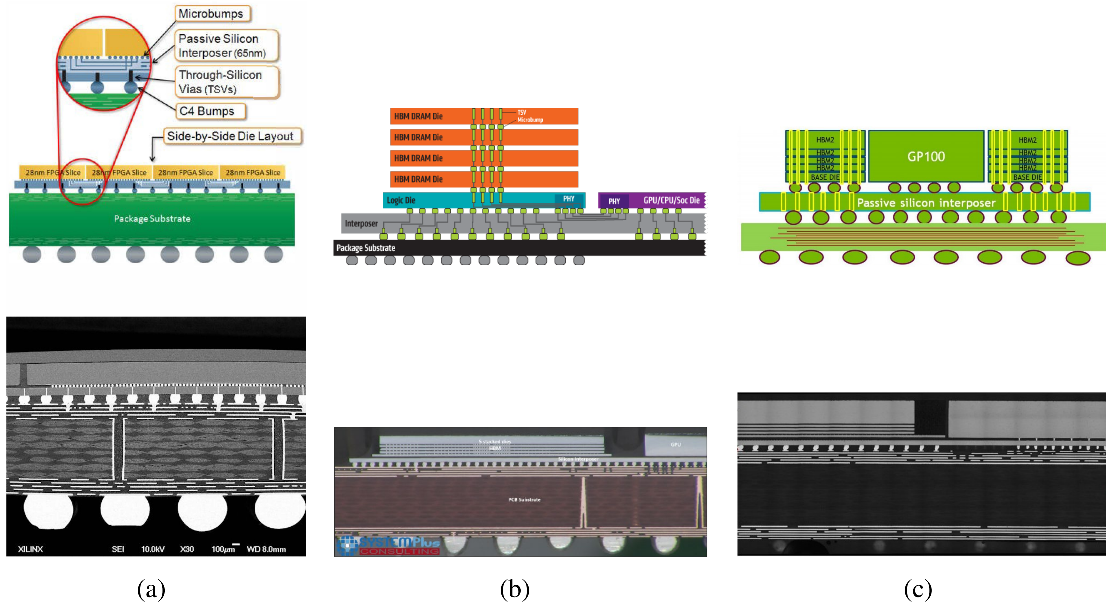
(b)

Figure 1.3: (a) ITRS projection for server current, (b) Thermal coupling between dice placed on the same package

its potential in 2.5-D integration [20, 21, 22, 19]. Fig. 1.4(a), 1.4(b), 1.4(c) show three such interposer based products offering 348 GBps of aggregate bandwidth, 512 GBps of memory bandwidth, and 160 GBps CPU-to-GPU NVlink bandwidth, respectively. This technology provides high density die-to-die interconnections. Alongside, this fabric also provides additional spread of current for power delivery [5]. Interposer is also utilized for 3-D integration technologies. However, owing to the increased power density [14], while interposers are being widely used for 2.5-D integration technologies [23, 22], their use in 3-D ICs is primarily limited to the memory devices [24, 25]. Some other noteworthy 2.5-D integration technologies are Intel's Embedded Multi-die Interconnect Bridge (EMIB) [26], Georgia Tech's Heterogeneous Interconnect Stitching Technology (HIST) [27], and imec's fan-out based bridging concept [6], as shown in Fig. 1.5. These bridging technologies increase the die-to-die signaling bandwidth while eliminating the 'reticle limitation' of the

Table 1.1: Key heterogeneous integration technologies

	Silicon interposer [17]	EMIB [18]	Bridge-chip [6]	Foveros [13]	Chip stacking [19]
Interconnection method	2.5-D	2.5-D	2.5-D	2.5-D & 3-D	3-D
I/O structure	Bump	Bump	Bump	Bump & Bump	Bump
Pitch	30-60 μm	55 μm	20 μm	-	>8 μm
Scalability	Limited	Scalable	Scalable	Limited	Limited



(a) (b) (c)

Figure 1.4: (a) Xilinx FPGA with interposer, (b) AMD GPU with HBM, and (c) NVIDIA GP100

interposers.

Recently, Fan-out Wafer Level Packaging (FOWLP) technology has shown its potential to significantly miniaturize the package[28]. The advantages of FOWLP technology are not only related to a significant package miniaturization in the lateral directions, but it also reduces the package thickness significantly. Package I/Os are redistributed across the entire package including the fan-out region outside of the silicon die for increased pin count at the package level. The absence of the substrate reduces the thermal resistance of the package, increases the electrical performance owing to the shorter interconnections and

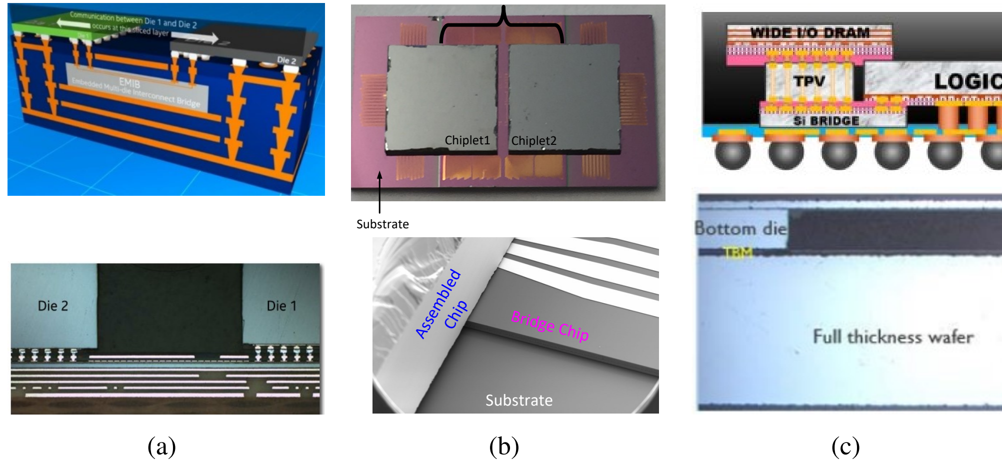


Figure 1.5: (a) Intel EMIB technology, (b) Georgia Tech HIST platform, and (c) imec bridge technology

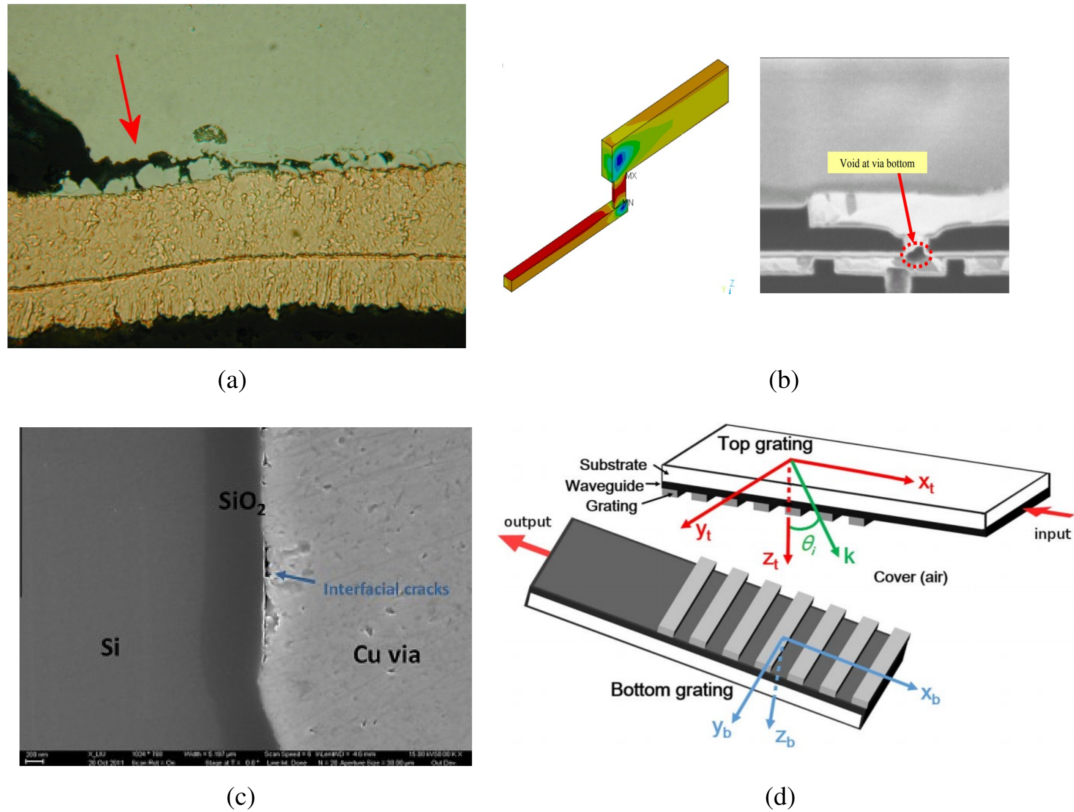
lowers parasitic effects. FOWLP offers as low as 8x reduced PDN impedance compared to a flip-chip pakcage [29]. Moreover, 3-D FOWLP, like conventional package-on-package (POP) configurations, enables added functionalities and miniaturization in the third dimension. TSMC's FOWLP technology shows 20% reduction in package thickness compared to a flip-chip package [8]. FOWLP has been under extensive investigation in recent years [28, 8, 30, 31, 32]. Some noteworthy examples include TSMC's Integrated Fan-out wafer level packageing (InFO WLP) [8], Infineon's embedded Wafer Level Ball Grid Array (eWLB) [31], and Freescale's Redistributed Chip Package (RCP) [32].

While heterogeneous integration is pushing for modular based package designs, recent trends indicate that a single die itself can be heterogeneous in nature [33] where different computing blocks are fabricated with different technology nodes. Besides, there is also co-integration of voltage regulators, inductors, etc. in the same package. In recent years, on-chip regulators have gained significant attention because of their fine grain voltage control, increased availability of power, increased performance, decreased inductor size, etc. [34, 35, 36]. An on-chip regulator with an inductor placed in the package is shown in [34]. A 2.5-D based integrated voltage regulator (IVR) where the inductors are placed right beneath the chip is presented in [35]. These technologies eliminate the need for multiple VRs in the

case of multiple supply voltage systems while reducing the parasitic length of the power delivery path, enabling active power management required by high-performance computing devices.

1.1.2 Thermo-Mechanical Reliability Analysis for Advanced Packaging Technologies

Numerous heterogeneously integrated systems are being used to assemble chips side-by-side, and thus allowing designers to put dice next to each other in a high-bandwidth, low-energy configuration. For such a system, thermally induced warpage is an increasing concern for device and interconnect reliability [37, 38, 39], as shown in Fig. 1.6. For example, a 2.5-D interposer-based integration technology requires an organic package to minimize the effects of CTE mismatch, employing Ball Grid Array (BGA) between the package and the board, and C4 bumps between the package and the interposer [20, 21, 40, 41]. These packages have multiple layers and typically use underfill to ensure the reliability of the C4 bumps between the package and the interposer. As functionally-diverse dice are packed into a smaller space, the corresponding increased thermal load is an added concern to the thermo-mechanical reliability of the solder joints [42, 43]. Thermally-induced warpage also affects 3-D integrated systems as through-silicon-vias (TSVs) are subjected to mechanical stresses and strains [20, 21, 44]. Similarly, thermally-induced warpage may negatively impact the coupling efficiency of optical grating couplers as this warpage offsets the necessary alignment for high coupling efficiency [45, 46, 47, 48, 39]. For example, Wan et. al. [39] reported a 25% reduction in diffraction efficiency for a 5.73^o angular displacement. This is an important consideration since silicon photonics is evolving as an enabling technology for high-performance computing which uses interlayer grating couplers for transferring optical signals between out-of-plane waveguides. Moreover, with the advancement of semiconductor processes, there are innovative packaging solutions that increase the interconnection complexity [22, 49, 50]. The International Technology Roadmap for Semiconductors (ITRS)[14] predicts the substrate-to-board pitch to be approximately 300 μm by



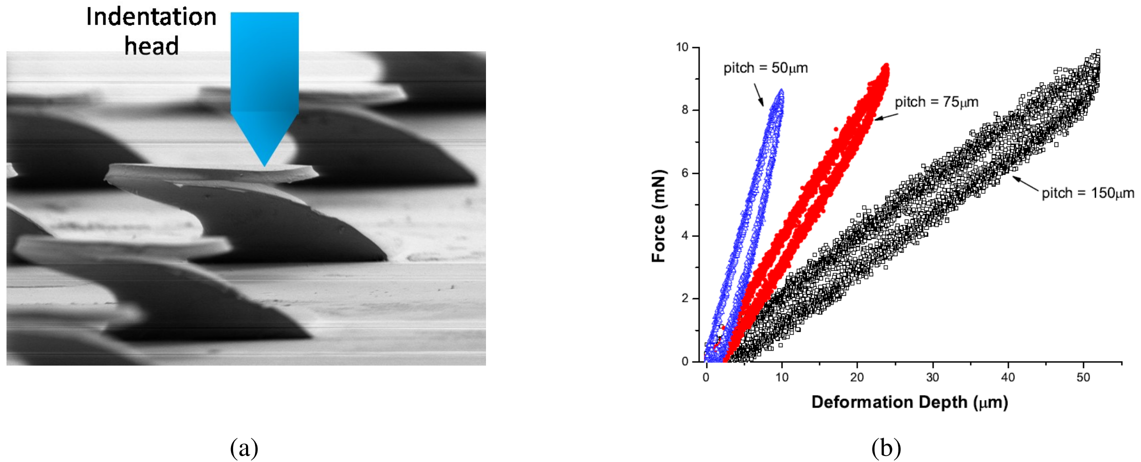
(a) (b)

(c) (d)

Figure 1.6: (a) Solder joint crack, (b) Interconnect and via delamination, (c) TSV crack, and (d) Grating coupler misalignment

2026, which may introduce numerous reliability concerns. ITRS also projects that the peak package warpage limit, which occurs during the solder ball reflow process, to be as low as 50 μm for 300 μm pitch BGA. Therefore, it is critical to minimize warpage as warpage-induced stress/strain only functions to negatively impact the reliability and performance of a wide variety of systems and technologies. There have been significant prior efforts to reduce substrate warpage. Raghavan et al. [51] outlined a temperature profile modification and external mold technique to reduce warpage. Mikael et al. [52] showed the impact of different process conditions from analytical and experimental data on substrate warpage. They proposed solutions including a thicker insulator layer, thinner metal1/metal2 layers, etc. in an effort to reduce warpage. Chaware et al. performed a reliability analysis of different underfill materials [21]. Murayama et. al, proposed possible solutions for warpage control including chip first process, usage of underfill with low T$_g$, etc.[53].

Compliant interconnects have been used to address some reliability issues generated from the CTE mismatch between organic/ceramic packages and silicon die[54, 55, 56, 57, 58]. Recently, Mechanically Flexible Interconnects (MFIs), as shown in Fig. 1.7, have



(a) (b)

Figure 1.7: (a) Mechanically flexible interconnect (MFI) and (b) Compliance measurement of MFI

been investigated as enablers for direct assembly of a Si interposer onto a motherboard to achieve a smaller profile and better electrical performance [59]. Apart from reducing the thermally induced warpage, MFIs could eliminate the secondary substrate in some applications, resulting in a smaller form factor, higher bandwidth, lower power, and shorter interconnections. Flexible interconnect design and optimization are also carried out to tackle CTE mismatch in bridging concepts, such as HIST [60, 61]. Moreover, component-level optimization has been carried out in numerous studies [61, 62, 63]. The primary focus of these studies is to design and optimize a single interconnect under a mechanical loading condition. For example, in [61], the focus is MFI optimization under a nano-indentation load. Multi-objective single interconnect optimization is carried out in [57]. However, design and optimization of a group of interconnects based on system level parameters, e.g. thermal loading, is missing in the literature.

1.1.3 Power Delivery Network Modeling for 2.5-D and 3-D ICs

Power requirements in modern high-performance computing systems are becoming increasingly stringent. Such systems typically contain several cores [64] to tens of cores [65] with multiple power supply domains [66]. Traditionally, the power supplies are placed off-chip to provide necessary load currents to the on-chip active circuitry. These systems typically have resistive and parasitic losses from the interconnects and metal pads. Large passive components (i.e., capacitors) are placed to somewhat compensate these effects. However, the power delivery challenges are becoming increasingly prominent as more and more transistors are being packed into a single chip, which eventually translate to increased load. A single package may include high-bandwidth memory with GPUs [25], FPGAs with server processors [67], high-performance GPUs with general purpose CPUs [68], etc. Despite the scaling of supply voltage in recent device technologies [69], these high-performance integrated modules inevitably lead to higher current demand and increased power density [58]. As a result, power delivery in high-performance digital systems is an increasingly difficult challenge [70]. In an electronic system, there are resonances from die-to-package, package-to-board, and board-to-supply interactions, as shown in Fig. 1.8. Meeting the target frequency over a wide frequency range is becoming increasingly dif-

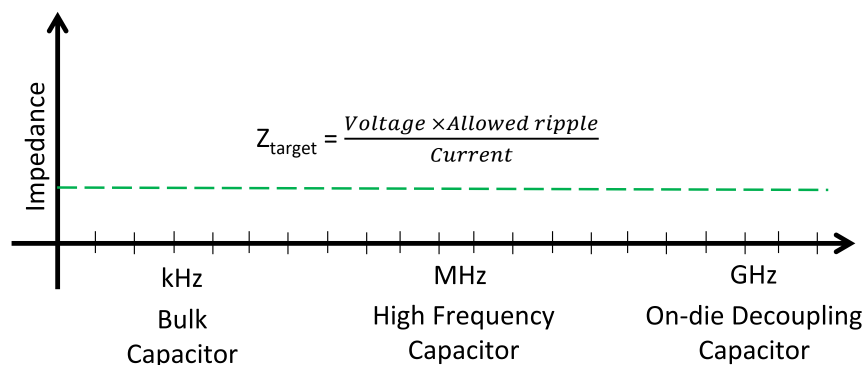


Figure 1.8: Target impedance of a PDN and capacitor requirement for different frequency ranges

ficult. Recently, on-chip regulators have gained significant attention because of their fine grain voltage control, increased availability of power, increased performance, decreased

inductor size, etc. [34, 35, 71, 72]. An on-chip regulator with inductor placed in the package is shown in [34]. These technologies eliminate the need for multiple VRs in the case of multiple supply voltage systems while reducing the parasitic length of the power delivery path, enabling active power management required by high-performance computing devices. In short, these are efforts to bring the power supply circuitry closer to the active circuits. There are a number of solutions to improve the efficiency and reduce the footprint of the active portion of a PDN [36, 5] and, one has to rethink the on-die PDN design to achieve the best out of both scaling trends and innovative packaging solutions. Specifically, scaling device technology poses several challenges. The resistivity and resistance of conventional metal layers and inter-metal vias are increasing rapidly with advanced technology scaling [73, 74], while PDN noise margins are becoming stringent [75]. Moreover, the power consumption of different computing blocks is increasing significantly [14, 76]. Power supply noise (PSN) negatively impacts the system performance; PSN induces clock jitter, which exacerbates the performance of a computing block [77]. Modern processors can create nanosecond level voltage droops that require different circuit techniques to ensure reliability. Moreover, advancement of the packaging technologies results in critical interfaces, which increases PDN impedance. For example, in a silicon interposer based 2.5-D integration [78, 79], there is a tradeoff between using additional PDN grids in the interposer to reduce PSN and added parasitics from the TSVs and microbumps. Likewise, in bridge-based 2.5-D integration technologies (EMIB, HIST, etc.), signal interconnections and I/O drivers are placed in the periphery of the dice. As such, the bridge may block the direct access of the package power/ground planes to the periphery of the dice [5], which increases the source-to-sink impedance for the die blocks in these regions. This effect is more prominent in a CPU-HBM or FPGA-HBM configuration since HBMs are wide I/O configurations with concentrated connections in the center of the die. This increases the overlap between a bridge-chip and the memory dice. While memory banks have power supply through TSVs, the base logic die suffers from longer PDN paths owing to this overlap re-

gion. Similar to bridge-chip configurations, FOWLP technologies have unique attributes as well. For example, owing to the dense redistribution layers (RDLs) in the package, the PDN in the RDLs is different from the power/ground planes in organic/ceramic package PDN. Moreover, some FOWLP technologies use copper pillars instead of coarse C4 bumps common in flip-chip packages. For all these innovative technologies, there is a need for evaluating the PDN early in the design cycle before it becomes expensive to adopt any changes in the latter stages.

Power supply noise (PSN) modeling has been under extensive research over the last decades [87, 81, 83, 7]. Some noteworthy contributions in PDN modeling are summarized in Table 1.2. DC IR-drop modeling of 2.5-D and 3-D integration systems is analyzed in [87]. However, Ldi/dt transient analysis or dynamic IR-drop and impedance analyses are missing in this work. DC and dynamic IR-drop is evaluated in [81]. However, the model uses a lumped model for the package PDN, which makes it harder to model emerging packaging technologies. Lumped model based PDN modeling is carried out in [83]. Recent work also addressed the power integrity modeling of fan-out wafer level packages [29, 30]. Chou et. al [30], provided impedance, DC resistance, and transient analysis results from experimentation. Wang et. al. [29], presented a power integrity model to investigate the scope of integrated voltage regulators in fan-out wafer level technologies. Yang et. al. [5] presented a PDN tool that can perform both steady state and transient analysis with distributed on-die, package, and board model. Different technologies have unique attributes which can negatively impact the PSN of a system. These efforts open the path towards accounting for such attributes in a quick and accurate manner.

1.1.4 Thermal-PDN Co-Anaysis Modeling

In a tightly integrated system, if multiple dice are placed side-by-side, there can be significant thermal coupling [7], as shown in Fig. 1.9. In 3-D ICs, owing to the vertical stacking, the temperature profile of the low power dice becomes an image of the temperature profile

Table 1.2: Relevant PSN work in the literature

	IR	Transient	AC	Distributed On-die PDN	Package PDN	Board PDN	VRM	Multi-VRMs	Die configurations	Package decaps	Power map
J.Xie [80]	Yes	No	No	Single-layer no vias	Distributed	Distributed	No	No	2.5/3-D	N/A	Non-uniform
R. Zhang [81]	Yes	Yes	No	Multi-layer no vias	Lumped	Lumped	No	No	3-D	Lumped	Non-uniform
S. Park [82]	Yes	Yes	Yes	Single-layer no vias	Lumped	Lumped	Yes	No	3-D	Lumped	Non-uniform
X. Zhang [83]	Yes	Yes	No	Lumped	Lumped	Lumped	No	No	2-D	Lumped	Non-uniform
H. He [84]	Yes	Yes	Yes	Single-layer no vias	Lumped	Lumped	Yes	No	3-D	Lumped	Uniform
Y. Shao [85]	Yes	No	No	Single-layer no vias	Distributed	Distributed	No	No	2.5/3-D	N/A	Non-uniform
C. Pan [86]	Yes	Yes	Yes	Multi-layer with vias	Distributed	No	No	No	2.5/3-D	Distributed	Uniform

of the higher power die [88]. Typically, a thermal model and a PDN model provide mutually exclusive results. However, there are inter-dependencies between these two models that require special attention, especially for heterogeneously integrated 2.5-D and 3-D ICs with advanced technology nodes. The inputs to a PDN modeling tool typically includes a power model. The power model includes both the dynamic power and the leakage power contributions of the active circuits. In an early analysis, the power is estimated based on

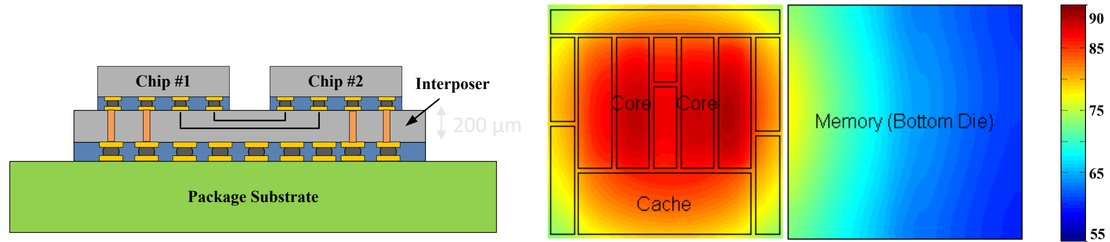


Figure 1.9: Lateral thermal coupling between dice in an interposer based 2.5-D configuration

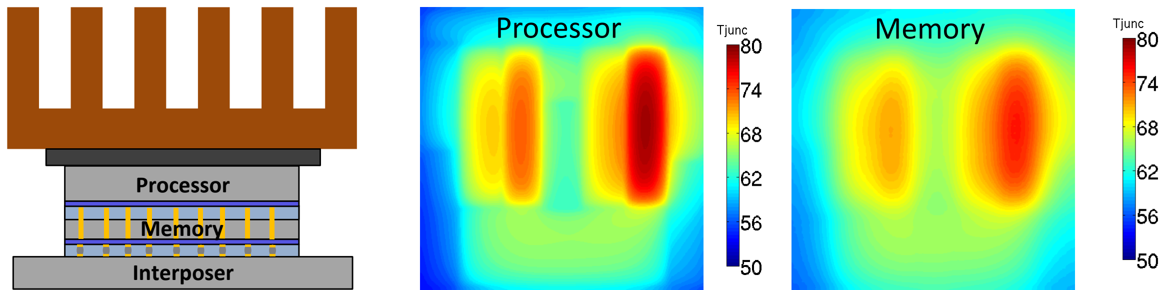


Figure 1.10: Vertical thermal coupling between dice in an 3-D IC configuration

Table 1.3: Relevant thermal-PSN co-simulation work in the literature

	Analysis type	Interactions
Y. Shao, J. Xie [85, 80]	Steady	Wire resistivity
Y.Liu [89]	Steady	Leakage power
H. Su [16]	Steady	Leakage power and dynamic power
S. Park [82]	Steady & Transient	Leakage power and wire resistivity
Y. Zhang [90]	Steady & Transient	Leakage power, dynamic power, and wire resistivity

architectural tools and data sheets which provides power specifications at different temperatures. However, from a realistic thermal map, the temperature across a single die can be different. Moreover, there are different thermal solutions [15] that can impact the temperature, and hence, impact the performance of a system. The power model, temperature of a die, and the PSN are interdependent. Fig. 6.1 shows the dependencies between power dissipation, temperature, and PDN. The temperature impacts the leakage power and the grid resistivity of the PDN. Conversely, the power supply voltage impacts both leakage

and dynamic power. Without considering the interactions between each of the components in Fig. 6.1 for emerging architectures with increased power density, the results from the standalone or partially integrated models could be overestimated.

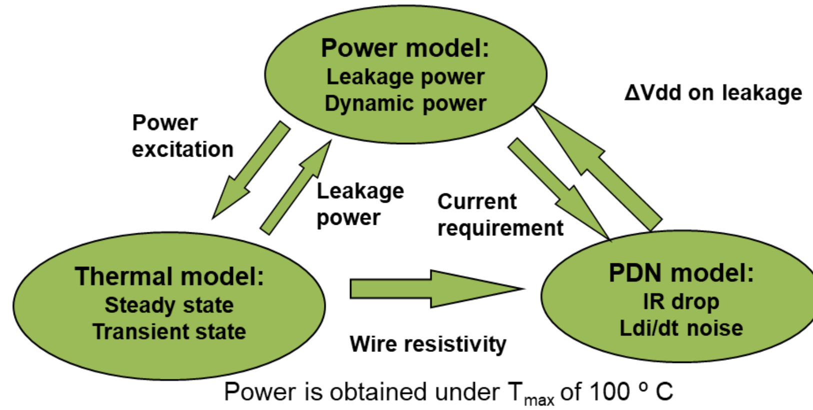


Figure 1.11: Thermal-PDN interaction models

Researchers have put efforts to address these inter-dependencies among different interaction models [16, 80, 7]. Some of these efforts are summarized in Table 1.3. Xie et. al. [80] studied the interaction between temperature distribution and the steady state IR-drop. Su et. al. [16] studied the impact of temperature and supply voltage on the power dissipation of the dice. Yang et. al. [7] incorporated the inter-dependencies of all the interaction models for a 3-D stacked processor-memory system. All these modeling techniques are significant efforts to address the issues related to the advanced technologies. However, investigation of 2.5-D technologies from the co-analysis perspective is missing in the literature. Also, most of the prior works mentioned in this section do not account for both steady state and transient analysis. With a complete and comprehensive thermal-PDN co-analysis tool, the state of the art design methodology will gain added momentum and reduce the design space significantly before moving to the full design cycle.

1.2 Organization of This Thesis

This document is arranged as follows:

- Chapter 2: This chapter explores different means by which both interconnect reliability is improved and interposer warpage is decreased for an interposer-on-board integration using mechanically flexible interconnects (MFIs). Central to this exploration is the design and distribution/orientation of the MFIs on the interposer. Using Finite Element based tool ANSYS, different MFI distributions and configurations are investigated. Using MFIs for interconnection, a minimum 43% improvement in warpage is reported. Employing a genetic algorithm based structural optimization technique, greater than 50% reduction in MFI stress is presented.

- Chapter 3: We present a PDN modeling framework with a focus on multi-die heterogeneous integration. We show the detailed formulation and analysis methods in this chapter. A design space exploration of power delivery networks is performed for 2.5-D and 3-D integrations. This chapter focuses on PDN modeling of different 2.5-D configurations including interposer and bridge-chip based technologies. We show that by splitting a bridge-chip into multiple smaller bridge-chips, on-die PDN impedance can be reduced. We also study these scenario with a PDN in the bridge-chip. If we use PDN in the bridge-chip, the DC IR-drop can be reduced by more than 20% compared to a configuration excluding a bridge-chip PDN. This chapter also includes a study regarding effective placement of the voltage regulator modules (VRMs) for power supply noise (PSN) suppression. Multiple on-package VRM configurations have been analyzed and compared. Additionally, 3-D IC chip-on-VRM and backside-of-the-package VRM configurations are studied. We also study the thermal implications of different VRM placements. We observe a steep rise in temperature when we place the VRM on the backside of the package. We also perform a study to evaluate the power excitation limit of different configurations for a specific PDN noise level.

- Chapter 4: We present a PDN modeling framework for Fan-out Wafer Level Packag-

ing (FOWLP) technologies with a focus on multi-die heterogeneous integration. Results are compared to conventional multi-die packaging and 3D package-on-package technologies. Owing to the shorter interconnections enabled by thinner packages and elimination of large C4 bumps with copper pillars, the package contributes less parasitics to the PDN path. Hence, the IR-drop, transient droop, and impedance are reduced in the evaluated FOWLP technologies. We perform a design space exploration to investigate the impact of different design parameters: BGA pitch, metal layers, via distribution, copper pillar pitch, etc. on PSN.

- Chapter 5: We present a PDN modeling framework for backside PDN configurations. A backside PDN approach separates the PDN from a conventional signaling network of the back-end-of-the-line (BEOL) and improves power integrity and core utilization. We benchmark this technology with conventional front-side BEOL PDN configurations. Owing to the lower resistivity compared to Cu metal lines for advanced technology nodes, we use Ruthenium (Ru) based buried power rail for PDN modeling. The framework results are validated with a place-and-route (P&R) based physical implementation flow. We quantify the area improvement in the actual flow and observe 25%-30% improvement in the backside PDN configuration. Moreover, we investigate the impact of package-to-die interconnect pitch, metal-insulator-metal cap density, and input pulse on PDN performance. Additionally, we perform thermal modeling to analyze thermal implications of a backside PDN configuration.

- Chapter 6: We present a PDN modeling framework for heterogeneous 2.5-D integration platforms. Both steady state and transient state (Ldi/dt) noise analyses have been presented for a conventional multi-die package and a bridge-chip based package. Compared to thermal-PDN co-simulations, we observe a 10-12% overestimation in the steady state temperature and IR drop results and a 20% overestimation in the Ldi/dt noise in standalone PDN simulations without thermal impacts.

- Chapter 7: This chapter presents the conclusions and summary of this thesis; future research topics are also discussed.

CHAPTER 2
THERMOMECHANICAL ANALYSIS AND PACKAGE LEVEL OPTIMIZATION OF MECHANICALLY FLEXIBLE INTERCONNECTS (MFIS) FOR INTERPOSER-ON-MOTHERBOARD ASSEMBLY

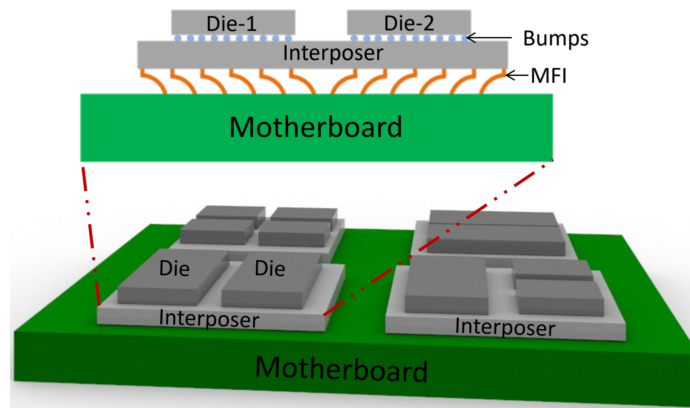


Figure 2.1: MFI-enabled large integrated system with interposer-on-motherboard

In this chapter, we perform a thermomechanical analysis of MFI-interposer assembly. The goal of this work is to extend the component-level optimization methodology presented in [63, 62, 61] to the optimization of interconnects in an assembled subsystem, which includes printed circuit board (PCB), a silicon interposer, and a large number of MFIs between the interposer and the motherboard, as shown in Fig. 2.1. First, we report an MFI distribution configuration to reduce MFI stress and also discusses a package-level optimization process. Moreover, a thermally-induced warpage comparison between solder bumps and MFIs in an interposer-to-motherboard assembled system is reported. Next, we describe the impact of chip size on thermo-mechanical warpage and stress. We show a comprehensive MFI distribution technique utilized for improved thermo-mechanical reliability. Finally, we investigate the impact of MFI pitch on thermo-mechanical reliability.

Table 2.1: Simulation setup

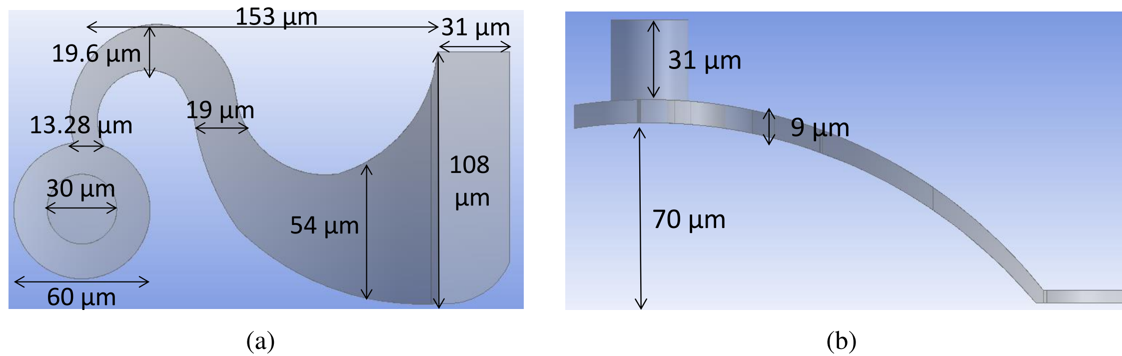
Parameters	
Interposer size	1 cm × 1 cm
Interposer thickness	100 μm
PCB thickness	1000 μm
Solder bump/ MFI height	110 μm
Solder/MFI pitch	400 μm

2.1 MFI Orientation and Package Level Optimization for Reduced Stress and Warpage

2.1.1 Simulation Specifications

The overall specifications of the test vehicle are specified in Table 2.1

MFI Configuration



(a) (b)

Figure 2.2: (a) top view, (b) side view of an MFI.

The overall dimensions of the baseline MFI are shown in Fig. 2.2. These dimensions are based on [59] and considered as the initial design for the analysis. NiW is chosen as the interconnect material because of its relatively high yield strength of 1930 MPa[91]. The MFIs are 9 μm thick and have a standoff height of 70 μm. The MFIs are permanently bonded, as is done in [92], to the interposer with a 30 μm diameter and 31 μm tall tip. As these interconnects are to be compared with solder bumps between the motherboard and the interposer, the total height for both the solder bump and the MFI is 110 μm. The purpose

of this study is to compare the thermo-mechanical performance of the MFI assembly to the solder bump assembly.

MFI Orientation

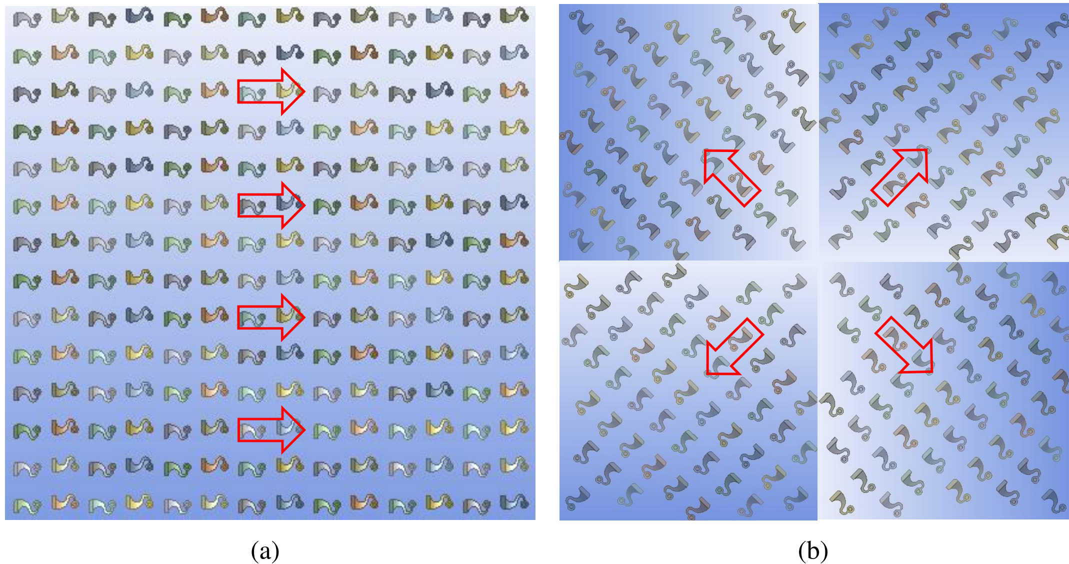


Figure 2.3: MFI distribution with (a) baseline orientation (b) radial orientation

Fig. 2.3 shows the overall MFI distribution for two different MFI orientations. In the baseline orientation design, as seen in Fig. 2.3(a), MFIs are distributed across the board in an array-like fashion. Different orientations were also implemented that consider the spring-like structure of the MFI. Specifically, since the MFI design under consideration has a greater in-plane compliance in the direction of its anchor-to-tip and since interposer warpage is largest along the substrate diagonal, aligning the MFIs from anchor-to-tip along the substrate diagonal may reduce the stress in the MFIs during warpage. Following this same logic, a radial orientation design, as shown in Fig. 2.3(b), is also implemented where the interposer is evenly broken into four symmetric sections. For each quarter, the MFIs are distributed along the direction of the substrate diagonal.

Because of the symmetry of this configuration, only one quarter of the whole assembly is considered for FEM simulations in ANSYS Workbench. Fig. 2.4 shows the overall

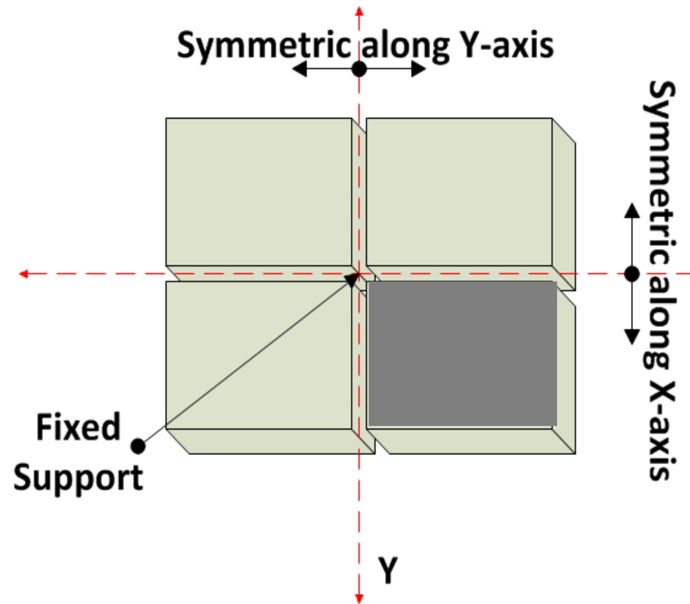
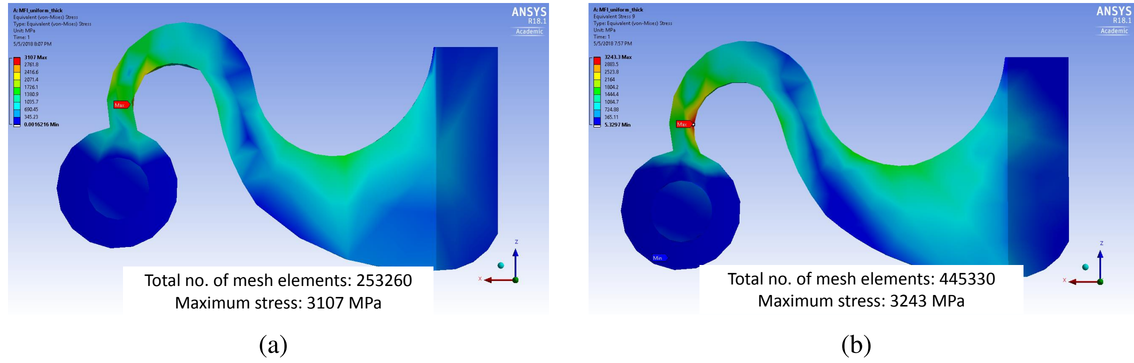


Figure 2.4: Boundary condition for the simulations. Only the interposer is shown. The MFIs (not shown in the figure) are on top of the interposer

boundary conditions. The point at the origin is considered fixed and the two orthogonal planes that pass through the origin are assumed to have no displacement in the direction normal to the corresponding plane (e.g. the X-Z plane in Fig. 2.4 has no displacement in the Y direction). In the figure, the red lines designate the two lateral axes of symmetry. The whole assembly is cooled down from 160°C to room temperature (25°C). The resulting steady-state warpage and stress are compared and analyzed.

2.1.2 Meshing Profile

The ANSYS built-in adaptive meshing mechanism is adopted to ensure high fidelity simulations are performed during the optimization processes. Fig. 2.5 shows von-Mises stress results from two loops of mesh refinement, where the mesh adaptively grows finer and finer for MFIs in the baseline orientation setup. As seen in Fig. 2.5, adaptively increasing the number of mesh elements to almost twice the number of mesh elements found in the initial mesh changes the max von Mises stress by less than 5 %.



(a) (b)

Figure 2.5: Worst case MFI stress distribution with adaptive meshing (a) Initial meshing and (b) denser meshing

Table 2.2: Material properties

Material	Young's Modulus (GPa)	CTE (ppm/^{o}C)
FR4	24	16
Silicon	130	2.7
SAC305	50	23.5
NiW	180	13
Copper	120	17.3

2.1.3 Thermally Induced Warpage and Stress Results

As a first step, the design incorporating MFI baseline orientation is compared with a solder bump assembly. In the latter case, the bumps are 170 μm in diameter, 110 μm tall, and 400 μm in pitch. The dimensions and pitch of the solder bumps are in accordance with [93] and the MFI dimensions noted in the previous section. Material properties used in the simulations are shown in Table 2.2. FEM simulation results for substrate warpage and MFI max von Mises stress are summarized in Fig. 2.6 for both MFI and solder bump configurations. In Fig. 2.6, each temperature data point is a standalone FEM simulation result, i.e., for each temperature data point in Fig. 2.6, the whole assembly is cooled down from the specified temperature to 25 ^{o}C. As expected, thermally induced warpage for solder bump assembly is much larger compared to that of the MFI assembly. Thermally-induced stress in the MFIs is also quantified. The worst-case stress increases monotonically with

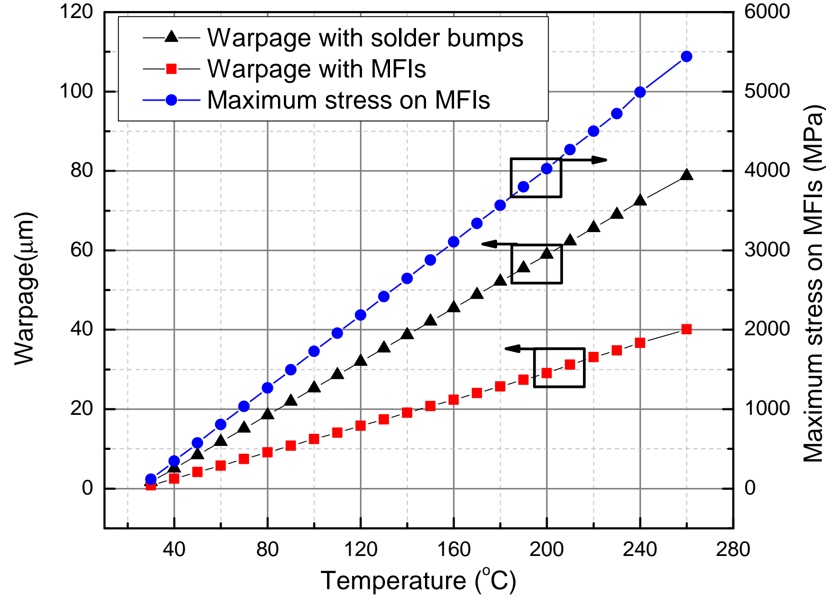


Figure 2.6: Thermally induced warpage and stress results for different configurations

increasing temperature. The maximum warpage is along the diagonal of the interposer and the maximum stress is in the MFI located at the farthest corner. At temperatures above 110°C, the maximum von Mises stress in the NiW MFI exceeds the yield strength of NiW and hence, plastic deformation occurs within the MFI. Given such important contributing factors, comprehensive design strategies such as the radial orientation described earlier are necessary to ensure that the MFIs will maintain reliable interconnections after being subjected to high temperature conditions.

2.1.4 Radially Oriented MFI Distribution

Fig. 2.7 shows the radially oriented MFI distribution scheme employed in the simulations. As shown in the figure, the maximum stress in the MFIs is 1717 MPa. Simply using this orientation method results in a 45% reduction in maximum von Mises stress (compared to the baseline orientation case). This result will be referred to as the 'unoptimized radial orientation' case for the rest of the chapter. As expected, the maximum MFI displacement

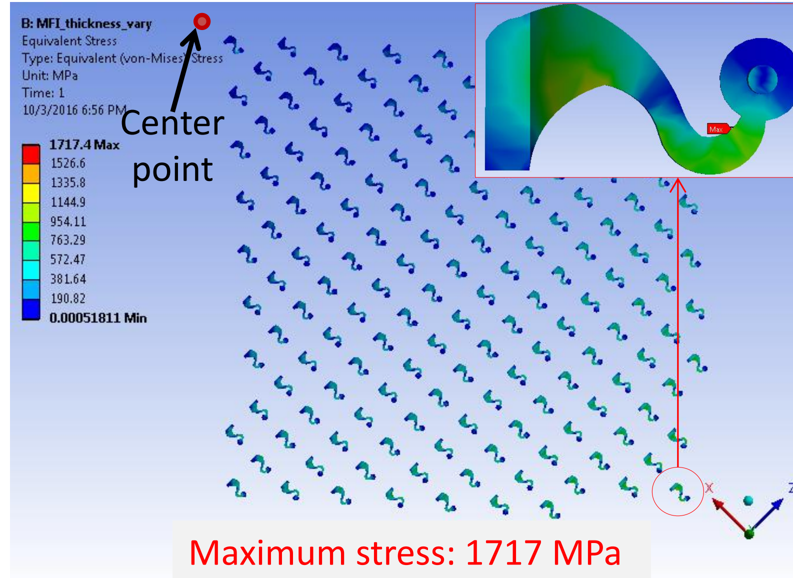


Figure 2.7: MFI distribution along the diagonal of the interposer

is along the diagonal. We also characterized the in-plane displacement of the MFIs, which is also called lateral displacement along the X and Y axes. The in-plane displacement of the MFIs is very minimal (0.04 μm).

2.1.5 System Level Optimization Methodology

Radial orientation improves the thermo-mechanical reliability of the system. To further improve this reliabilty, in this study, an interconnect optimization technique has been considered along with the different orientation methodologies. By engineering the MFI geometry, the stress distribution can be improved to reduce the overall maximum stress value. Since the warpage of an assembled system includes many different force vectors exerted upon each individual off-chip interconnect, it is difficult to simulate all of these external forces on a single MFI. Therefore, rather than optimizing a stand-alone MFI and attempting to simulate the correct environmental stimuli, we optimize the MFI geometry via modeling the entire assembled system. The general flow diagram for the methodology is specified in Fig. 3.1(b). A single MFI ('master') with all the parameter variables is placed on the board near the origin of the coordinate system (i.e. center of the board). A distributed

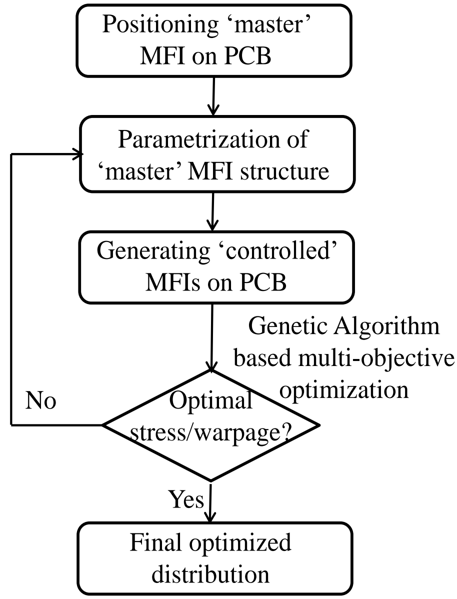


Figure 2.8: Flow diagram for optimization methodology

array of MFIs is generated based on the master MFI design and distributed according to the technique described in the previous section. A Genetic Algorithm(GA) based optimization tool from ANSYS is utilized for the optimization process [61]. Permanently-bonded interconnects have many factors to consider, e.g. substrate material, interconnect materials, substrate thickness, size of the chip, temperature cycle, in-plane forces/displacements, and out-of-plane forces/displacements. In this chapter, either maximum interposer warpage, maximum MFI stress, or both are taken into account as optimization objectives. The optimization methodology begins with a parametrization of our initial design, as seen in Fig. 2.9(a). This parametrization includes different widths, lengths, and radii of certain regions of the MFI body. These parameters are used as inputs to our optimization problem. A variation of these parameter inputs, which effectively modifies the geometry, results in different outputs that we wish to optimize. In this case, these outputs include the maximum von Mises stress of the worst-case MFI and the maximum warpage deformation of the die/interposer/package. After selecting the lower and upper limits for the input and output objectives, a design of experiment (DOE) is developed that attempts to thoroughly explore the design space to form a strong basis upon which the optimization process builds from.

Specifically, an optimal space filling DOE is used. Finally, a direct optimization process is employed that simultaneously minimizes both the maximum von Mises stress in the MFIs and warpage in the interposer. Ultimately, the optimization process results in many Pareto-optimal solutions from which one among them is chosen according to the objectives that we prioritize (minimizing warpage over minimizing MFI stress for example). After running the process for multiple generations, the optimization process tends to preferentially select better designs that fulfill system-level parameter objectives, ultimately converging to optimized MFI designs. A number of different parameters can be incorporated into the optimization process, e.g., electromigration, electrical behavior, creep analysis [94], etc. This chapter presents a methodology that has been adopted to address stress/warpage considerations. However, this methodology can be applied to other objectives as well.

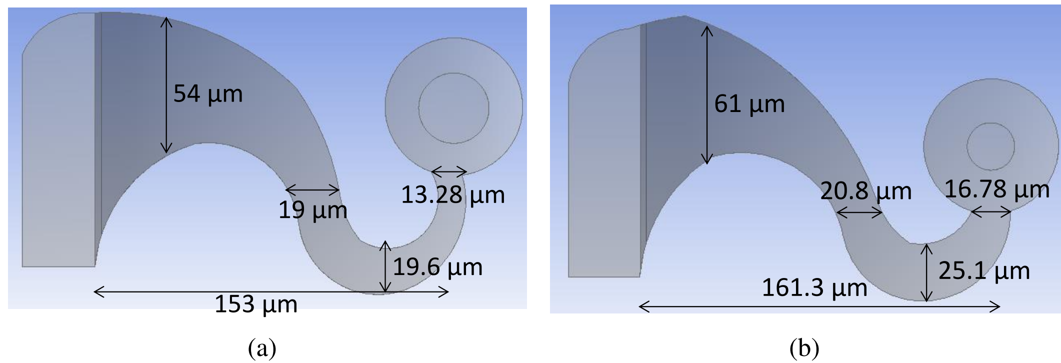


Figure 2.9: (a) Initial design and, (b) Optimized design of MFI

Fig. 2.9(b) shows the optimized MFI structure after the optimization process. As seen in the figure, the optimized geometry of the MFI has changed during the optimization process, hence leading to an interconnect structure that is more mechanically robust and reduces the interposer warpage. The maximum von-Mises stress of this optimized worst-case MFI is shown in Fig. 2.10. As seen, the maximum stress is along the neck of the MFI, which is also the case for the initial MFI. It is likely that the optimized design distributes stress more evenly in this region compared to its predecessor, which is, in part, why maximum stress is decreased. The stress is reduced to 1511 MPa from an initial stress value

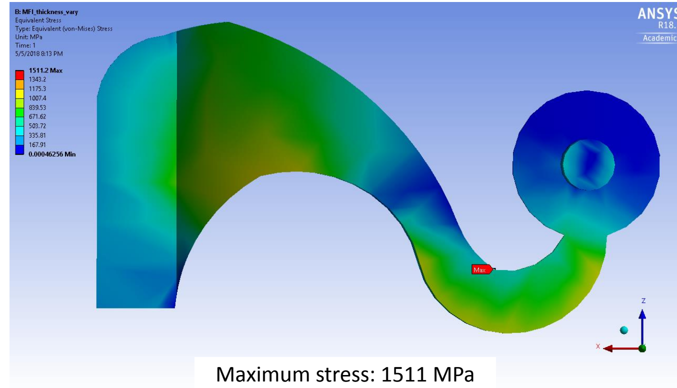


Figure 2.10: Stress distribution in worst case MFI for optimized design

of 1717 MPa in the unoptimized radial orientation case (Fig. 2.7), a 12.0% reduction in maximum stress. Compared to the baseline orientation case (Fig. 2.6), this improvement translates to a 51.3% decrease in maximum stress.

2.2 Impact of Interposer Size on Thermo-Mechanical Reliability

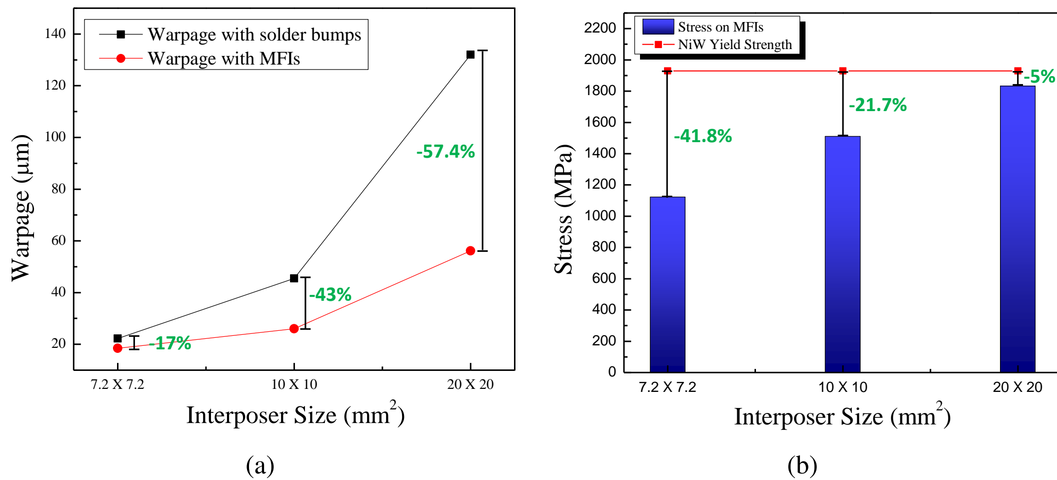


Figure 2.11: For different interposer sizes, (a) warpage comparison between different configurations and, (b) worst case MFI stress

The simulation setup is applied to interposers with different sizes maintaining the same thickness of 100 μm to investigate the reliability issues resulting from interposer size variation. As is evident from the previous sections, stress can be minimized using different

orientations and optimization methods. In this case, a comparison study is performed between the MFI assembly and the solder bump assembly. Along with the $10\,mm \times 10\,mm$ interposer, smaller ($7.2\,mm \times 7.2\,mm$) and larger ($20\,mm \times 20\,mm$) cases have been used to emulate different applications. In each of the cases described above, the MFI and solder bump pitches are maintained at 400 μm. The MFIs are distributed using the radial distribution methodology described in the previous section. According to Fig. 2.11(a), as the interposer size increases, the reduction in warpage using MFIs becomes more prominent. For the smallest interposer case, 17% reduction in warpage is seen whereas for the largest interposer, 57.4% reduction is observed. Specifically, in the solder bump assembly, the warpage is $132\,\mu$m, which is more than twice the requirement set by ITRS. On the other hand, MFI assembly results in $56.18\,\mu$m deformation, which is lower than the limit. The latter case can be further improved, if necessary, by running a warpage-specific optimization, which we ignore here. For each of the interposer size cases, worst case MFI stress analysis has also been carried out in Fig. 2.11(b). The gap between maximum stress and NiW yield strength is defined as 'stress headroom'. As interposer size increases, the overall displacement along the diagonal also increases, which results in higher stress. Hence, for larger interposers, the stress headroom is expected to be smaller. However, the maximum stress remains below the limit specified by the yield strength.

2.3 MFI Orientation (Radial) Along the Thermal Expansion/ Contraction Contour

In the design with radially distributed MFIs, the MFIs are oriented along the diagonal of the interposer. With only isotropic thermal expansion, this is sufficient to improve the thermo-mechanical performance of the system. However, in more pragmatic considerations, the expansions are mostly anisotropic. This expansion can be decomposed into planar expansion and vertical deformation. Because of the spring like behavior of the specified initial cases, the MFIs are better suited to handle the deformations along their axis (i.e. parallel to the vector connecting the anchor and the tip). In the radial orientation case, the overall

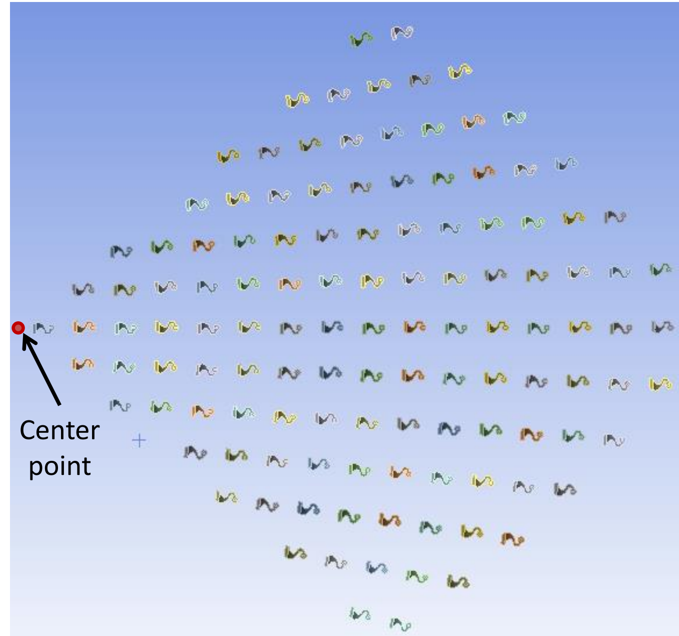
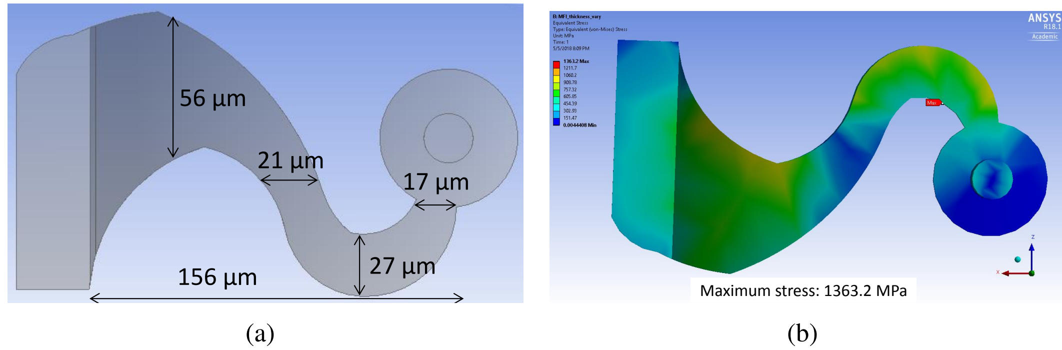


Figure 2.12: MFI distribution following the deformation contours

stress will decrease; but the worst case MFI may no longer be located along the interposer diagonal. As a result, these off-diagonal MFIs may still experience larger in-plane displacement which contributes to additional stress. One of the solutions to this problem is to design MFIs for better in-plane compliance for all directions at the cost of worse out-of-plane compliance. A comprehensive approach has been taken in this analysis to exploit the full advantage of the higher out-of-plane compliance of the MFIs. The thermal expansion or contraction patterns of the interposer are taken into account to better distribute the MFIs. The scheme is carried out and described in Fig. 2.12.

The optimized MFI structure is shown in Fig. 2.13(a). The basic parameter variables that have been modified by the optimization process are specified in the figure as well. Fig. 2.13(b) shows the overall stress distribution of the worst case MFI, which is located along the diagonal of the interposer. As can be seen from the figure, the maximum stress is further reduced to 1363.2 MPa, representing a 10.25% improvement compared to the optimized radial orientation case (Fig. 2.7) and a 56.3% improvement compared to the baseline design. As described earlier, the optimization process can be carried out to optimize either stress



(a) (b)

Figure 2.13: (a) Optimized MFI structure and, (b) worst case MFI stress distribution

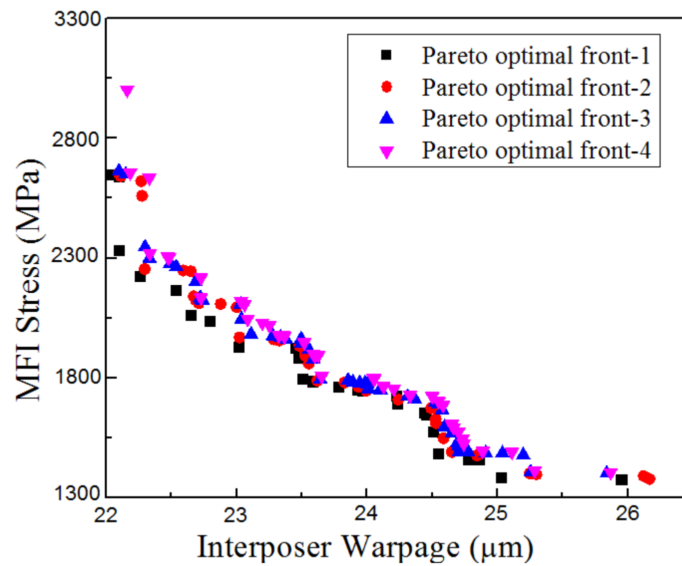


Figure 2.14: Interposer maximum warpage and MFI maximum stress tradeoff

or warpage or both. Accordingly, a multi-objective optimization process was attempted for minimizing both stress and warpage. Fig. 2.14 shows the overall tradeoff trend between warpage and stress minimization. It is hypothesized that the more compliant interconnects tend to mitigate warpage since they can more easily move in the necessary directions to effectively transfer the strain from the interposer to the MFIs. This strain transfer has the effect of lowering the interposer warpage while increasing the stress in the MFIs. From the final Pareto front observed in Fig. 2.14 (Pareto-optimal front-1), the designer can choose which Pareto-optimal solutions best fit their scenario. In this case, the chosen designs are the solutions which minimize interposer warpage without inducing plastic deformation in

Table 2.3: Interposer warpage and worst-case MFI stress for different MFI orientations

Parameters	Baseline MFI Orientation	Radially Distributed MFIs (optimized)	Modified Radial Distribution of MFIs (stress-focused optimization)	Modified Radial Distribution of MFIs (warpage-focused optimization)
Interposer Size	10 mm × 10 mm	10 mm × 10 mm	10 mm × 10 mm	10 mm × 10 mm
MFI Pitch	400 μm	400 μm	400 μm	400 μm
Interposer Warpage with MFIs	22.4μm	26μm	25μm	24.217μm
Warpage reduction with MFIs compared to solder bump case	50.68%	42.75%	44.96%	46.7%
Maximum MFI Stress	3107MPa	1511MPa	1363MPa	1479MPa
Minimum MFI Stress headroom relative to NiW yield strength	-61%	21.7%	29.37%	23.37%

the NiW MFIs. Since we are assuming a NiW yield strength of 1930 MPa, the MFI structure with the highest stress that is below 1930 MPa has been selected, minimizing interposer warpage while avoiding any plastic deformation of the interconnects. This chosen design is seen in Fig. 2.14 and reported below. Although the yield strength of NiW has been used as the cut-off criteria in choosing optimized MFIs, it might very well be acceptable to choose interconnects that will plastically deform, and perhaps this might even be preferred since it would lower the interposer warpage even further. For simplicity however, plastic deformation has been neglected in this study.

Both warpage-centric and stress-centric optimization have been performed. From the warpage-centric run, maximum MFI stress increases minimally but the interposer warpage

decreases. All results from the relevant FEM simulations are summarized in Table 2.3 and obtained from Fig. 2.14. For simplicity, only the $10\ mm \times 10\ mm$ interposer cases are shown in the table. These results follow the tradeoff pattern that was outlined in Fig. 2.14. With respect to all MFI designs (including optimized designs) and configurations/orientations, the minimum improvement among these relative to the solder bump assembly case, is 42.75%. After optimization, the minimum reduction in MFI maximum stress compared to the baseline design is 51.3%.

2.4 Impact of MFI Pitch on Warpage and Stress

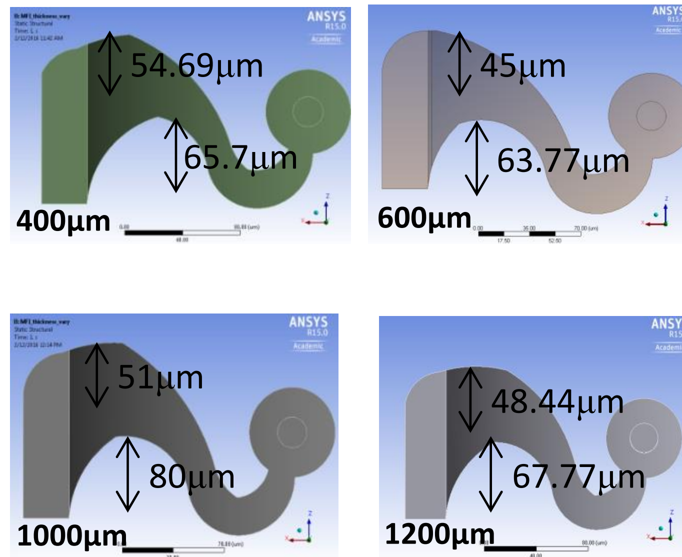


Figure 2.15: Optimized MFI shape for different pitches

Throughout the chapter, an MFI pitch of 400 μm is evaluated. However, conventional board level solder bump pitches can be greater than 1 mm. In this section, different MFI pitches have been investigated as part of a sensitivity analysis to determine the impact of increasing MFI pitch from a thermo-mechanical reliability point of view. Four different pitches of 400 μm, 600 μm, 800 μm and 1200 μm are simulated. For each case, a radial MFI distribution following the expansion/contraction contours is used. Due to the fixed size of the interposer, the number of MFIs varies for each case. The MFI optimization

process has been run for each individual case. Fig. 2.15 describes the shape of the MFIs for different cases showing some parameter variations.

Fig. 2.16 shows the overall FEM simulation results for different pitches. As MFI pitch

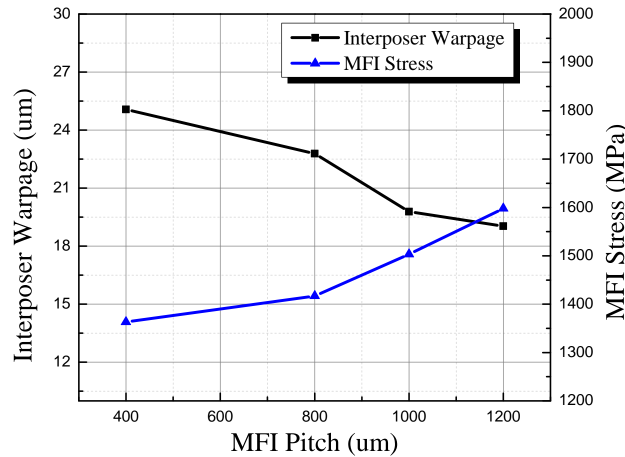


Figure 2.16: Optimized MFI warpage and stress for different pitches

increases, the number of fixed connections between the interposer and motherboard decreases. This eventually reduces the overall warpage of the interposer, but thermal load per MFI increases as well, resulting in higher maximum stress in the MFIs. Despite the additional stress, the maximum stress among all four cases is 17.2% less than the NiW yield strength.

2.5 Conclusion

In this chapter, we present a thermomechanical analysis of MFI-interposer assembly where the interposer is directly assembled on the motherboard. We compare the results with a conventional solder bump based assembly. For the MFI assembly, a minimum of 42.75% improvement in warpage is shown. This impact becomes larger as interposer size increases. Only permanently-bonded MFIs have been considered for the analyses. Rematable contact MFIs (e.g., not permanently-bonded) would reduce interposer warpage further. Simply changing the orientation of the MFIs along the diagonal of the interposer results in a 51.3%

improvement in MFI stress compared to the baseline MFI orientation case. Both stress-centric and warpage-centric optimization have been investigated and a tradeoff analysis has been performed. An MFI distribution technique has been employed following the interposer expansion/contraction contour, which further reduces the stress on the interconnects. This orientation also reduces interposer warpage. Finally, a sensitivity study of the MFIs has been performed to investigate the impact of different MFI pitches on thermo-mechanical performance.

CHAPTER 3
POWER DELIVERY NETWORK MODELING FOR EMERGING HETEROGENEOUS INTEGRATION TECHNOLOGIES AND DESIGN SPACE EXPLORATION OF POWER DELIVERY INCLUDING VOLTAGE REGULATOR MODULES

There is an increasing interest in heterogeneous integration of multi-functional dice into a single package. These high-performance integrated modules inevitably lead to higher current demand and increased power density [7] despite the down-scaling of supply voltage in recent device technologies [69]. As a result, power delivery in high-performance digital systems is an increasingly difficult challenge [70]. Moreover, in order to maintain, if not improve, the performance of heterogeneously integrated dice compared to monolithic integration, one must carefully consider the interconnect channels in emerging heterogeneous integration platforms. Before taking full advantage of emerging 2.5-D and 3-D integration technologies, we must first understand and address the challenges of the power delivery network (PDN) and power supply noise (PSN). Specifically, 2.5-D integrated electronics have several unique attributes that require modeling and benchmarking. For example, in a silicon interposer based 2.5-D integration [95], using a PDN grid on the interposer will enhance current spreading; however, overall impedance of the interposer PDN may increase if the parasitics of the TSVs and microbumps are large enough to offset the resistance decrease of the PDN grid. Likewise, for Embedded Multi-die Interconnect Bridge (EMIB) or similar bridge-chip based technologies, signal interconnections and driver circuits are placed, generally, on the peripherals of the dice and above the die-to-die interconnect carrier (i.e., bridge-chip), which may lead to less power/ground (P/G) C4 bumps that are connected to the package-level power/ground planes. Therefore, in this chapter, a power delivery network modeling framework is presented. Several emerging 2.5-D integration technologies

are benchmarked for power delivery. Moreover, a design space exploration of power delivery with VRM placement study is reported. Section 3.1 to 3.3 are based on the work reported in [5, 96]; this methodology is the foundation of all the PDN research reported in this thesis.

3.1 Modeling Methodology

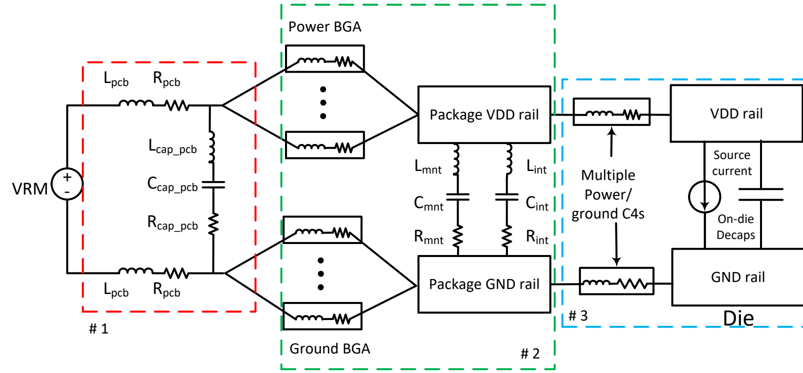
Fig. 3.1(a) shows the PDN structure of an IC. Unlike most prior work [70, 83, 81] that utilizes a lumped package model, we implement a distributed package-level PDN model to reflect the spreading effects of current in the package and the coupling between different P/G bumps. This is critical in multi-die packaged systems in which dice share the package-level PDN planes. Fig. 3.1(b) presents the flow diagram for different analysis types: steady-state IR-drop analysis and simultaneous switching noise based transient analysis. The analysis begins with the generation of the RLC network models of the board, package, and on-die PDNs. Subsequently, these models are combined to solve for nodal voltages and branch currents. Each step is detailed in the following subsections.

3.1.1 Board-Level PDN

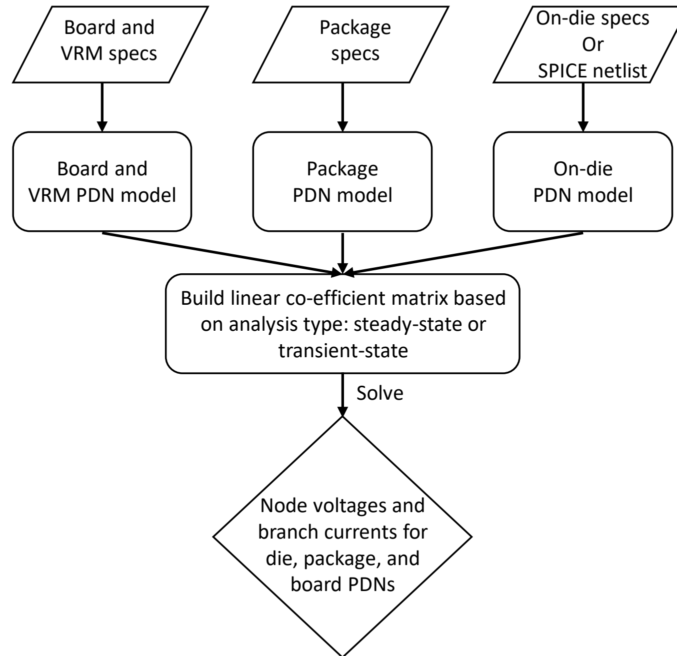
In this model, we do not explicitly model the VRM; instead, we assume an ideal Voltage Regulator Module (VRM) that is supplying a stable voltage and use a lumped resistor/inductor network to model the board-level current spreading. Moreover, the equivalent series resistance (ESR) and inductance (ESL) of the board-level decoupling capacitors are included in the model.

3.1.2 Package-Level PDN

Fig. 3.2 shows the detailed distributed package-level PDN model. The package power/ground planes are modeled as two layers, where the bottom layer is connected to the motherboard by BGAs, and the top layer is connected to an on-die PDN by C4 bumps. Each node in the



(a)



(b)

Figure 3.1: (a) The PDN modeling hierarchy. From left to right: lumped model of the board-level PDN, distributed model of the package-level PDN, and the distributed model of the on-chip PDN. (b) Flow diagram of the PDN analysis showing different steps of the framework.

two layers is connected to six adjacent nodes using a resistor-inductor pair either due to the package traces or inter-layer vias. It is assumed that the surface mounted decaps are only connected to the top layer in the designated areas.

Each R_{sp} and L_{sp} pair in the distributed model represents the current spreading effects,

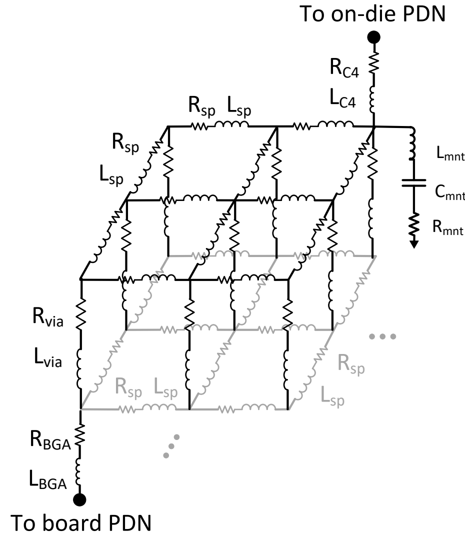


Figure 3.2: The two-layer (two power and two ground) package PDN model of power/ground planes

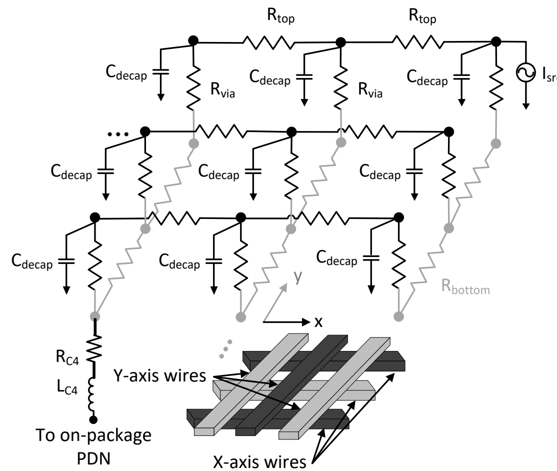


Figure 3.3: The on-die PDN model. Only one current source and one C4 bump is shown.

while each R_{mnt}, C_{mnt} and L_{mnt} set of values represent a surface mounted decap, as shown in Fig. 3.2. For bump inductance, L_{C4}, we consider both self and mutual inductances, where the mutual inductances are assumed to be dominated by the nearest 8 neighbors [97].

3.1.3 On-Die PDN

On-die PDN consists of several metal layers, where the power/ground wires are parallel to each other in each layer, but each layer is orthogonal to the layer below/above it (interleaved

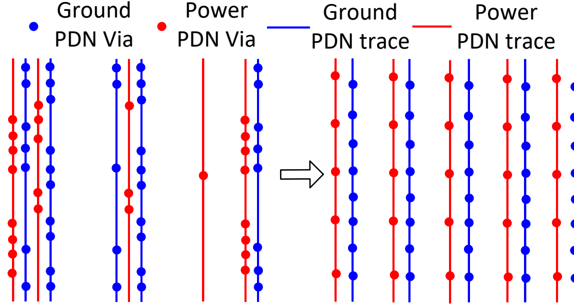
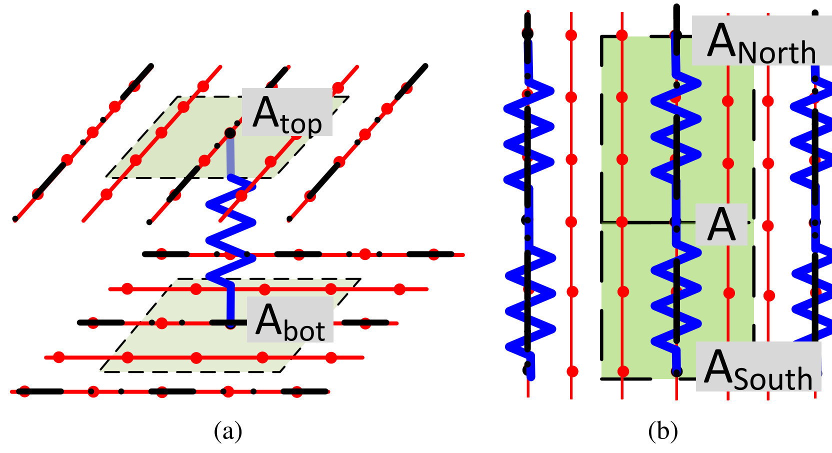


Figure 3.4: Re-organization of a non-uniform PDN layout



(a) (b)

Figure 3.5: Map fine-grained power PDN layout to coarse meshing grids (a) vias (b) wires.

structure, as shown in the inset of Fig. 3.3). Prior work has proposed a virtual PDN mesh design using C4 bump granularity with only one metal layer [70, 83, 81]. However, to better reflect the nature of the interleaved PDN design as well as the impact of on-die vias, we model the on-die PDN as a two-layer structure, as shown in Fig. 3.3. The resistance of R_{top}, R_{bottom} and R_{via} can be extracted from the design layout using the process described below.

For each layer within the on-die PDN, the metal wires and vias are typically uniformly distributed[81]. If the actual layout is non-uniform, we can calculate the effective wire pitch and via density and re-organize the PDN layout[81], as shown in Fig. 3.4.

Next, for each layer, we map the fine-granularity PDN layout to coarse mesh grids, which are in C4 bump granularity. Fig. 3.5(a) and 3.5(b) illustrate the mapping proce-

dure. For each coarse grid containing multiple vias and metal wires, the equivalent parallel resistance is calculated and assigned using the models described in [81].

Finally, all coarse PDN layers with X-axis metal wires are mapped onto the top layer, and all Y-axis metal wires are mapped onto the bottom layer, as shown in Fig. 3.3. R_{via} in Fig. 3.3 is the effective resistances of vias between adjacent metal layers. Likewise, R_{top} and R_{bottom} are the total parallel resistances between adjacent nodes in all layers with X-axis and Y-axis wires, respectively.

3.1.4 PDN Analysis Formulation

The supply voltage noise formulation is shown as follows:

$$\begin{bmatrix} G & A_L \\ -A_L & R \end{bmatrix} \cdot \begin{bmatrix} V(t) \\ I(t) \end{bmatrix} + \begin{bmatrix} C & 0 \\ 0 & L \end{bmatrix} \cdot \begin{bmatrix} \dot{V}(t) \\ \dot{I}(t) \end{bmatrix} = \begin{bmatrix} i_s(t) \\ 0 \end{bmatrix} \tag{3.1}$$

where G is the PDN grid conductance matrix; A_L represent the coefficients of branch current $I(t)$ in Kirchhoff's voltage and current equations, respectively. C and L are matrices reflecting the capacitive and inductive elements, respectively; $i_s(t)$ is the source current.

For steady-state analysis, the time-varying terms are omitted and the branch current $I(t)$ can be expressed in the form of $V(t)$. Hence, all the branch currents other than the source currents will be converted to a nodal voltage based representation. Thus, A_L will be merged with G in 3.1. Eq. 3.1 is then derived in the form of $G \cdot V(t) = i_s(t)$, where matrix G is positive symmetric definite. Therefore, the above linear equation can be solved using the Choleskey factorization method.

For transient analysis, the trapezoid difference scheme can be used to formulate Eq. 3.1, as shown below:

$$(\frac{K}{\Delta t} + \frac{U}{2}) \cdot X^{n+1} = (\frac{K}{\Delta t} - \frac{U}{2}) \cdot X^n + \frac{{I_s}^{n+1} + {I_s}^n}{2} \tag{3.2}$$

Table 3.1: Validation results (modeling vs. open source benchmarks)

Circuits (# of Nodes)	Metal Layers	Bump Current Error (%)	Max IR-Drop Error (%)	Transient Error (%VDD)
IBM1 (31 K)	2	21.75	20.29	1.84
IBM2 (127 K)	4	7.14	11.11	0.67
IBM3 (852 K)	5	3.59	2.21	0.54
IBM4 (954 K)	6	7.60	0.71	0.12
IBM5 (1.08 M)	3	6.12	3.03	0.22
IBM6 (1.67 M)	3	7.29	1.23	0.22
IBM7 (1.46 M)	6	5.34	5.71	N/A
IBM8 (1.46 M)	6	5.34	5.71	N/A

where

$$U = \begin{bmatrix} G & A_L \\ -A_L & R \end{bmatrix} \quad K = \begin{bmatrix} C & 0 \\ 0 & L \end{bmatrix}$$
$$X = \begin{bmatrix} V \\ I \end{bmatrix} \quad I_s = \begin{bmatrix} i_s \\ 0 \end{bmatrix} \tag{3.3}$$

To accelerate the simulations, we fix Δt which would eventually make $\frac{K}{\Delta t} + \frac{U}{2}$ a constant coefficient matrix. Therefore, we pre-factorize this matrix before transient simulations using LU factorization. In the solving steps, the triangular factors can be used to solve the linear equations efficiently. The framework is implemented using MATLAB because of the necessity for dense matrix operations and scientific computations.

3.2 Validation

To validate the PDN framework, open-source *IBM* power grid benchmarks [98] are used. The benchmarks are provided in the *HSPICE* netlist format. There are eight benchmarks

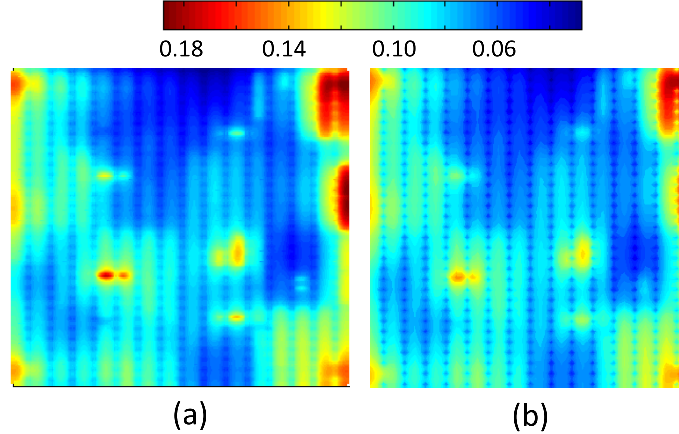


Figure 3.6: The noise profile of IBM3 benchmark (a) results from open-source *IBM* PG benchmarks and (b) our modeling results.

for steady-state analysis and six benchmarks for transient analysis. For steady-state results, the benchmarks provide the overall noise profile including the noise level of each node. On the other hand, for transient results, the benchmarks provide the waveforms of 20 randomly selected nodes throughout the whole circuit. The benchmark size and the number of metal layers are summarized in the first two columns of Table 3.1.

We use scripts to extract the layout and RLC information and then we map the PDN layout onto the coarse mesh grids at the granularity of C4 bump pitch. Next, we solve for the supply voltage noise in both steady-state and transient-state using the above mentioned framework. We compare the modeling results to the IBM open-source data using three sets of metrics: current of each C4 bump, IR-drop of each node, and transient noise of all the 20 randomly selected nodes.

3.2.1 Steady-State Results

The steady-state validation results are summarized in the third and fourth columns of Table 3.1. Except for the small benchmark cases IBM1 and IBM2, which have highly non-uniform PDN structure, all cases obtain maximum relative errors of less than 7.60% and 5.71% in bump current and IR-drop, respectively. The noise profiles are also compared and the results are well matched. Fig. 3.6 shows an example of the noise profile comparison of

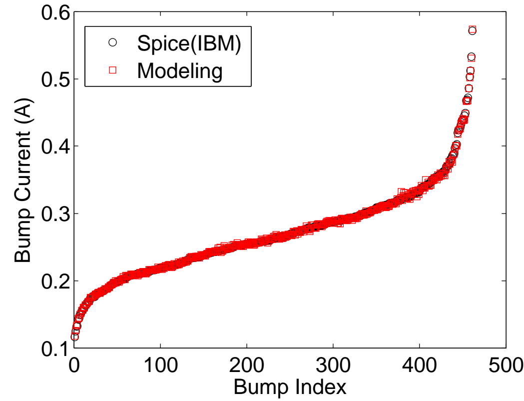


Figure 3.7: Bump current comparison for IBM3.

IBM3 as this case has the largest noise gradient. The model accurately captures the distribution of the noise. Fig. 3.7 shows the bump current comparison of IBM3 in which we sort the current of each bump in an ascending order and plot both *IBM* provided and our modeling results [81]. Likewise, although the current value spans a wide scale (approximately 5X), the bump current is very well matched.

3.2.2 Transient-State Results

Transient validation results are summarized in the last column of Table 3.1. We normalize the error to supply voltage because some of the benchmark noise values are small and thus, the relative error can be high. Except for IBM1, the maximum error for all cases is less than 0.67% VDD. Fig. 3.8 shows the node with the maximum error for IBM2. Even for this case, the peak noise and waveform are well captured.

3.3 PDN Evaluation of Emerging Heterogeneous Integration Platforms

In this section, we use the above PDN framework to evaluate and compare different heterogeneous integration approaches, as shown in Fig. 3.9. The PDN design challenges for 2.5-D integration are investigated and summarized.

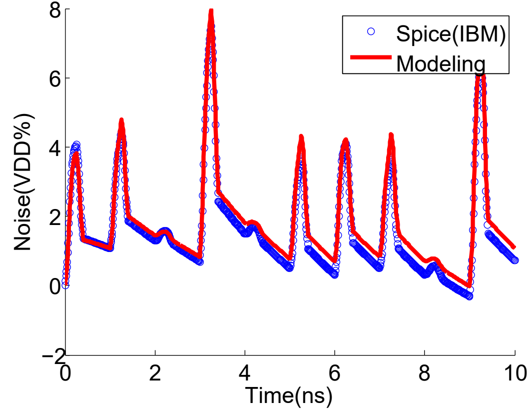


Figure 3.8: The transient noise of the node with the maximum error for IBM2.

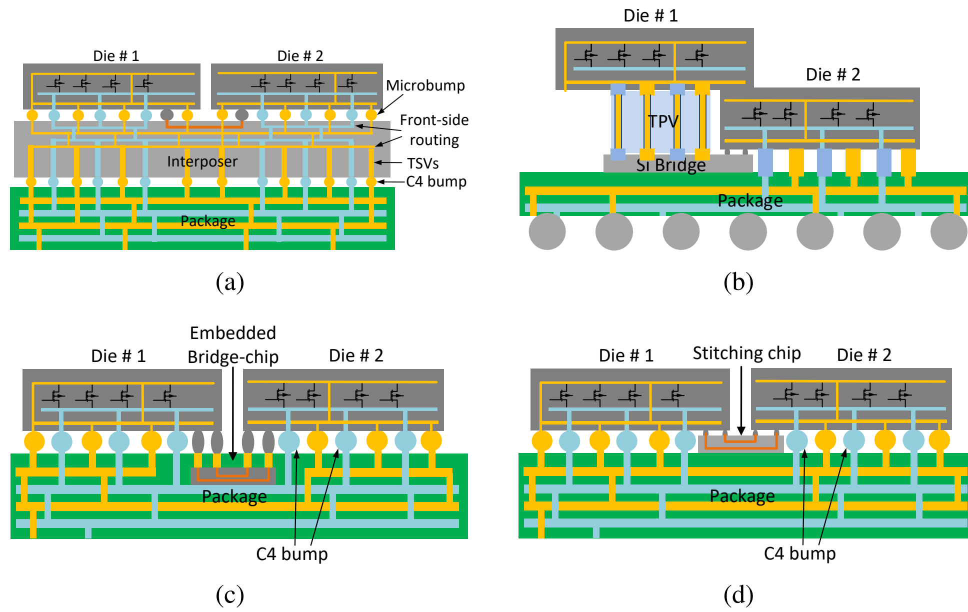


Figure 3.9: Various heterogeneous integration platforms (a) interposer, (b) bridge-chip within fan-out, (c) EMIB, and (d) HIST

3.3.1 2.5-D/3-D Integration Scenarios

Fig. 3.9 shows various heterogeneous integration technologies with different approaches for chip-to-chip interconnections. The first approach utilizes silicon interposer technology. In our study, we assume that the interposer contains front-side PDN routing that is interconnected to uniformly distributed fine-pitch microbumps [99]. Fine-pitch and bundled TSVs in the interposer are used to connect the interposer-PDN to the landing pads of the

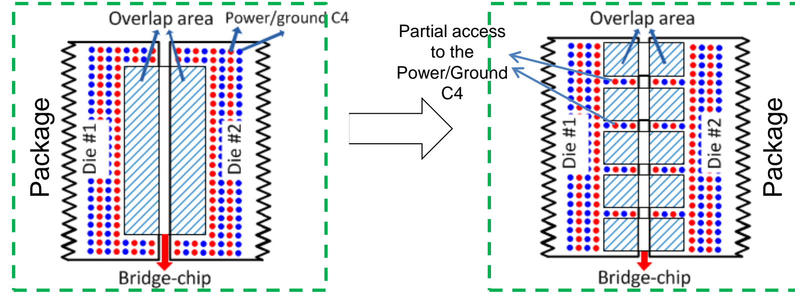


Figure 3.10: Illustration of bridge-chip placement (a) a single large bridge-chip (b) five small bridge-chips.

C4 bumps. By varying the number of TSVs, we can evaluate the best and worst interposer scenarios.

The second approach, HIST [27], is based on placing 'stitch' chips above the package substrate between the active dice to route high-density chip-to-chip interconnects. Another approach is EMIB technology as described in [26], which utilizes embedded silicon chips within the package to route the chip-to-chip interconnects. In imec's bridge-chip concept [6], a bottom die uses a bridge-chip and through silicon vias (TSVs) to communicate with an upper die within a fan-out package. As shown in [5], for these bridge-chip based technologies, if no through vias are used, this limits the number of bumps that are connected to the package-level power/ground planes especially at the edges. As a result, the PSN in those regions is impacted. Under this assumption, our results show that the PDN modeling of all these approaches as relatively comparable, and thus, we refer to these approaches as 'bridge-chip' for the remainder of the chapter. From [5], multiple smaller bridge-chips can reduce the PSN, and therefore, in this chapter, we assume a single large bridge-chip configuration as the worst case and five smaller bridge-chips with the same aggregate area as the optimal case, as shown in Fig. 3.10.

3.3.2 Design Parameters and Specifications

The modeling framework under consideration may be used to model any heterogeneously integrated microsystem, including co-integrated processor-memory and processor-accelerator

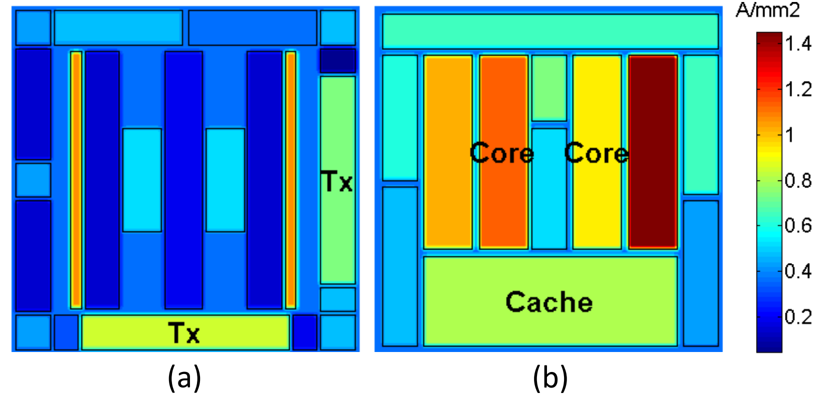


Figure 3.11: The current density of each die. (a) die #1 (b) die #2

modules. In this study, we emulate a field-programmable gate array (FPGA)-processor 2.5-D integrated package. In the two-die package, Die #1 emulates a 14 nm FPGA die and is assumed to have a peak total current of 49.78 A [100, 101]. The emulated FPGA power map is based on *Intel* Stratix 10 FPGA [102]. Die #2 emulates a 22 nm processor with a peak total current of 82.77 A. The current density maps are shown in Fig. 3.11. The supply voltage is assumed to be 0.9 V [101, 69].

Both dice are assumed to be 1 cm $\times$ 1 cm, and the package is assumed to be 2.45 cm $\times$ 1.8 cm. The two dice are placed side-by-side with a die spacing of 0.5 mm. The bridge-chip has a total area of 1.5 mm $\times$ 6 mm and the total overlap area with each die is assumed to be 0.5 mm $\times$ 6 mm (I/O area), as shown in the shaded region of Fig. 3.10. Table. 3.2 summarizes the parameters used in the PDN simulations. Since the FPGA and processor dice may have different supply voltages, they are assumed to have separate power delivery domains in each package layer and the PDN area in the package is equally assigned for simplification.

Moreover, as a reference to the best achievable results for bridge-chip and interposer cases, we consider an ideal baseline case where the two dice are assumed to be bonded to the package without an interposer or bridge-chips. This baseline is referred to as 'standalone' case. For all cases, we assume the packages are the same and utilize a C4 bump pitch of $200\ \mu m$. For the interposer case, the microbump pitch is 40 μm. Moreover, we

Table 3.2: Parameters of the PDN model

Parameter	value
On-die metal resistivity ($\Omega \cdot m$)	1.8e-8
On-die global wire **P**itch/**W**idth/**T**hickness (μm)	39.5/17.5/7
On-die intermediate wire P/W/T (nm)	560/280/506
On-die local wire P/W/T (nm)	160/80/144
on-die decap density (nF/mm^2) microbump pitch/R/L (μm/$m\Omega$/pH)	335 40/30.9/11.1
C4 bump pitch/R/L (μm/$m\Omega$/pH)	200/14.3/11.0
Package effective decap R/L/C (($m\Omega$/pH/μF)	541.5/220.7/52
Package resistivity/inductance ($m\Omega/mm$/pH/mm)	1.2/24
BGA pitch/R/L (μm/$m\Omega$/pH)	500/38/46
TSV R/L ($m\Omega$/pH)	54.2/77.78
PCB R/L ($\mu\Omega$/pH)	166/21
PCB Decap R/L/C (($\mu\Omega$/nH/μF)	166/19.54/240

assume the interposer worst case scenario to be when only one TSV is used per C4 bump while the best case scenario to be when 25 TSVs are used per C4 bump.

3.3.3 Benchmarking

IR-Drop

IR-drop profiles of each case are shown in Fig. 3.12. For the interposer case, even with additional fine-pitch P/G grids in the interposer, whether the PSN is improved relative to the standalone case depends on how many TSVs are used. This is because while the fine-pitch P/G grid and microbumps cause the current to spread more uniformly, the addition of TSVs may effectively increase the total PDN impedance. With only one TSV per C4 bump, the interposer case has a 6.27% and 7.79% larger IR-drop compared to the standalone case for Die #1 and Die #2, respectively. However, with 25 TSVs per C4 bump, the IR-drop is approximate 3.42% (Die #1) and 4.44% (Die #2) smaller than the standalone case. Moreover, the IR-drop distribution is more uniform than the case with only one TSV per C4 case. Nevertheless, interposer with fine-pitch TSVs will have higher fabrication costs, signal integrity challenges for high speed ICs and mechanical reliability challenges [26],

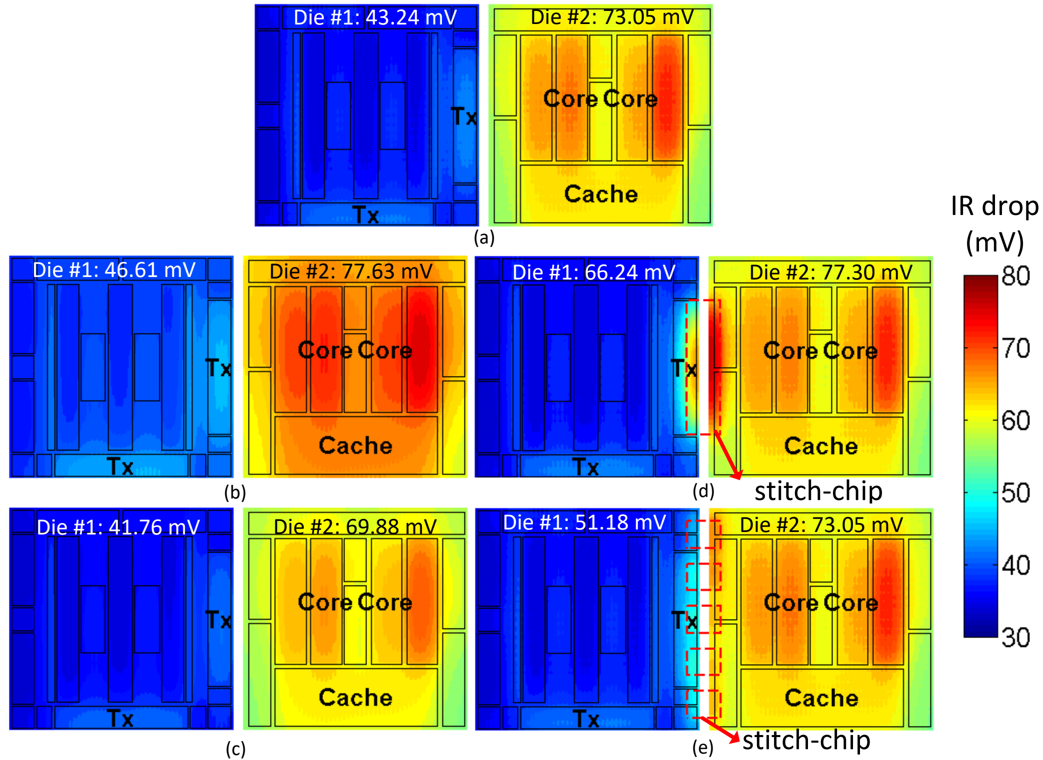


Figure 3.12: The IR-drop profiles of both dice for (a) standalone, (b) interposer case with one TSV per C4 bump, (c) interposer case with 25 TSVs per C4 bump, (d) single bridge-chip, and (e) five bridge-chips.

which make bundled-TSVs per C4 bump potentially difficult to use in practice.

For the bridge-chip cases, compared to the standalone and interposer cases, the additional noise is mainly due to the absence of C4 bumps in the regions overlapping with the bridge-chips. The IR-drop is approximately 53.2% (Die #1) and 5.8% (Die #2) larger than the standalone case. When five bridge chips are used as shown in Fig 3.12(e), there are no PSN hotspots at the edges of Die #2, and the maximum IR-drop of the system is the same as the standalone case. Similar to the interposer case with dense TSVs, there is also a larger manufacturing complexity to using multiple bridge-chips instead of a single large bridge-chip due to the requirement for multiple high-accuracy alignment assembly steps.

When comparing the interposer and bridge-chip technologies from a PDN perspective, it is hard to come up with a fair criterion since both are affected by multiple parameters. For the interposer case, besides the parasitics of the TSVs, the microbump pitch also plays

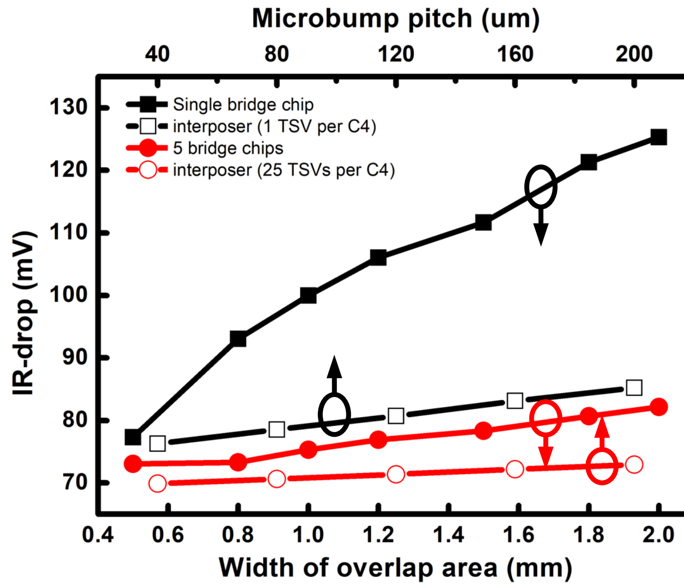


Figure 3.13: IR-drop comparison of interposer and bridge-chip technologies as a function of key parameters for each case

an important role, and for the bridge-chip case, the width of the overlap area is the critical factor. Therefore, we plot the maximum IR-drop of the system as a function of the above variables using the same Y-axis, as shown in Fig. 3.13. We sweep the microbump pitch of interposer from $40\ \mu m$ to $200\ \mu m$; on the other hand for the bridge-chip case, we sweep the overlap area from $0.5\ mm\ \times\ 6\ mm$ to $2\ mm\ \times\ 6\ mm$ (width of overlap area changes from $0.5\ mm$ to $2\ mm$).

For the interposer case, with a larger microbump pitch, the IR-drop gradually increases. This is because the additional microbump and TSV resistances will offset the spreading effects of the interposer PDN. When the microbump pitch is increased from $40\ \mu m$ to $200\ \mu m$, there is an 11.7% and a 4.38% IR-drop increase for the one TSV per C4 bump case and the 25 TSVs per C4 bump case, respectively. This indicates: first, without fine-pitch microbumps, there are not many benefits to using interposer, and second, using bundled-TSVs, even when the microbump pitch is limited, the IR-drop will reduce.

For the bridge-chip case, as the overlap area increases, the IR-drop inevitably increases since the center of the overlap area becomes further away from the nearest C4 bumps. However, with multiple bridge-chips, the IR-drop is less sensitive to the overlap area than

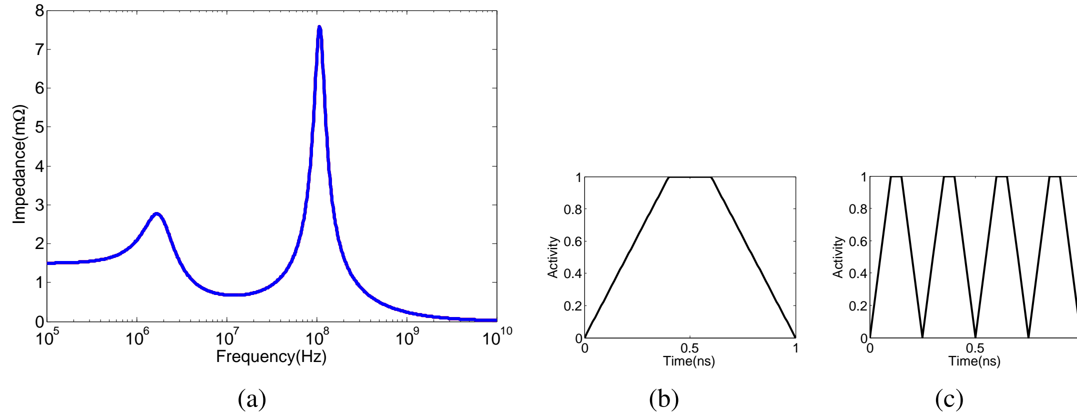


Figure 3.14: (a) Impedance analysis of a single on-die PDN node and illustration of the switching current activity (a) waveform #1 1 GHz frequency (c) waveform #2, 4 GHz frequency

the single bridge-chip case and it incurs an IR-drop increase of 12.5% while for single bridge-chip case, the IR-drop almost doubles when the overlap region is 2 mm wide instead of 0.5 mm.

In summary, there are challenges and opportunities for both interposer and bridge-chip technologies from an IR-drop and manufacturing perspectives. For interposer technology, the key parameters are fine-pitch microbumps and high density TSVs, while for the bridge-chip based technologies, the key parameters are overlap area, single versus multiple smaller bridge-chips, and the location of power hotspots.

Transient Droop

For transient analysis, the supply noise results from the switching current. Fig. 3.14(a) shows the impedance analysis results of an on-die node. The chip operating frequency (> 1 GHz) is higher than the resonant frequency (about 150 MHz), therefore we only consider two waveforms with different frequencies (1 GHz and 4 GHz). The two waveforms are illustrated in Fig. 3.14(b) and 3.14(c). Waveform #1 has a rise time, pulse time, fall time and period of 400 ps, 200 ps, 400 ps, and 1000 ps, respectively and waveform #2 if four-times the frequency of waveform #1, as shown in Fig. 3.14(c). Since we already benchmarked

Table 3.3: Transient state analysis results

Unit: mV	Waveform #1		Waveform #2	
	Die #1	Die #2	Die #1	Die #2
Single-die	103.19	160.46	91.95	146.03
Interposer	99.93	155.37	91.61	143.06
Bridge-chip	109.44	160.46	97.50	146.06

interposer and bridge-chip cases with different technology parameters, for simplification of transient droop analysis, we only investigate an interposer with 25 TSVs per C4 bump and a bridge-chip based 2.5-D heterogeneous integration using five bridge-chips between the dice with an overlap to be 0.5 mm $\times$ 6 mm.

The results are summarized in Table. 3.3. The droop curves of the worst node for both waveforms are shown in Fig. 3.27. The frequency of waveform #1 (1 GHz) is much closer to the resonant frequency (approximately 125 MHz from Fig. 3.14(a)) than that of waveform #2 (4 GHz), therefore, waveform #1 produces much larger on-die noise swing and relatively larger first droop. Therefore, in the following analysis, we focus on the results of waveform #1. Compared to the standalone case, the interposer achieves a PSN reduction of approximately 3.16% and 3.17% for Die #1 and Die #2, respectively. For the bridge-chip case, there is only a 6.04% increase in Die #1 and a minimal increase in Die #2 (which can also be seen by the lack of noise hotspots in the peripherals of Die #2). Another observation is that the difference between the evaluated cases is not as significant as was in the IR-drop analysis since the switching noise principally results from inductive parasitics of the package. Note, all PSN results in this chapter are strongly dependent on the power-maps assumed; for example, if a bridge-chip is within the footprint of a large power density region, PSN will be impacted more severely.

3.4 Impact of PDN in the Bridge-Chip

Thus far, we assumed that the bridge-chip in a 2.5-D configuration is used for die-to-die signaling; bridge-chip does not have any PDN. Recent studies show that a bridge-chip can contain multi-layer PDN [103]. In this section, we investigate the impact of adding PDN in

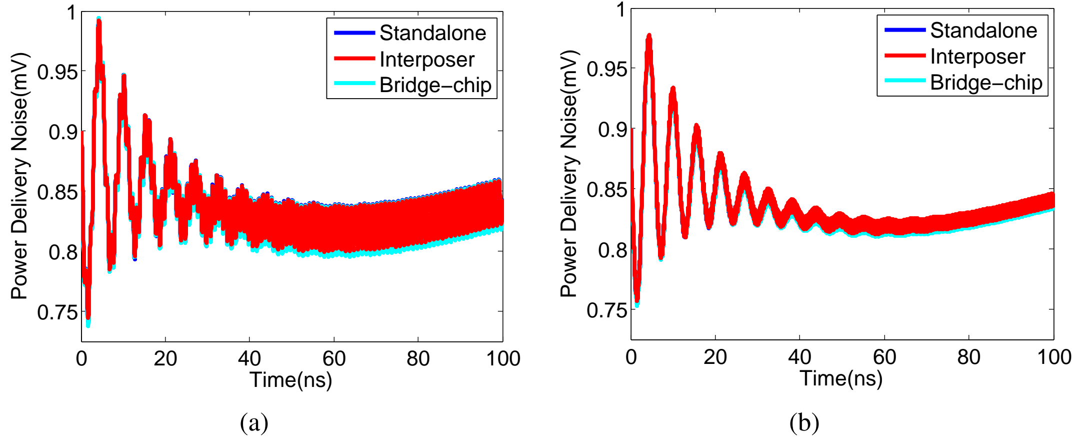


Figure 3.15: Transient analysis results of the point with largest droop (a) waveform #1 (b) waveform #2

the bridge-chip. We perform various case studies for two different bridge-chip based 2.5-D configurations: CPU-FPGA integration and stacked memory-FPGA integration.

3.4.1 PDN Schematics with Bridge-Chip PDN

Fig. 3.16(a) and 3.16(b) show the schematic diagram for two different scenarios under consideration. Prior sections consider the scenario described in Fig. 3.16(a). Similar to the previous studies, owing to the overlap region between the dice and the bridge-chip, the package PDN still does not have direct access to the peripheral circuits on the die. In this section, we make a few assumptions. First, we assume that the microbumps are part of the PDN as well; some microbumps are interconnecting the on-die PDN to the bridge-PDN. Second, we assume that the peripheral circuits of different dice might share the same voltage domain. Finally, we assume that the bridge-chip metal-stack is multi-layer to accommodate PDN as well as the signaling network. Specifically, we investigate three scenarios.

- Inclusion of the VSS (ground) network in the bridge-chip: Our assumption of sharing the voltage domain across different dice might not hold for all cases. However, two dice can always share a ground. Fig. 3.17(a) presents this scenario.

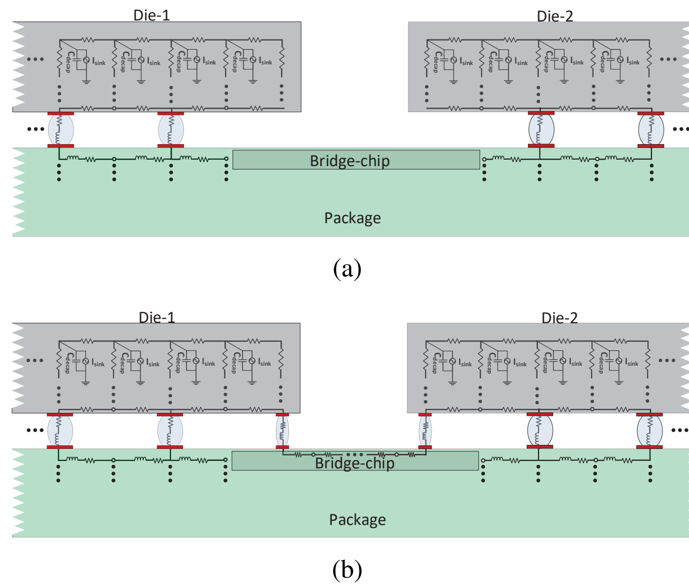


Figure 3.16: PDN schematic diagram (a) excluding bridge-chip PDN and (b) including bridge-chip PDN

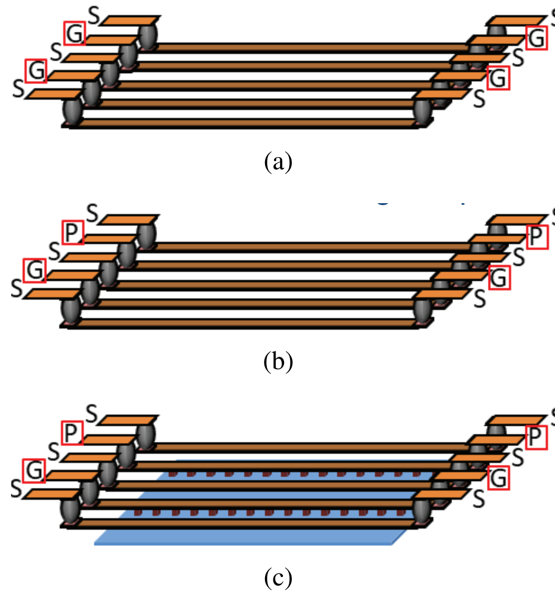


Figure 3.17: (a) Ground net in the bridge-chip, (b) power and ground nets in the bridge-chip, and (c) metal-insulator-metal capacitors in the bridge-chip

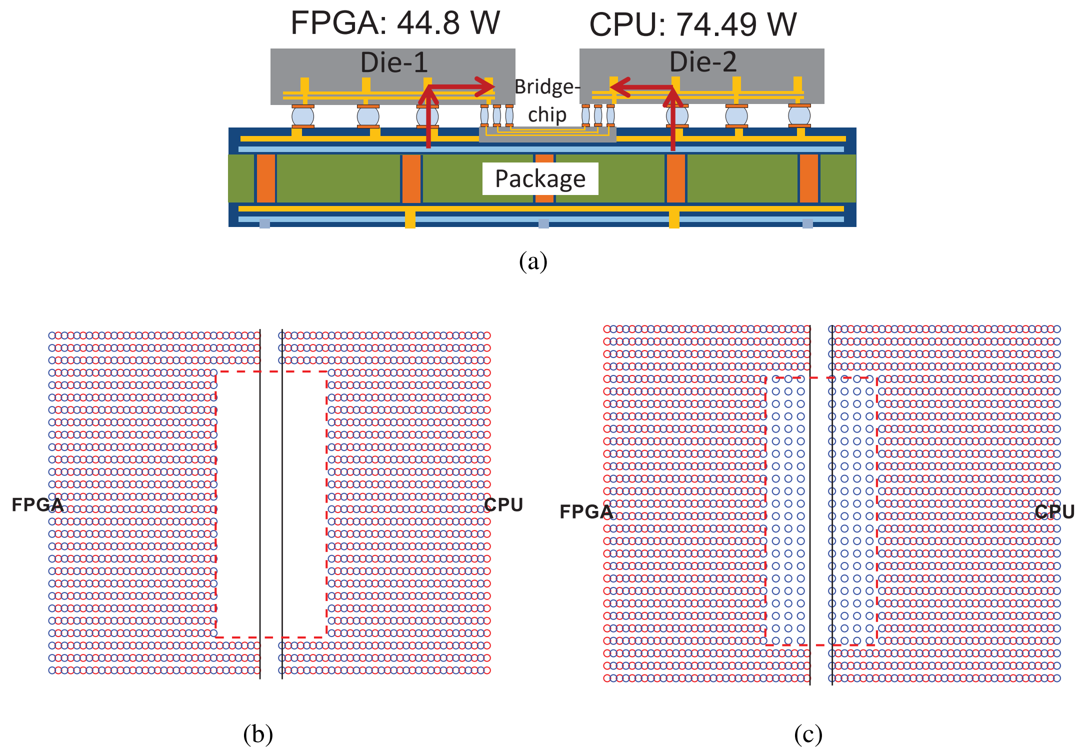


Figure 3.18: (a) CPU-FPGA configuration with re-routed PDN for the peripheral circuits, (b) die-to-package bump map with no PDN in the bridge, and (c) die-to-package bump map with ground net in the bridge-chip

- Inclusion of the power and ground network in the bridge chip: Fig. 3.17(b) illustrates this concept.
- Inclusion of metal-insulator-metal decoupling capacitors in the bridge-chip: In the event where we have both power and ground available on the bridge-chip, it might be possible to embed metal-insulator-metal (MIM) decoupling capacitors in the bridge-chip. Fig. 3.17(c) presents this scenario.

3.4.2 Bridge-Chip PDN Analysis for 2.5-D of CPU-FPGA Integration

Similar to prior studies in this chapter, we consider a bridge-chip based 2.5-D integration of a CPU die and an FPGA die, as shown in Fig. 3.18(a). The figure also shows how the current has to re-route through the nearest package-to-die bumps to deliver power to the

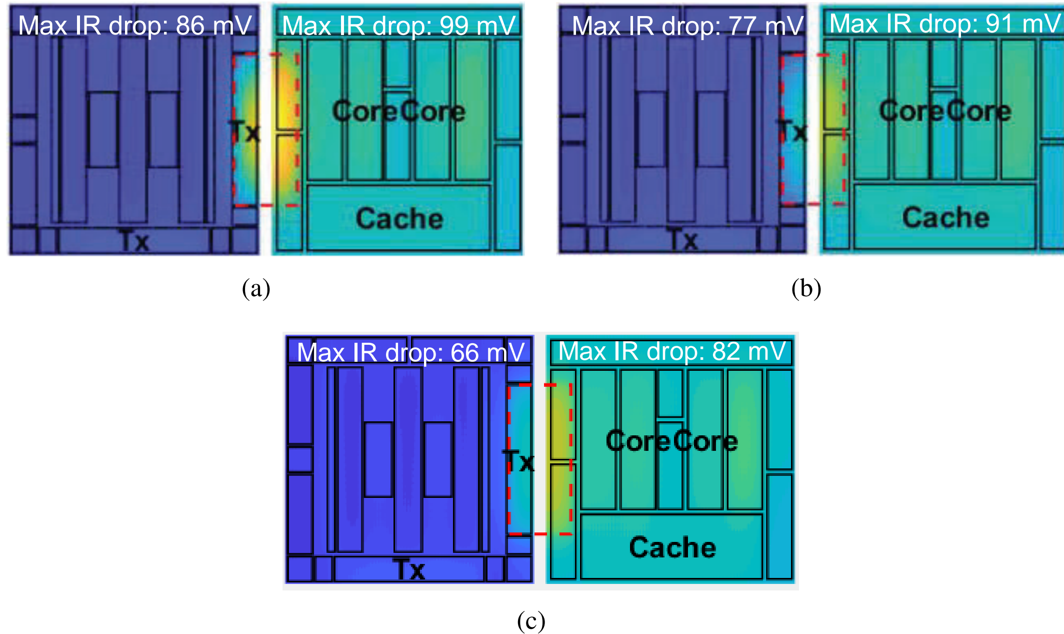


Figure 3.19: DC IR-drop results for (a) no PDN in the bridge-chip, (b) ground network in the bridge-chip, and (c) both power and ground network in the bridge-chip

peripheral circuitry. Fig. 3.18(b) shows the package-to-die bump patterns for 'no bridge-chip PDN' case. The microbumps in the overlap region are cut-off from the PDN. Fig. 3.18(c) shows the bump pattern with the inclusion of the ground network in the bridge-chip. Including both power and ground network will have a bump pattern similar to this.

Power and ground network in the bridge-chip

Fig. 3.19 summarizes the steady-state IR-drop results for the bridge-chip based configuration under consideration. With no bridge-chip PDN, the CPU die and the FPGA die have 99 mV and 86 mV IR-drop, respectively. If we include the ground network in the bridge-chip, as shown in Fig. 3.19(b), there is an 8% and a 10% reduction in IR-drop for the CPU die and the FPGA die, respectively. This reduction is further enhanced by the inclusion of both power and ground network in the bridge-chip. Fig. 3.19(c) presents this result. Compared to the 'no bridge-chip PDN' case, this case improves the IR-drop by 17% for the CPU die and 23% for the FPGA die. The on-die PDN is more resistive than the package

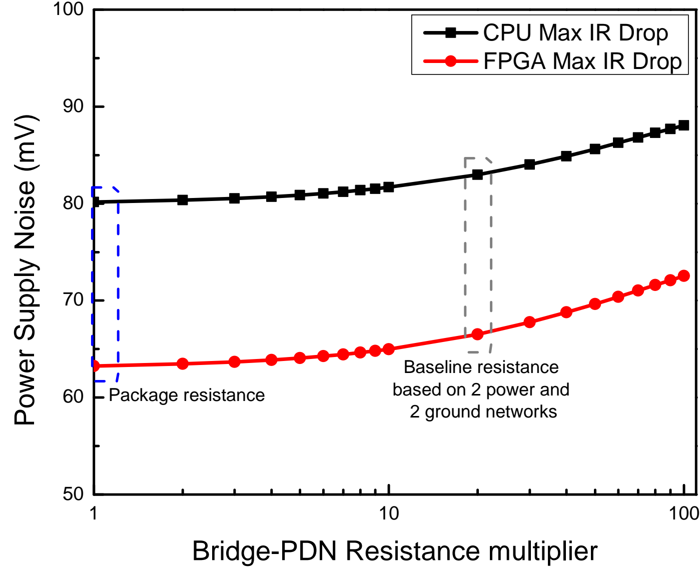


Figure 3.20: Impact of bridge-chip PDN resistance on DC IR drop

PDN. However, the PDN in the bridge-chip acts as a parallel resistance to the on-die PDN path for the peripheral circuits. Hence, we observe the reduction in the IR-drop for each die. We also analyzed the impact of bridge-chip PDN resistance on the performance of a CPU-FPGA system. Fig. 3.20 shows the results for this analysis. For our baseline case, we calculate the PDN resistance in the bridge under the assumption of two and ground networks in the bridge-chip. In the figure, this shows the case where we have a multiplier of 20. A multiplier of 1 means the resistance equivalent to the resistance of the package PDN. We observe that regardless of the resistance of the PDN in the bridge-chip, this will always improve the PSN.

We also investigate the $L\frac{di}{dt}$ based transient analysis results for this configuration. We use a 1 GHz on-die stimulus for this study (Fig. 3.21). For the CPU die and the FPGA die, we observe a 5% and 9% decrease in the first droop noise, respectively. The bridge-chip PDN reduces the resistance of the on-die network. However, the transient first droop noise depends on the package inductance and the on-die decoupling capacitors. Inclusion of the bridge-chip PDN does not impact any of these two factors. Hence, we observe lesser impact on transient noise as we observed in the steady-state IR-drop analysis. Moreover,

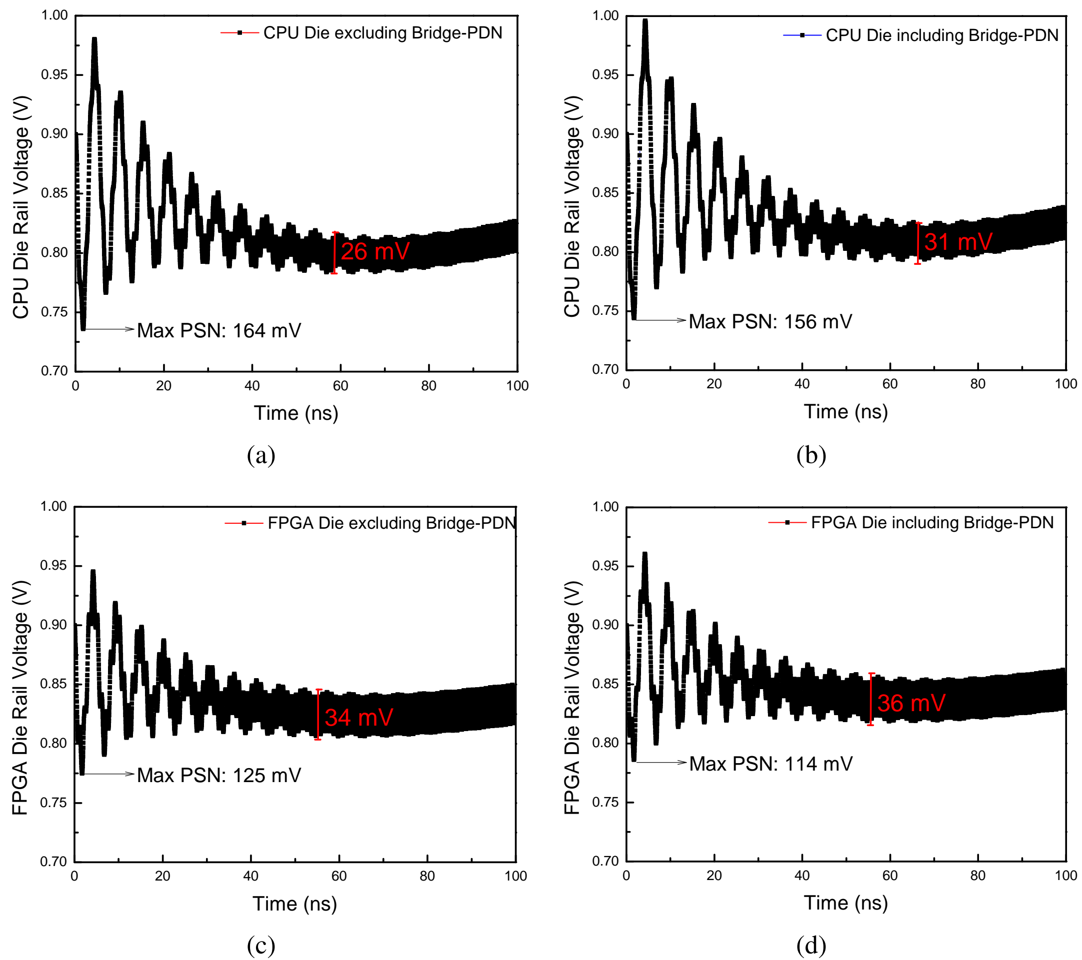


Figure 3.21: Transient analysis results for a 1 GHz pulse on-die excitation for (a) CPU die excluding bridge-chip PDN, (b) CPU die including bridge-chip PDN, (c) FPGA die excluding bridge-chip PDN, and (d) FPGA die including bridge-chip PDN

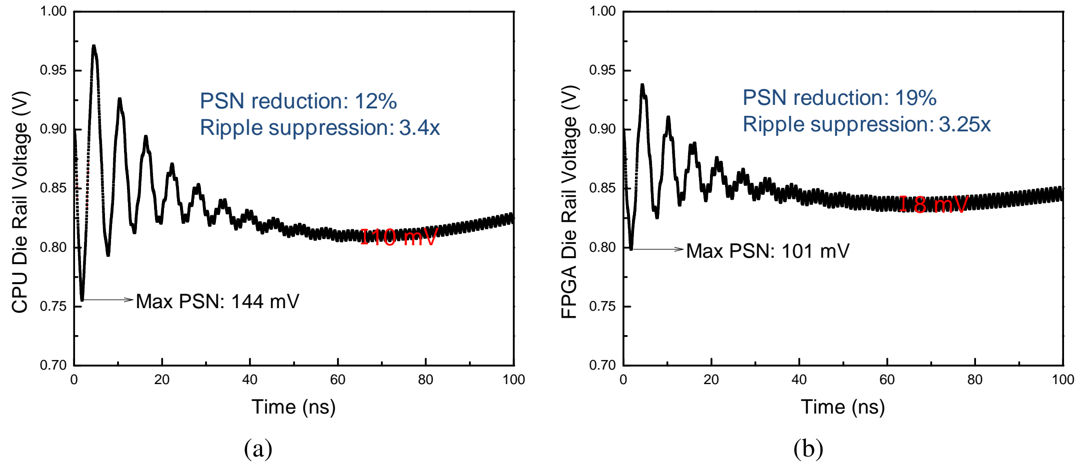


Figure 3.22: Transient analysis results including metal-insulator-metal capacitors in the bridge-chip for (a) CPU die and (b) FPGA die

since the resistance is lesser with the inclusion of the bridge-chip PDN, there is a slight increase in high-frequency ripple across the noise profile.

Decoupling capacitor in the bridge-chip

On-die decoupling capacitors reduces the first droop noise. However, owing to the limited on-die space, the number of on-die decoupling capacitors that can be added is very limited. If both power and ground networks are available in the bridge-chip, then MIM capacitors can potentially be embedded within the bridge-chip metal layers. In this study, we use a decoupling capacitor density of 5 nF/mm^2 in the bridge-chip. Fig. 3.22 shows the results for this scenario. For the FPGA die, the first droop noise reduces by 19% compared to the 'no bridge-chip PDN' case. For the CPU die, this reduction is 12%. Compared to the 'no bridge-chip capacitor' case, this is an 11.4% and a 7.6% improvement for the FPGA die and the CPU die, respectively. We observe a significant reduction in high frequency ripple in the supply voltage. For both dice, this high frequency ripple is reduced by greater than 3x compared to the results shown in Fig. 3.21. MIM capacitor density depends on its structure, dielectric material, etc. Hence, we also vary the decoupling capacitor density in

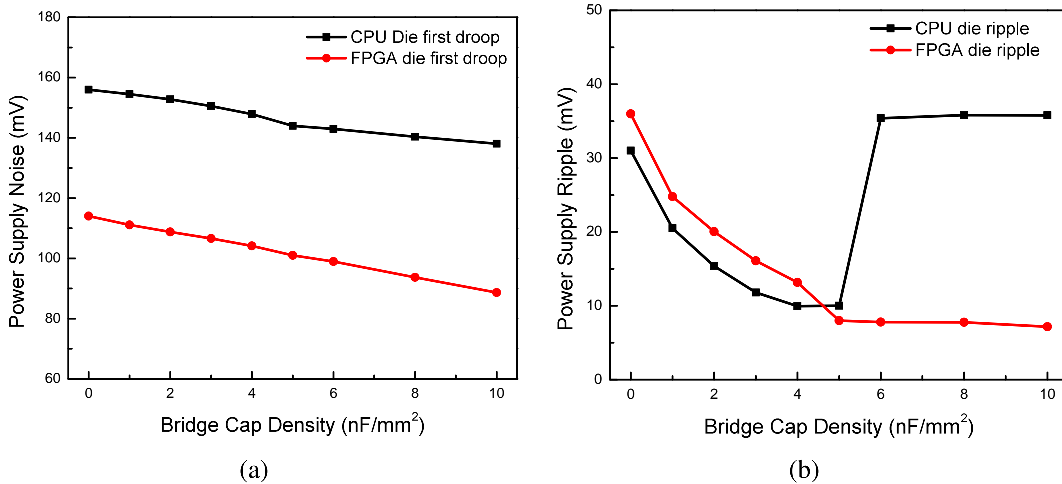


Figure 3.23: Impact of MIM capacitor density on (a) PSN and (b) high frequency ripple

the bridge-chip from 0 nF/mm^2 to 10 nF/mm^2. Fig. 3.24 summarizes these results. While adding decoupling capacitor helps reduce the PSN for both chips, beyond 5 nF/mm^2, we observe a diminishing return for the CPU die. From the high frequency ripple perspective, we observe that for the CPU die, beyond 5 nF/mm^2, the high frequency ripple increases. This can be attributed to the shift in the high PDN noise region for the CPU die. As shown in Fig. 3.24, with no or little decoupling capacitor in the bridge-chip, the PDN 'shadow' region in both of the dice persists. However, beyond 5 nF/mm^2, the shadow region in the CPU die is non-existent. The maximum PDN noise spot moves away from the edge of the die where the PSN profile looks similar to that of the multi-chip module configuration discussed in this chapter. Since we only report the maximum noise for all decoupling capacitor densities, we observe a sharp shift in the high frequency ripple in the CPU die. However, for the FPGA die and in the overlap region for the CPU die, additional decoupling capacitors still help reduce the PSN. We do not see a similar trend in the FPGA die owing to the assumed power map with higher power regions in the edge of the die. If our bridge-chip structure allows us to have a higher capacitor density or the power map changes to put lower power blocks in bridge-chip overlap region of the die, we might observe a trend similar to the CPU die under consideration.

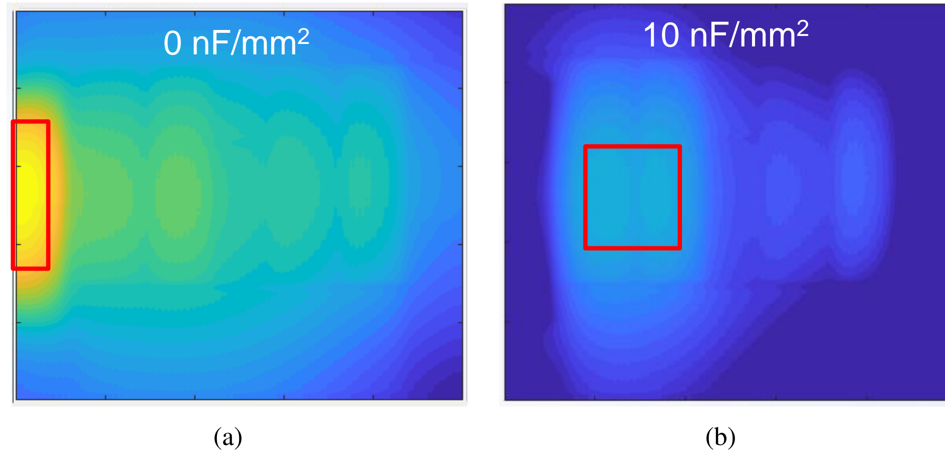


Figure 3.24: Impact of MIM capacitor density on maximum noise location for (a) no MIM capacitors and (b) 10 nF/mm^2 MIM capacitor density

3.4.3 PDN Analysis for a 2.5-D Integration of Stacked Memory-FPGA Configuration

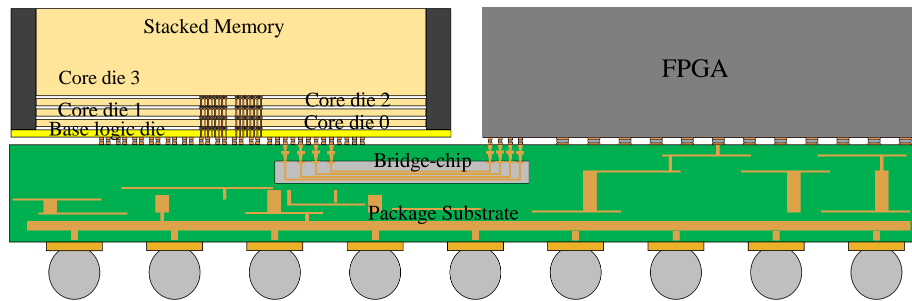


Figure 3.25: HBM-FPGA configuration with bridge-chip

Fig. 3.25 shows a bridge-chip based 2.5-D configuration with stacked memory and an FPGA die. From prior analysis in this chapter, we observe that the increased overlap region between a bridge-chip and a die leads to increased supply noise. For a stacked memory based configuration, the memory banks are centrally interconnected to the package using TSVs. The base logic die has an area array distribution of the package-to-die bumps. However, even for the base logic die, we assume that the bumps in the bridge-chip overlap region provide mechanical support; they are not electrically connected. Owing to the centrally distributed memory PDN, the bridge-chip extends farther towards the memory die than the FPGA die. In this study, we investigate the impact of this overlap region on

power delivery to the memory dice.

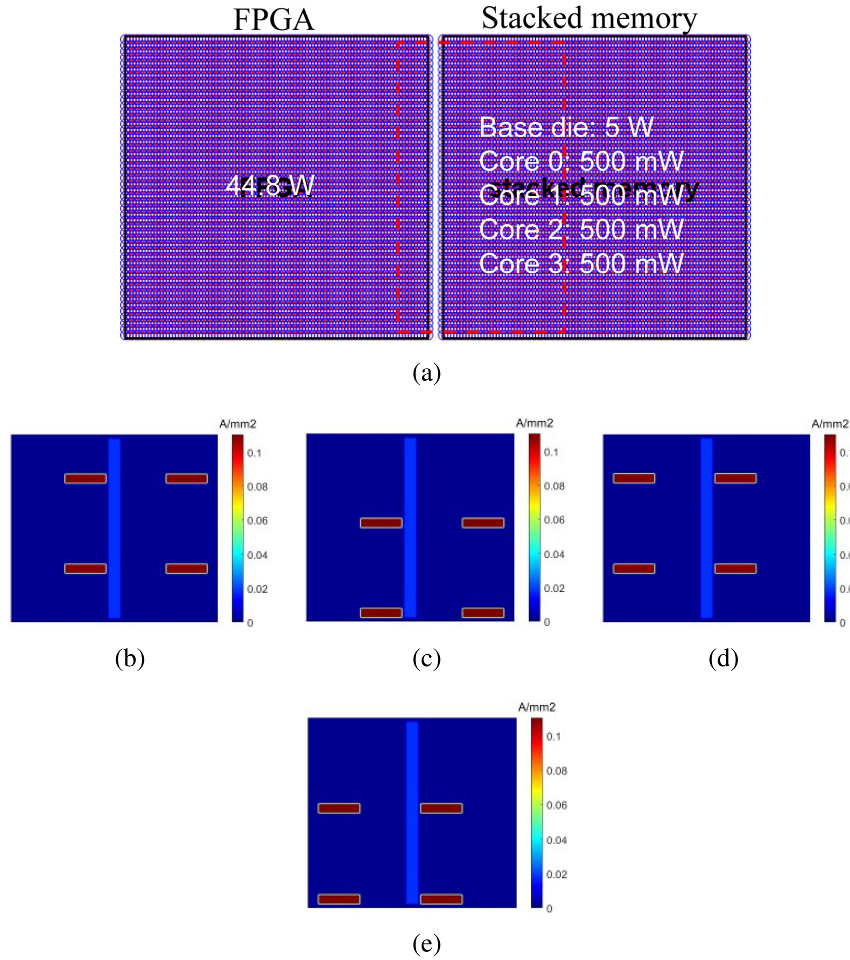


Figure 3.26: (a) FPGA-stacked memory power specifications and power map for memory (b) core die 0, (c) core die 1, (d) core die 2, (e) core die 3

Fig. 3.26 shows the power specifications of different dice under consideration. For the FPGA die, we use the same specifications as the prior studies in this chapter. For the stacked memory, we assume that the base logic die consumes 5 W. Moreover, we assume that the memory die has four core dice. Each core die has two channels and four pseudo-channels [104]. Each pseudo-channel has eight memory banks on each side of the central I/Os. Fig. 3.27(a) and 3.27(b) summarizes the steady-state IR-drop results for the base logic die and the FPGA die with and without bridge-chip PDN, respectively. Owing to the larger overlap region towards the memory die, the base logic die has highly resistive PDN path to the

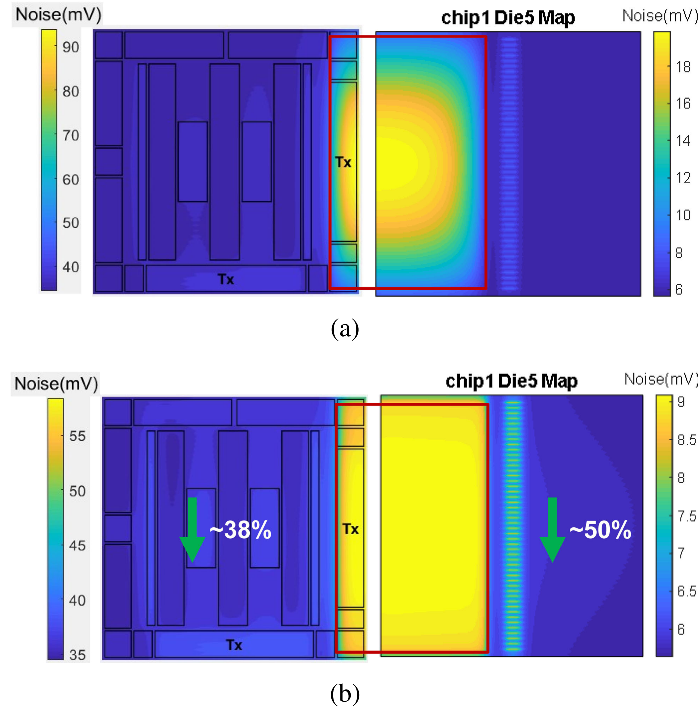


Figure 3.27: DC IR-drop results for the FPGA and memory dice (a) excluding PDN in the bridge-chip and (b) including PDN in the bridge-chip

peripheral circuits. With the inclusion of the bridge-chip PDN, the IR-drop in the FPGA die and the base logic die reduces by ∼38% and ∼50%, respectively. However, for the memory core dice, the IR-drop is almost invariant to this overlap region. Fig. 3.28 presents these results. This non-sensitivity to the bridge-chip PDN can be attributed to the centrally distributed TSVs for power supply. If the bridge-chip extends more toward the memory dice so that it overlaps with the memory I/O region, we would see the impact of bridge-chip overlap on memory bank power delivery as well. However, owing to this configuration, the die-to-die signaling channels between an FPGA die and the memory banks are longer. This will impact the signaling network of this configuration. To investigate the impact of the bridge-chip overlap on the PSN of a die similar to the base die in this configuration, we evaluate a two die system where the Die-1 is the aforementioned FPGA die and the Die-2 is a chip with variable power consumption. We varied the power of the Die-2 from 5 W to 100 W with the same overlap region. Fig. 3.29 presents the results of this analysis. We observe

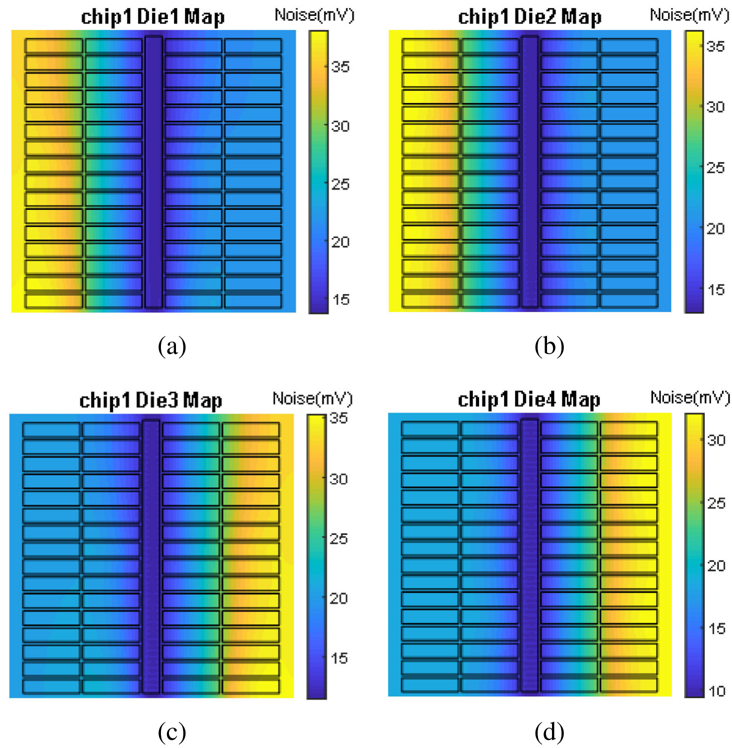


Figure 3.28: DC IR-drop results for different memory dice in the stacked memory; (a) core die 0, (b) core die 1, (c) core die 2, and (d) core die 3

that for different power values, there is a significant benefit in using PDN in the bridge-chip. For example, for the 100 W case, we observe a 42% VDD PSN for a configurations which excludes PDN in the bridge-chip. Including the bridge-chip PDN, this PSN is 17% VDD (i.e. 2.5x reduction). Hence, bridge-chip PDN can be vital depending on applications and power consumption of dice.

3.5 Design Space Exploration of Power Delivery Including Voltage Regulator Modules

In this section, based on prior PDN modeling efforts [5, 58, 105], different voltage regulator module (VRM) placement methodologies e.g., on-package, 3-D stacked VRM-chip, VRM placed on the backside of the package, etc. are explored.

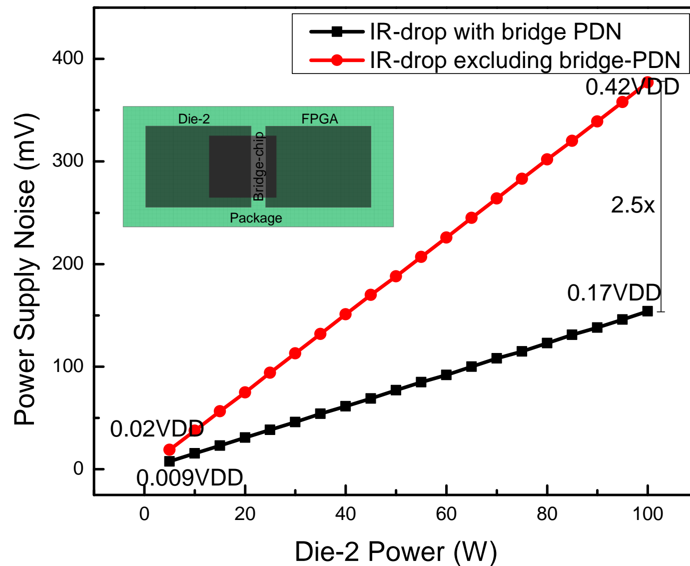


Figure 3.29: Impact of bridge-chip overlap for a die with varying power

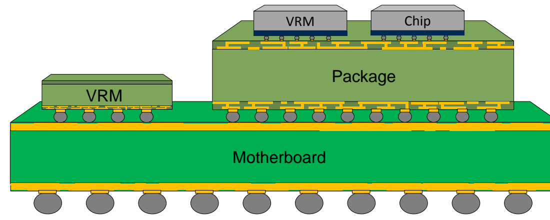
3.5.1 Benchmark Architectures

Several benchmark configurations have been analyzed in this chapter. A brief description of each of the configurations is given below.

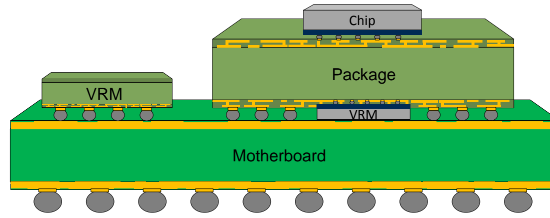
- **Side-by-side VRM-chip configuration:** In Fig. 3.30(a), the VRM chip is placed next to the active chip on the same package and thus, the long interconnect distance from the power supply to the chip is reduced. This, in effect, is expected to reduce the overall IR-drop compared to the case with the off-chip VRMs placed in the motherboard.

- **Backside-of-the-package VRM technology:** The configuration shown in Fig. 3.30(b) considers a VRM chip placed on the backside of the package. In such an approach, the parasitics of the board PDN and the package PDN are mostly eliminated from the noise calculation.

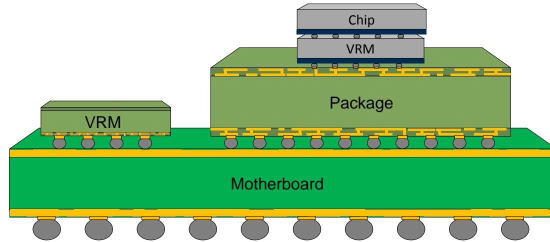
- **3-D-IC Chip-on-VRM topology:** Fig. 3.30(c) shows the 3-D IC stacking of a processor chip on top of the VRM chip. The parasitics in the path of power supply consist of only TSVs and bumps. In effect, the PSN is expected to decrease further.



(a) On-package VRM configuration



(b) Backside-of-the-package VRM Configuration



(c) 3-D IC Chip-on-VRM Configuration

Figure 3.30: Benchmark architectures

3.5.2 PDN Topology and Specifications

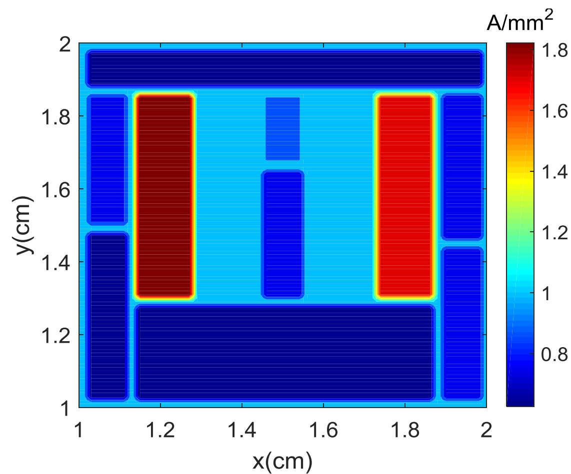


Figure 3.31: The non-uniform current density map used for the analysis

Table 3.4: PDN parameters

TSV pitch	100 μm
TSV resistivity [106]	80×10^{-9} Ωm
TSV contact resistance [107]	0.45 $\Omega\mu$m^2
Package wire thickness (metal planes)	10 P/G metal layers, 0.010 mm per layer (5 for Power and 5 for Ground)
Package wire resistivity [108]	180×10^{-9} Ωm
On-chip PDN wire dimensions	5 um thick, 3.3 um wide, 30 um pitch
On-chip PDN wire resistivity	17.1×10^{-9} Ωm
On-chip decap	5.3 nF/mm^2
C4 bump diameter	40 μm
C4 bumb pitch	100 μm
C4 bump material	Solder (alloy; 100×10^{-9} Ωm)

The overall analysis flow is as described in the prior work [109, 7]. Throughout this chapter, a 1 cm $\times$ 1 cm chip is considered. The active chip is assumed to have a 1 V supply voltage rail and a total power of 100 W. VRM parasitic resistance and inductance are extracted from the literature [108, 110]. The simulation specifications of different parameters are described in Table 3.4. The package level decoupling capacitors are discrete and have capacitance (C_esc_pkg), resistance (C_esr_pkg), and inductance (C_esl_pkg) associated with them. Non-uniform current density map with distinct high-power blocks is used for the simulations. The power map (or current density map) is specified in Fig. 5.4(a) which is taken from [88], but is modified according to [111, 76].

3.5.3 DC IR-Drop Comparison of Different Benchmark Configurations

In this section, the DC IR-drop for different configurations i.e., on-package VRM, 3-D IC chip-on-VRM, and the backside-of-the-package VRM, etc., has been analyzed. In each configuration, adding additional VRMs to the system reduces DC IR-drop. The impact of multiple VRMs on PSN suppression is more pronounced if there are hotspots in the chip. Fig. 3.32 shows one such example. Fig. 3.32(a) and 3.32(b) show the relative positions

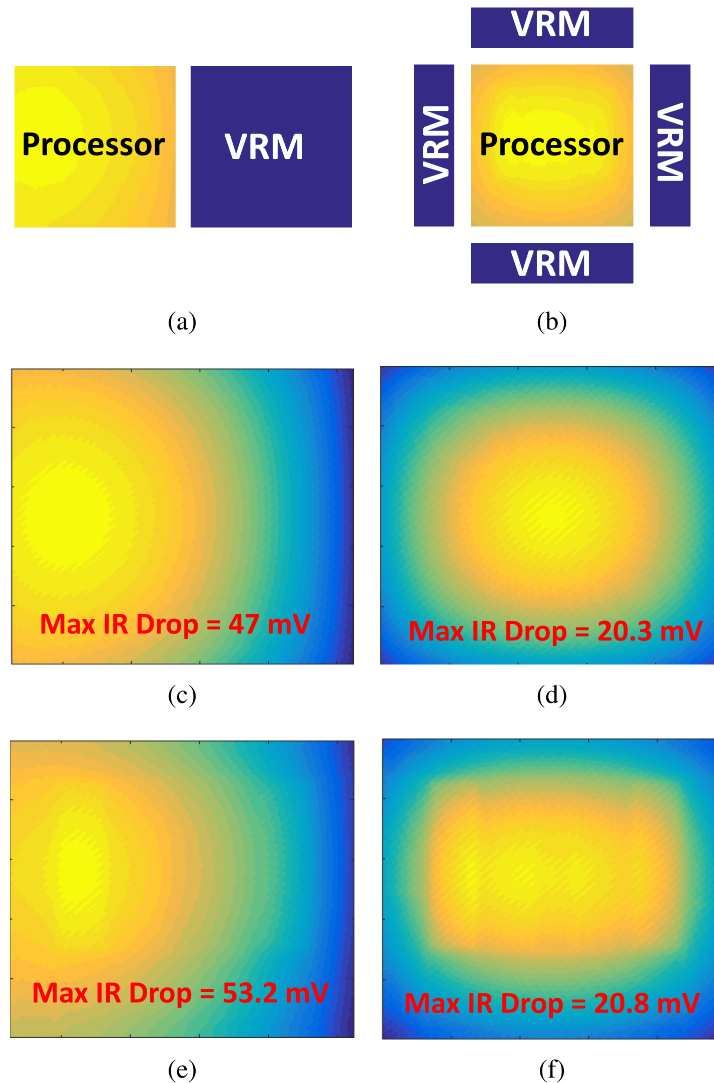


Figure 3.32: a) Single on-package VRM configuration, b) four on-package VRM configuration; DC IR-drop for c) single on-package VRM case, d) four on-package VRMs case with uniform current density map; e) single on-package VRM case, and f) four on-package VRMs case with non-uniform current density map

of the VRMs with respect to the active chip. With the single VRM placed farther from the high current density region, there is a significant increase in the IR-drop, as shown in Fig. 3.32(c) and 3.32(e). On the other hand, multiple VRMs suppress the hotspot issues as the effective distance between the high power load and the voltage regulator is less than that with the prior case. Fig. 3.32(d) and 3.32(f) show the IR-drop suppression effect using four on-package VRMs. The maximum IR-drop is reduced by 60.9% using four on-package

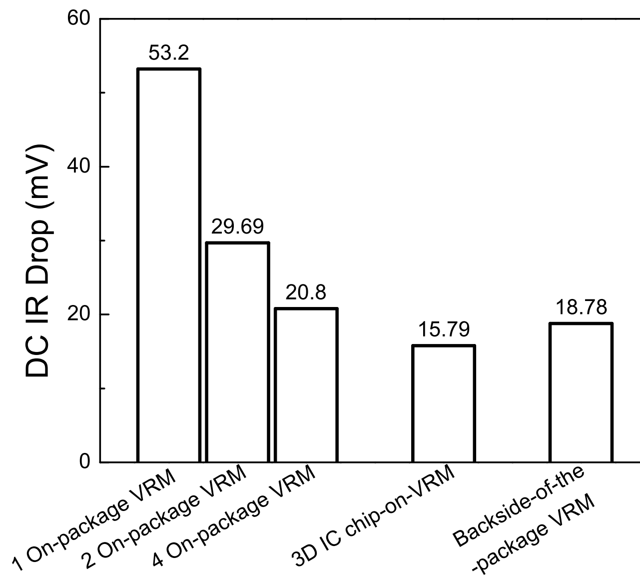


Figure 3.33: Comparison of DC IR-drop for different configurations

VRMs instead of one. The IR-drop for single on-package VRM case with a non-uniform current density map is 13.2% larger than the case with a uniform current map. However, for a similar condition, the four on-package VRMs case has a 2.4% increase in the IR-drop.

Fig. 3.33 summerizes the results for different VRM-processor configurations. All results are obtained using the non-uniform current density map specified in Fig. 5.4(a). In the backside-of-the-package VRM and 3-D IC chip-on-VRM cases, owing to the shorter distance, the IR-drop is smaller compared to the prior on-package VRM cases. In the backside-of-the-package configuration, the through package vias and metal layers in the package PDN are important components of the power delivery path. In the 3-D IC case, however, due to the dense bumps between the chips, the TSVs in the VRM chip and the microbumps between the VRM and the active chip are the only contributors of the parasitics in the PDN path. As a result, the IR-drop for the 3-D IC case is 24% and 15.9% smaller than that of four on-package VRMs and backside of the package VRM cases, respectively.

Since the multiple on-package VRMs case brings the regulator circuits closer to the active chip, there are different trade-off analyses which determine how close we can bring these chips. Fig. 3.34 shows the DC IR-drop results for three different distances between

the chips.

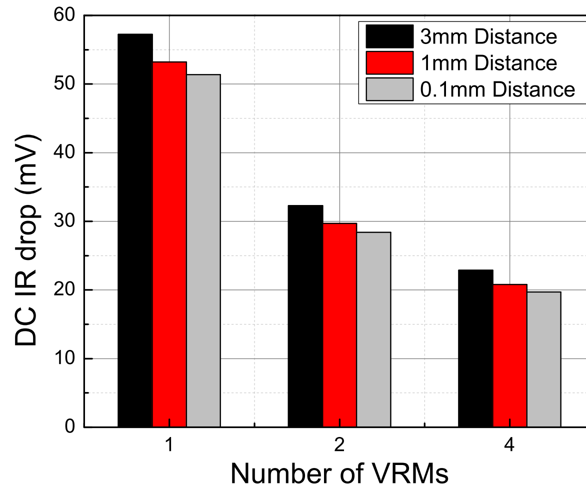


Figure 3.34: Comparison of maximum IR-drop for different VRM-chip gaps in the on-package VRM configurations

In the baseline model, the distance between the VRM and the processor chip was fixed to 1 mm. Increasing the distance is beneficial for reducing thermal coupling between two chips of different power density [15]. However, it increases the signaling and power delivery path lengths [112]. To investigate the impact of this on power delivery performance, the distance was varied from 3 mm to 0.1 mm. Fig. 3.34 summarizes the results for 3 mm, 1 mm and 0.1 mm distances. As expected, if the distance is increased, the interconnect length for power supply increases, which eventually increases the IR-drop. Conventionally, decoupling capacitors (decaps) are placed in these regions. Reducing this inter-chip distance would result in lesser decaps in the vicinity of switching. This is another trade-off that has to be considered. Generally, PSN suppression is more important and hence, a greater emphasis is placed on a lesser inter-chip distance.

In this study, bump pitch is held at 100 μm for the 3-D IC chip-on-VRM case. In this section, the impact of bump pitch scaling is investigated. The bump pitch is scaled from 100 μm to 500 μm. For each of the configurations, as the bump pitch increases, bump diameter is increased with the same factor. We use a bundled TSV approach and hence,

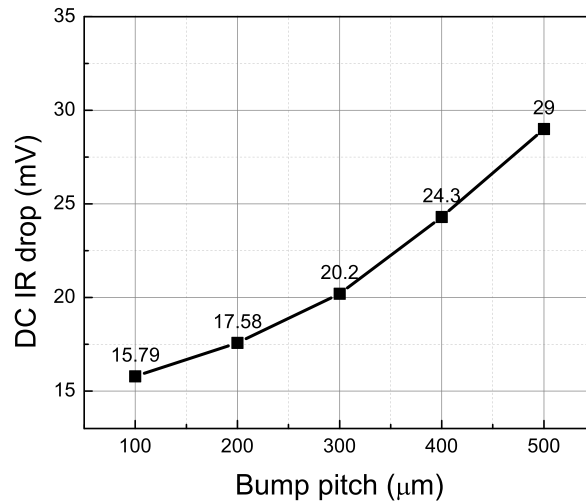


Figure 3.35: Comparison of IR-drop for different bump pitches in the 3-D IC chip-on-VRM configuration

the number of TSVs under each bump is also increased in quadratic progression. With the increased pitch, the overall number of bumps will decrease. Therefore, each bump will carry more current. If the number of TSVs is not increased in quadratic progression, then the current density in each TSV will increase, leading to increased joule heating [113] and potentially reducing the meantime to failure (MTTF)[114]. Fig. 5.6 summarizes the results. Only the 3-D IC case has been considered here for the analysis as we believe other cases will follow the same trend. For larger bump pitches, since the total number of TSVs is the same for all cases, and the bump resistance is decreased with increased diameter, the on-die loss is the only differentiating factor.

3.5.4 Comparison of Transient Noise for different configurations

In this section, we investigate the transient analysis results for the VRM-processor configurations under consideration. Owing to the area constraints, the amount of on-chip decap is very limited [115]. However, we can control the second and third droop since these are controlled by the integrated and mounted decaps in the package and the board. In all the transient simulations, the on-chip decap is fixed to the specified value as noted in Table

3.4. The controlling parameters are the discrete decaps on the board and the package. In each case, a small number of decaps is considered. Since this chapter is about the impact of different benchmark architectures on DC IR-drop and simultaneous switching noise (SSN), a detailed analysis of decap allocation for optimized result is out of scope. In this section, for different configurations, step response of the system will be shown. The supply voltage rises from 0 V to 1 V with a rise time of 1 ns.

Fig. 3.36 shows the transient noise profile for multiple on-package VRMs. As expected, with increased number of VRMs surrounding the chip, there is less PSN. In all cases, the transient noises generated from the interaction of capacitive and inductive (mainly package) elements oscillate and settle down to the DC IR-drop value of the corresponding case. The second droop is suppressed by the discrete decaps placed on the package. Also, the parasitic inductance from VRMs is somewhat suppressed by the low pass filter integrated with the regulator circuit. That is why the most dominant transient droop in all the cases is the first droop noise. The four on-package VRMs case achieves almost 24.45% improvement in PSN compared to the single on-package VRM case.

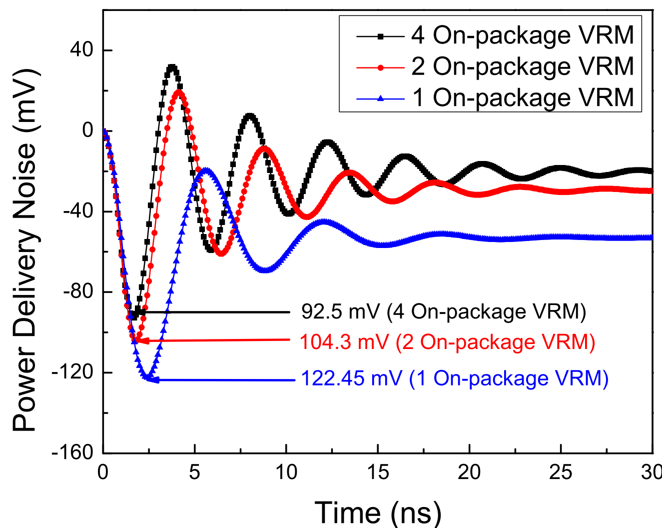


Figure 3.36: Comparison of transient noise for different on-package VRM configurations

When the VRM chip is placed on the backside of the package, VRM-to-chip PDN is

mostly dominated by package vias and bumps. Package vias typically have low aspect ratio compared to TSVs, so they contribute less to the resistance and more to the inductance of the system. Solder bumps between the package and the board play a similar role compared to the microbumps. Also, the number of microbumps is higher than the number of solder bumps. In the 3-D IC chip-on-VRM case, the VRM is directly supplying power from the bottom of the chip. Hence, the inductive components are the TSVs in the VRM chip and the microbumps between the VRM and the active chip. These are minimal compared to the inductive components in the other cases described in this study. In both cases, the package is less involved, which reduces the overall package parasitics in the PDN. Fig. 3.37 compares the best case from the on-package VRM cases with the backside-of-the-package and 3-D IC chip-on-VRM configurations. For the backside-of-the-package VRM case, the maximum PSN is 82.64 mV. This itself is 10.65% improvement compared to the four on-package VRMs case. The 3-D IC chip-on-VRM case provides a maximum PSN of 58.8 mV.

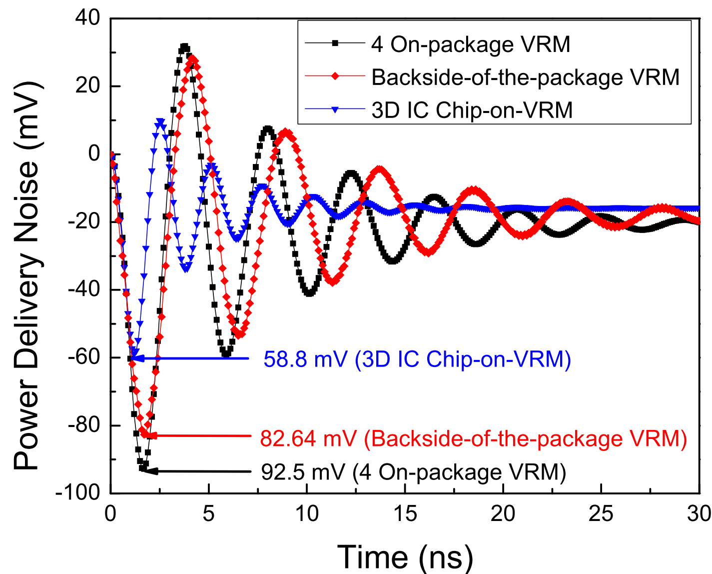


Figure 3.37: PSN comparison for different key benchmarks

Throughout the chapter, it's been obeserved that the transient noise is dominated by the

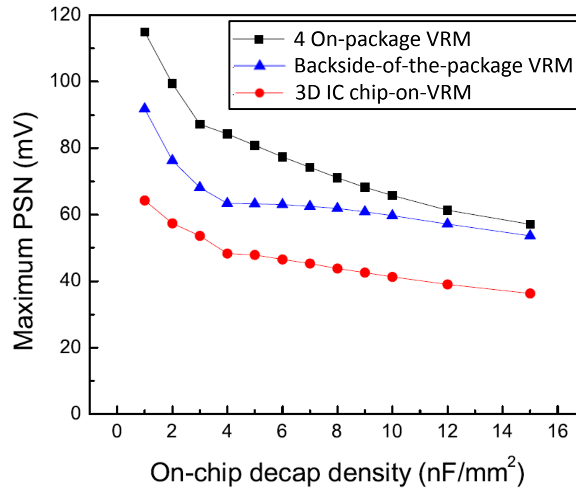


Figure 3.38: Maximum PSN of some key configurations for different on-chip decap density

first droop noise. This noise is dependent on the on-chip decap allocation. Throughout this chapter, a decap density of 5.3 nF/mm^2 has been used for the analysis. Typically, on-die decap can take 20-30% area depending on the available space [115]. Moreover, depending on the type of capacitors used, the decap density can vary [116]. Typically using MOS capacitors, a decap density of 10-20 nF/mm^2 can be achieved. The four on-package VRMs case, the 3-D IC chip-on-VRM case, and the backside-of-the-package VRM case have been simulated for a varying decap density. The density is varied from 1 nF/mm^2 to 15 nF/mm^2. To simplify the analysis, uniform power density has been considered. Fig. 3.38 summarizes the results from this study. As expected, with increased decap allocation, the PSN decreases. For the 3-D IC case, the maximum PSN reduced from 64 mV for 1 nf/mm^2 to 36 mV for 15 nF/mm^2. The other two cases follow the same trend as well. So, if the available space after floorplanning can be utilized for decap allocation, then using higher decap density enabled by MOS capacitors or metal-insulator-metal capacitors will suppress the PSN further along with the different configurations mentioned in this chapter.

Table 3.5: Thermal simulation parameters

Layer	Conductivity (W/mK)		Thickness (μm)
	In-plane	Through-plane	
TIM 2	3		30
Heat spreader	400		1000
TIM 1	3		30
Processor	149		100
VRM	149		100
Microbump and ILD [117]	1.6		40
Package	30.4	0.38	1000

3.5.5 Thermal Implications of Different Architectures

The different configurations studied in this chapter span from 2.5-D to 3-D integration. Depending on the type of circuitry and inductor placement, voltage regulators typically have 70-90% efficiency [35, 34, 72]. Hence, the VRMs are typically low power active chips which contributes to the overall temperature distribution of a given configuration. In this section, steady-state thermal analysis of different configurations is carried out in ANSYS. The parameter specifications used for thermal simulations are given in Table 6.1. There are two thermal interface material (TIM) layers in the system (i.e. one in each side of the heat spreader). The system is considered air-cooled, and the case-to-ambient thermal resistance is assumed to be 0.218 K/W along with an effective convective coefficient in the backside

Table 3.6: Thermal results

Configuration	Processor temperature (^oC)	VRM temperature (^oC)
1 On-package VRM	56.5 ~ 78	33 ~ 58.6
2 On-package VRM	55 ~ 74.6	42 ~ 56.6
4 On-package VRM	54.2 ~ 74.1	44.5 ~ 55.7
Backside-of-the-package VRM	55.8 ~ 75.2	91 ~ 117
3-D IC Chip-on-VRM	57.4 ~ 78.2	54.1 ~ 78.3

of the package [90]. There is a microbump layer between the VRM and the processor in the 3-D IC case. The package-to-chip connection is established by C4 bumps and underfill material. For both the bump layers, same thermal conductivity is used as specified in the table. The ambient temperature is assumed to be $22^{o}C$. The processor power is 100 W. As a starter, assuming 91% efficient regulators, the VRM power is estimated as 10 W. The thermal results are summarized in Table 3.6. The configuration with four on-package VRMs has the minimum temperature both in the chip and the VRMs. In the backside-of-the-package VRM case, because of the lower thermal conductivity of the package material, the VRM at an elevated temperature compared to the other cases. Since the VRM power is low relative to the processor chip power, 3-D stacking causes only a minor increase in chip temperature. This configuration minimizes power supply noise, while still being thermally feasible. However, the VRM chip itself is at an increased temperature, e.g. the VRM chip is 31% higher temperature compared to the 4 on-package VRMs case. Because of the thin layer of bumps between the processor and the VRM, these chips have similar temperature distribution. Moreover, it's observed that the lower power chip helps in heat spreading off the higher power chip.

We use 91% efficient regulators for simulations, which are towards the high-end regulators reported in the literature. As specified before, the overall efficiency can vary, which eventually means a higher power VRM die. In this analysis, VRMs with different power densities are simulated. The results are summarized in Fig. 3.39. The top and bottom figures report the maximum temperature of the VRMs and the processor for different VRM power densities, respectively. As the regulator efficiency decreases, the on-package VRM configurations and the backside-of-the-package VRM are almost invariant to the increased power density in the VRM die. However, the 3-D IC case is a bit more sensitive to this variation. For a 15 W/cm^2 change (10 W/cm^2 to 25 W/cm^2) in the VRM power density, there is approximately 10% increase in the maximum temperature of the dice. From this analysis, we can conclude that, with higher efficiency VRMs, the 3-D IC case is a feasible

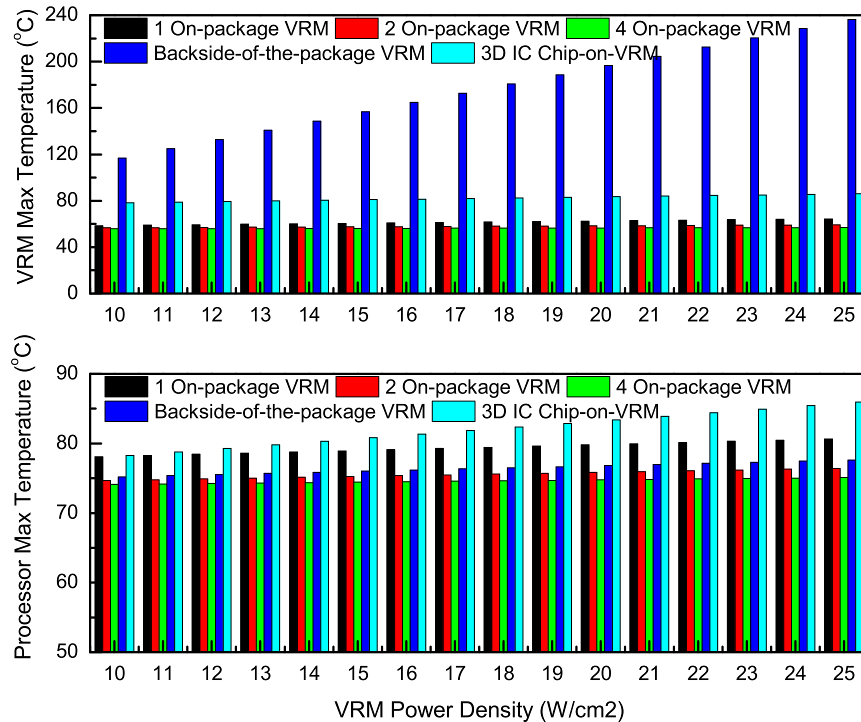
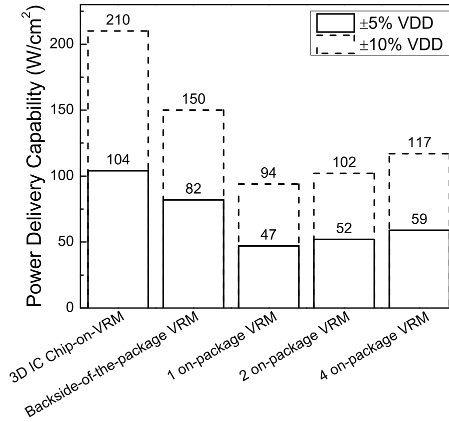


Figure 3.39: Processor and VRM maximum temperature for different configurations with respect to different VRM power density

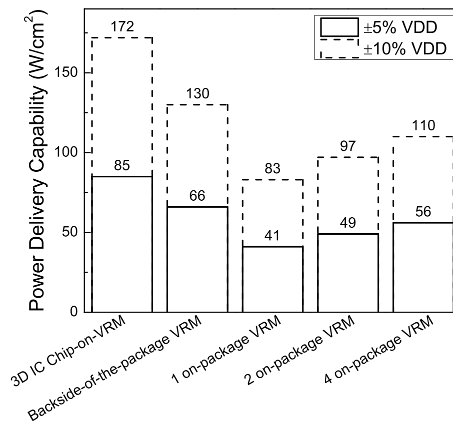
option. The best performing option of all the configurations regardless of the VRM efficiency is the side-by-side on-package VRM configurations. With a more advanced package cooling technique, the backside-of-the package VRM case can be a viable solution as well.

3.5.6 Power Delivery Capabilities of Different Architectures

In the preceding sections, a power delivery analysis is performed with the assumption of an overall 100 W/cm^2 power density in the active chip. However, in a real design, there is a limit up to which designers will allow the supply voltage to fluctuate. This supply voltage tolerance is typically 5% of the supply voltage [108]. Since 1 V supply has been considered in this analysis, a $\pm$50 mV is used as the threshold. In this section, the processor power for all the configurations is swept to observe the power delivery capability of different architectures.



(a)



(b)

Figure 3.40: Power delivery limit of different configurations for (a) uniform and (b) non-uniform current density map

Fig. 3.40(a) summarizes the results from the simulations with a uniform current density map in the chip. The four on-package VRMs case is capable of achieving a power density of 60 W/cm^2 without reaching the 50 mV limit. But in the 3-D IC chip-on-VRM case, the power delivery capability is more than 100 W/cm^2. Fig 3.40(a) also shows the power limits of different configurations if the designers allow a higher supply voltage fluctuation. The results follow the same trend of power delivery. In each case, the power delivery capability is almost doubled. Fig. 3.40(b) shows the results for a non-uniform current density map specified in the previous sections. As can be seen from the analysis, with respect to the hotspots on the chip, the 3-D IC chip-on-VRM case and the backside-of-the-package VRM

case can push the power density with a larger margin compared to the on-package VRM configurations.

3.6 Conclusion

This chapter presents a PDN analysis framework for emerging 2.5-D/3-D heterogeneous integration platforms. Interposer and bridge-chip based integration technologies are benchmarked and compared from a PDN point of view. Interposer based integration, with the right technology parameters, can exhibit a smaller IR-drop and transient droop than the standalone case. However, if the TSV pitch is close in value to that of the C4 bumps, the results may be worse. While bridge-chip based interconnection platforms present PDN challenges, especially to the active die regions that overlap with the bridge-chips, results suggest minimizing this overlap region and using multiple bridge-chips instead of a single large bridge-chip to mitigate PSN. Moreover, we perform PDN analysis including a PDN in the bridge-chip. We perform three case studies on two different configurations; the studies are (1) inclusion of ground network in the bridge-chip, (2) inclusion of power and ground network in the bridge-chip, and (3) inclusion of MIM capacitors in the bridge-chip. Besides the CPU-FPGA integration, we also study a stacked memory-FPGA configuration. Moreover, we perform a power delivery network analysis for different benchmark configurations including voltage regulator modules. Multiple on-package VRMs, 3-D IC chip-on-VRM, and backside-of-the-package VRM cases are studied. The latter two cases enable supplying power directly from the bottom of the chip. Because of the proximity from the power supply to the active circuitry, the power delivery noise of the 3-D IC chip-on-VRM case and the backside-of-the-package VRM case are the least. With distributed on-chip decoupling capacitors and package-level discrete decaps, the PSN is minimized in all the configurations. The impact of on-chip decap density variation is also quantified. For 3-D IC chip-on-VRM case with uniform current density, 25% improvement in PSN is possible if three times more decap is used compared to the one used for this analysis. Thermal implications of different

configurations are evaluated using ANSYS. Despite the 3-D nature, owing to the low power VRM die, the temperature distribution in the 3-D IC case is comparable to the on-package VRMs cases. Finally, power delivery limits of different configurations are also analyzed. The 3-D IC chip-on-VRM case and the backside-of-the-package VRM case are relatively less sensitive to the hotspots compared to the other configurations discussed in this chapter.

CHAPTER 4
BENCHMARKING POWER DELIVERY NETWORKS FOR FAN-OUT WAFER LEVEL PACKAGING (FOWLP) TECHNOLOGIES

Fan-out wafer level packaging (FOWLP) technologies have gained significant attention, especially in the low power computing space. Fig. 4.1 demonstrates an example of a heterogenously integrated system that integrates a wide number of functionalities including, but not limited to, stacked memories, RF devices, application processors, MEMS, power management ICs, etc. These state-of-the-art applications demand smaller form factors, lower power/signaling losses, and stricter resource requirements (metals, capacitors, etc.). While transistors continue to shrink, the limited scalability of traditional organic package substrates is the bottleneck for such a multi-functional integration scheme. For low power applications, a lot of these dice are available in FOWLP. As such, a power delivery network (PDN) model must reflect the unique features of this technologies for accurate modeling. Recent work has addressed some power integrity modeling aspects of FOWLP [**ase:ectc2017** , 29]. However, a detailed analysis including distributed on-die and on-package PDN models with comprehensive design space exploration is missing in the literature. In this chapter, based on prior PDN modeling efforts for 2.5-D and 3-D ICs [109, 5, 105], we propose and analyze a PDN modeling framework for steady-state and

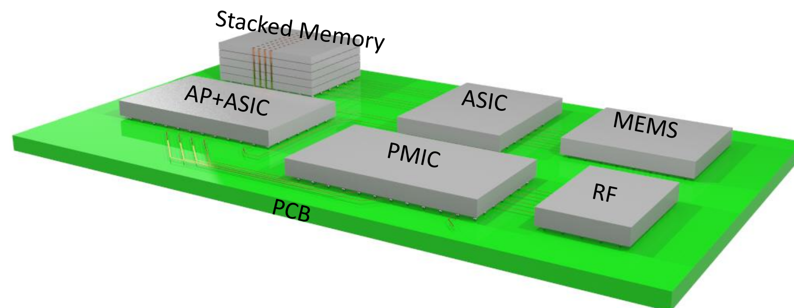


Figure 4.1: Packaging trend including FOWLP technology

transient-state analysis to evaluate and benchmark different FOWLP technologies.

4.1 Modeling Framework

4.1.1 Simulation Configurations

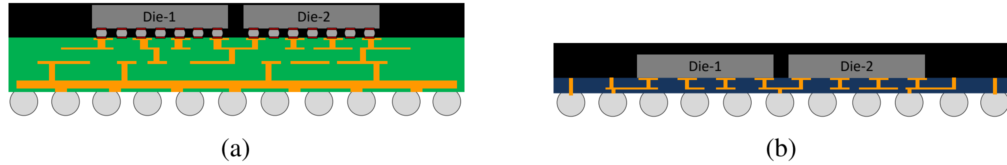


Figure 4.2: (a) Conventional multi-die flip-chip configuration and (b) Conventional multi-die FOWLP configuration

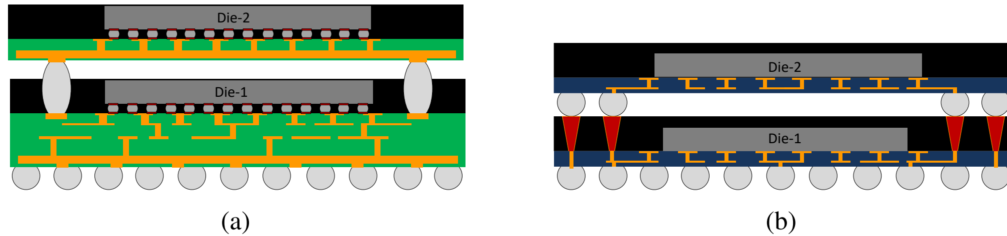


Figure 4.3: (a) 3-D flip-chip POP configuration and (b) 3-D FOWLP POP configuration

Fig. 4.2 and Fig. 4.3 present the simulation configurations under consideration in this chapter. Fig. 4.2(a) and 4.2(b) illustrate the conventional multi-die flip-chip configuration with organic package and multi-die FOWLP configuration with fine-pitch RDLs, respectively. Likewise, Fig. 4.3(a) and 4.3(b) present the flip-chip POP and 3-D FOWLP configurations, respectively. Each configuration consists of two dice. Table 4.1 presents the specifications of the two dice under consideration. The power consumption of Die-1 and Die-2 is assumed to be 3 W and 2 W, respectively. The power specifications of Die-1 and Die-2 are based on an ARM cortex A9 application processor[118] and a hypothetical memory/ASIC die [119, 120]. Moreover, we assume a uniform power distribution in each die. The general PDN parameters are provided in Table 4.2. For the 3-D POP configurations, flip-chip POP is formed using a solder bump-on-solder bump stack while the 3-D FOWLP stack is formed using through mold via (TMV)-solder bump stack.

4.1.2 PDN with Multiple Voltage Domain

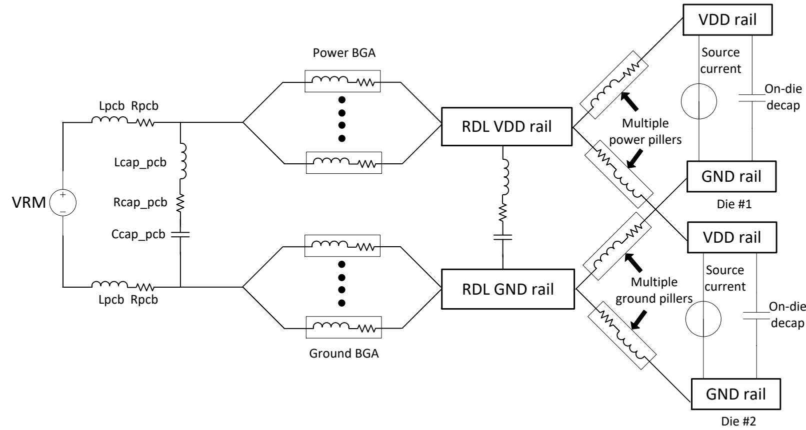


Figure 4.4: PDN structure for FOWLP technology

We present the generalized PDN structure of a FOWLP system in Fig. 4.4. As evident from the figure, there are three distinct modeling domains. We use Altera PDN tool [75] for PCB, VRM, and BGA parameters. In all our configurations discussed in this chapter, we use distributed package and on-die PDN model. The FOWLP PDN is connected to the on-die PDN using power/ground copper pillars. The flip-chip counterpart uses an organic package with power/ground planes instead of a distributed RDL in the FOWLP case.

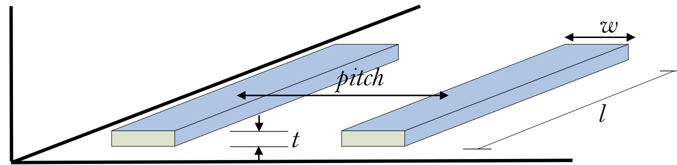


Figure 4.5: Loop inductance structure for FOWLP PDN

We assume that the package PDN in FOWLP has a similar configuration as the on-die PDN. Hence, in the distributed model, each grid is modeled as an RLC circuit. The

Table 4.1: Chip specifications

-	Voltage	Size	Power
Die-1	1.05 V	$0.5\ cm \times 1\ cm$	3 W
Die-2	1.3 V	$0.5\ cm \times 1\ cm$	2 W

Table 4.2: General parameters for PDN model

On-die global wire and FOWLP PDN **P**itch/**W**idth/**T**hickness (μm)	39.5/17.5/7
On-die decap density (nF / mm^2)	3.35
C4 bump diameter/pitch (μm)	60/130
BGA inner diameter/outer diameter/pitch (μm)	250/300/1000
PCB and VRM lumped R/L ($\mu\Omega$/pH)	1000/120
PCB decap R/L/C ($\mu\Omega$/nH/μF)	166/19.54/240

resistance of each grid is calculated based on the dimensions of different metal layers and meshing information. We use analytical formulae to calculate the inductance of the PDN. Fig. 4.5 shows the parameters that are used for the inductance calculation. For each layer, we first compute the Geometrical Mean Distance (GMD) [121]. Based on this GMD and the analysis provided in [121, 122], we calculate the inductance. We also calculate the grid capacitance based on the parallel plate capacitor model.

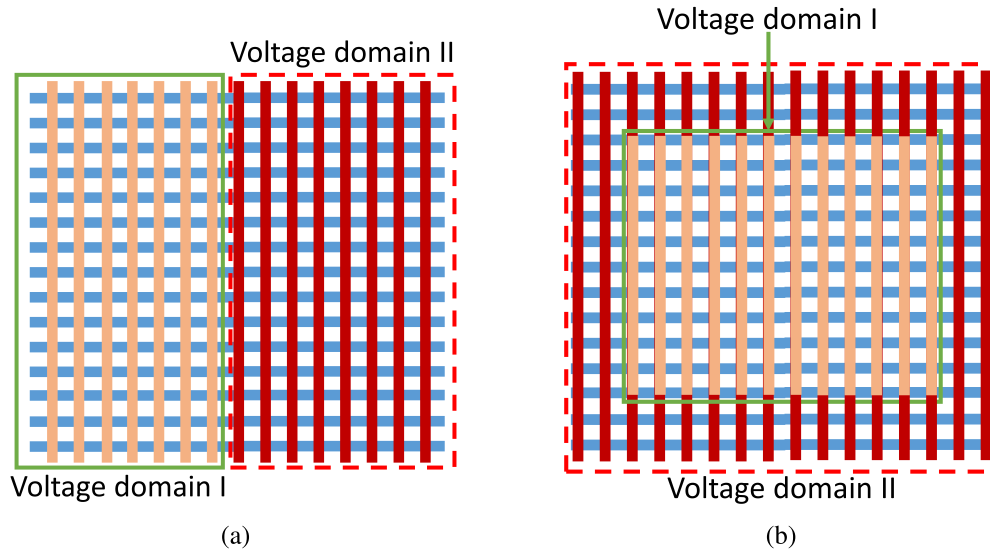


Figure 4.6: Voltage domains for each die in (a) Multi-die FOWLP and (b) 3-D FOWLP POP

As shown in Table 4.1, we consider different power supply domains for each die. We assume that there is limited utilization of on-die regulators [36] and the power in different

dice is centrally distributed from the PCB. Hence, the FOWLP PDN is split into different voltage domains to support the individual voltage requirement of each die. Fig. 4.6 provides the details related to the PDN splitting. In Fig 4.6(a), we show the PDN for a conventional multi-die FOWLP configuration for the two-die assembly under consideration. Likewise, Fig. 4.6(b) shows the PDN splitting in a 3-D FOWLP package. In the latter configuration, we assume that the top die is supplied power using the TMVs and solder bumps. These TMVs are only available surrounding the bottom die (Die-1), as shown in the figure.

4.1.3 Analysis Type

We perform time domain and frequency domain analyses. The details of each analysis are reported in the following subsections.

DC IR-drop and transient analysis

We follow a similar formulation as [5] for our steady-state IR-drop analysis and transient analysis. In the IR-drop analysis, only the resistive elements of the entire network are used. However, for the step-response based transient simultaneous switching noise (SSN) analysis, we consider the inductive and capacitive elements along with the resistive ones to characterize the power supply noise. In the latter analysis, we assume that all on-die nodes are simultaneously switching from zero current to a current value determined by the total power and the rail voltage of each die. The rise time of is assumed to be 1 ns.

Frequency domain impedance analysis

The frequency domain analysis is similar to the steady state analysis with some modifications. First, we exclude the on-die PDN from the overall network. Second, we convert the PDN into an impedance network. Third, we group together all the C4 bumps/copper pillars. Fourth, we apply an AC current source to the group of bumps/pillars. Finally, we sweep

Table 4.3: Specifications for conventional multi-die FC-BGA and FOWLP PDN modeling

Parmaters	FC-BGA	FOWLP
Package size	$2\ cm \times 1.5\ cm$	$2\ cm \times 1.5\ cm$
Package thickness [30]	250 μm	50 μm
Mold height	140 μm	140 μm
Number of chips	2	2
Solder ball height[30]	170 μm	195 μm
Number of package layers[30, 29]	4	2
Die-to-package bumps[29, 8]	C4 bumps	Copper pillars
Total thickness on top of the PCB	560 μm	385 μm

the frequency to get the desired response. For each frequency, the framework regenerates the PDN and solve for the impedance.

4.2 FOWLP Benchmarking

In this section, we discuss the design rules and present the power supply analysis results for conventional different FOWLP configurations. Moreover, we compare the analysis with traditional FC BGA packages. We refer to multi-die packages as 'FOWLP' and 'FC', respectively. For vertical stacking, we refer to configurations as '3-D FOWLP' and 'FC POP' respectively. Additionally, we analyze an additional 'baseline' case. In this design, we assume that each die under consideration has a standalone fan-out based package. Hence, there is no resource sharing in this configuration.

4.2.1 Specification

Table 4.3 provides the detailed specifications of a conventional multi-die FOWLP configuration. There are four package layers in the flip-chip configuration whereas only two layers are used in the FOWLP configuration. As shown in the table, the FOWLP package is $\sim$30% thinner than the flip-chip package under consideration. Table 4.4 provides the details of analyzed 3-D FOWLP configuration. Each package in a POP structure contains a single

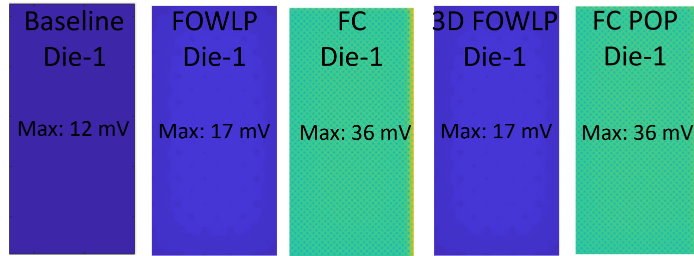
Table 4.4: Specifications for 3-D FC-POP and FOWLP-POP PDN modeling

Parameters	FC-POP	FOWLP-POP
Package size	$1.5\ cm \times 1.5\ cm$	$1.5\ cm \times 1.5\ cm$
Bottom package thickness[30]	250 μm	50 μm
Top package thickness [30]	125 μm	25 μm
Mold height	140 μm	140 μm
Number of chips	2	2
Solder ball height [30]	170 μm	195 μm
Number of bottom package layers [29, 8]	4	2
Number of top package layers [29, 8]	2	1
Chip-to-package bumps [30, 29]	C4 bumps	Copper pillars
Total thickness on top of the PCB	995 μm	745 μm

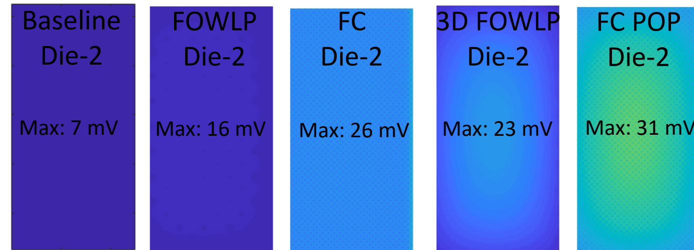
die. As mentioned in Section II, the bottom package is split into two power supply domains such that the PDN surrounding the bottom chip has a power supply domain corresponding to the top die. The top package has two metal layers in the flip-chip case and one metal layer in the FOWLP case, respectively. Compared to the FC POP stack, the FOWLP POP stack is $\sim$25% thinner.

4.2.2 PDN Analysis Results

Fig. 4.7 presents the DC IR-drop results for each configuration. Between multi-die FOWLP and FC configurations, Die-1 has the maximum IR-drop in both cases. This can be attributed to the higher current requirement of the die owing to the higher power and lower rail voltage. We observe more than 50% reduction in the IR-drop for die-1 in the FOWLP configuration compared to its flip chip counterpart. The IR-drop reduction in the die-2 is also a significant 38%. Unlike the results of the multi-die package, we see that the lower-



(a)



(b)

Figure 4.7: DC IR-drop results for (a) Die-1 and (b) Die-2 for baseline, multi-die FOWLP, multi-die FC, 3-D FOWLP, FC POP configurations

power die has higher IR-drop in both 3-D configurations. In both FC POP and 3-D FOWLP configurations, the PDN path to the lower power die (Die-2, i.e., top-most die) consists of package-to-package interconnections (solder bump-to-solder bump in FC POP and TMV-to-solder bump in 3-D FOWLP). Hence, the PDN-path for Die-2 is more critical compared to that of Die-1. Compared to FC POP case, we can see more than 50% reduction in Die-1 and 25% reduction in Die-2 in the 3-D FOWLP case. This pattern is very much dependent on the different package configurations under consideration and the number of BGAs allocated to each die (recall, Die-1 and Die-2 have separate voltage domains), etc. In the baseline configuration where each die is a standalone fan-out based configuration, we report the best achievable IR-drop for each die. IR-drop results for FOWLP configurations are closer to these hypothetical lower limits than the FC configurations under consideration.

Fine pitch RDL in FOWLP technology can also increase the interface bandwidth in a multi-die package. Fig. 4.8 shows the frequency domain impedance analysis results seen from the package-to-die connections excluding the on-die PDN. Fig. 4.8(a) and 4.8(b) show the impedance response for Die-1 and Die-2, respectively. It is evident from the

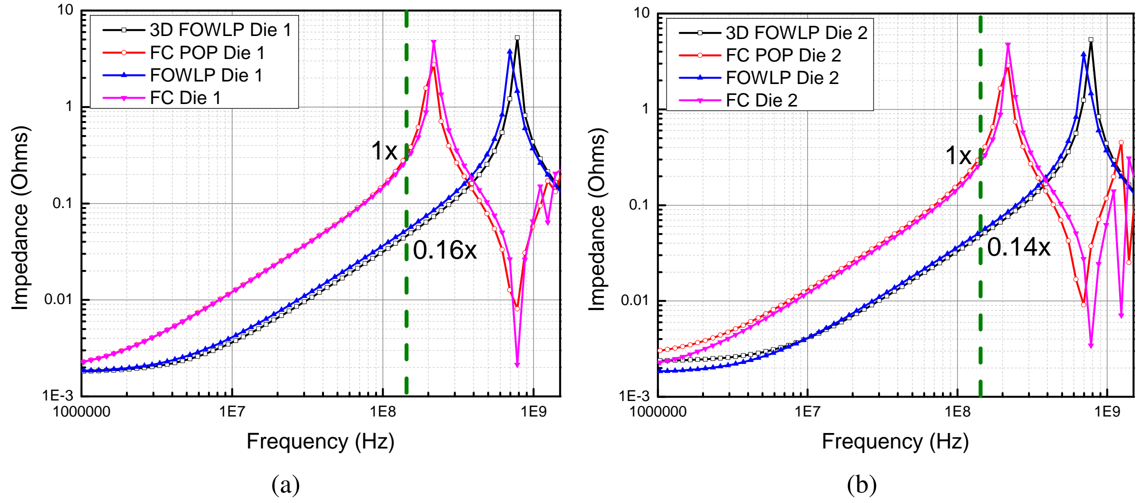


Figure 4.8: Impedance analysis results (a) Die-1 and (b) Die-2 for FC, FOWLP, FC POP, and 3-D FOWLP configurations

figure that there is a significant reduction in the PDN impedance for a wide range of frequencies. For example at 150 MHz, both dice in FOWLP and 3-D FOWLP configurations have impedance $\sim 6\times$ lower than that of the FC based configurations. The impact is a little more pronounced in the 3-D FOWLP case since the Die-2 network is more critical than the multi-die FOWLP configuration. This sort of low PDN impedance can lead to lower power required for transistor switching and hence, better power integrity.

We also characterized the FOWLP and FC configurations for transient SSN. Fig. 4.9(a) shows the results for both FC and FOWLP multi-die configurations. For both Die-1 and Die-2, there is a 17% and 22% reduction in PSN in the FOWLP case compared to the FC case, respectively. This reduction can be attributed to the lower package inductance in the FOWLP technology along with lower inductance of the package-to-die interconnections. Additionally, the plated through hole vias in the organic packages are inductive as well. Owing to the thinner package (absence of the package core), there is no need for these vias in the FOWLP technologies. Hence, for a combination of all these reasons, the power integrity performance of the FOWLP configurations are superior to the FC BGA configurations. Compared to the baseline configuration, the FOWLP case provides $1.1\times$ and $1.9\times$ PSN ($1.4\times$ and $2.3\times$ for FC configurations) for Die-1 and Die-2, respectively. Fig. 4.9(b)

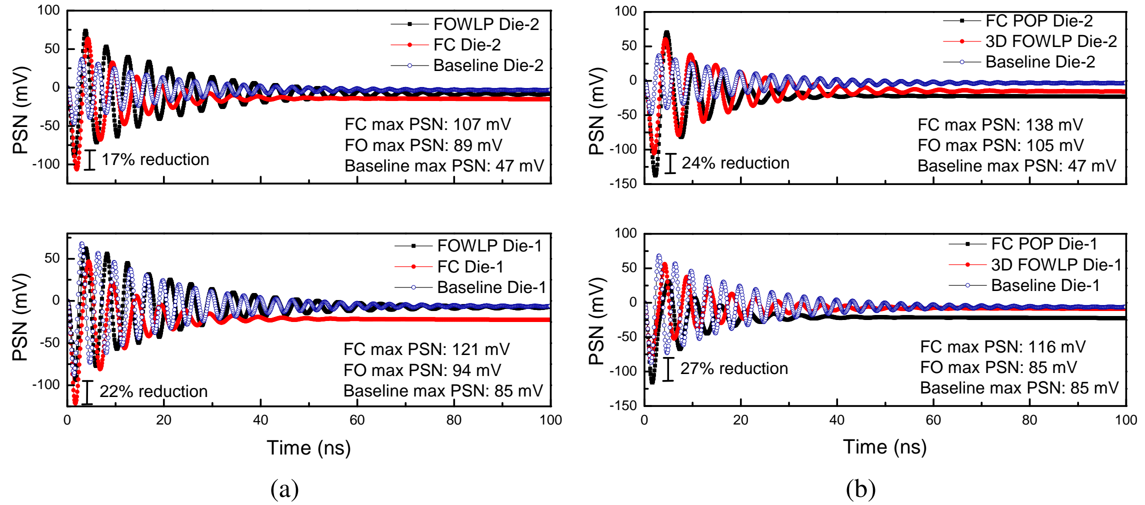


Figure 4.9: Simultaneous switching noise based transient analysis results for (a) multi-die package and (b) 3-D package-on-package configurations. Each figure shows PSN results for both Die-1 and Die-2. The baseline configuration is a single die in a single package case.

presents the transient analysis results for different POP packages under consideration. As evident from the figure, there is a significant reduction in the transient PSN for FOWLP configurations. In the FC POP configurations, there are two organic packages each contributing to the parasitics in the PDN path. Moreover, there are more bump parasitics in series in the PDN path. The reduction of the package thickness in FOWLP configurations reduces the PDN parasitics even more compared to the FC POP configurations. Moreover, TMVs are denser than solder bumps which reduces the effective parasitics of these components. All these different contributing factors result in a lower PSN in the FOWLP POP configurations. Between multi-die FOWLP and 3-D FOWLP configurations, we observe a 9 % reduction in Die-1 PSN whereas a 18 % increase in Die-2 PSN for 3-D FOWLP configuration with respect to the multi-die FOWLP configuration. However, owing to the vertical stacking, 3-D FOWLP configurations might reduce footprint while increasing the thickness of the stack. It is up to the designers to consider these different design rules to perform a trade-off analysis suitable for a specific design.

4.3 Design Space Exploration of Fan-out Wafer Level Technology

In this section, we perform a comprehensive design space exploration of power delivery in FOWLP and 3-D FOWLP technologies. This includes impacts of solder bump distribution, RDL resistivity, copper pillar pitch, RDL distribution, through mold via distribution. Finally, we show a comparison of PSN in FOWLP, FC POP, 3-D FOWLP, and TSV based 3-D IC configurations.

4.3.1 Impact of Solder Bump Distribution

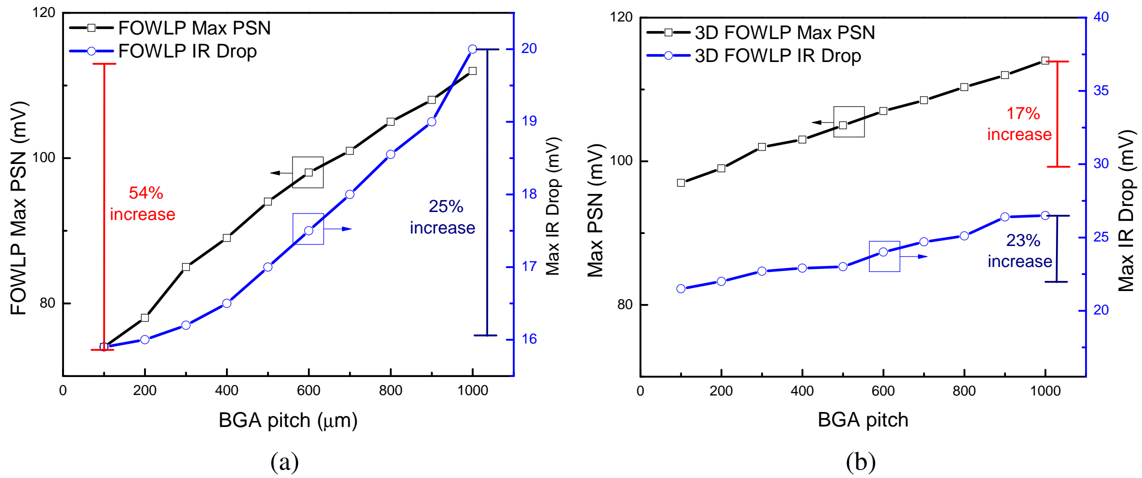


Figure 4.10: Impact of solder bump pitch on (a) multi-die FOWLP and (b) 3-D FOWLP configurations. In both cases, we report the worst-case scenario; we report Die-1 results for multi-die FOWLP and Die-2 results for 3-D FOWLP configurations.

Thus far, we have considered a solder bump pitch of 500 μm in this chapter. In this study, we sweep the solder bump pitch from 100 μm to 1000 μm. For each BGA pitch lower than the baseline case, there are more bumps available for power delivery. Fig. 4.10(a) reports the IR-drop and transient analysis results for conventional multi-die FOWLP configuration. If we increase solder bump pitch from 100 μm to 1000 μm, there is a 54% variation in the transient PSN. Moreover, our DC IR-drop analysis shows a 25% variation for a similar change in the solder bump pitch. For 3-D FOWLP configurations, we assume that different bump pitches change the bump distribution in both tiers. From Fig. 4.10(b),

we can see that across different bump pitches, there is ∼20% variation in both DC IR-drop and the transient SSN.

4.3.2 Impact of RDL Density

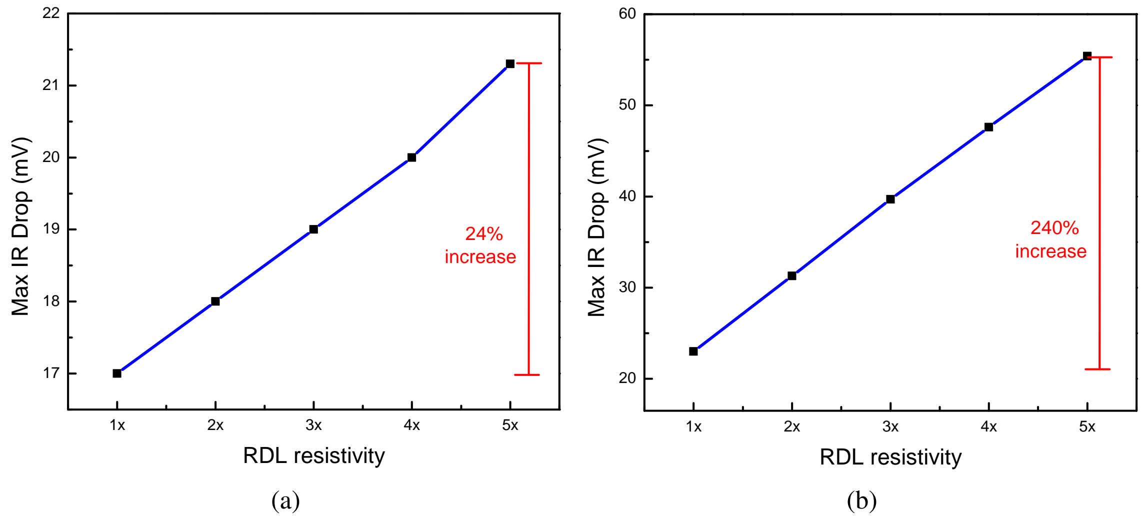


Figure 4.11: Impact of RDL resistivity on PSN for (a) multi-die FOWLP and (b) 3-D FOWLP configurations.

High density RDL is one of the key advantages in fan-out packages. There are two processes which impose restrictions on RDL density in a FOWLP package: Mold-first and RDL-first. Each process has inherent advantages over the other one. However, one key advantage of the RDL-first process is that the RDL density can be extremely high compared to the Mold-first process. These different processes change the effective sheet resistance and hence, the effective resistance of the RDLs. Moreover, scaling technology nodes can significantly increase the PDN resistivity [123]. Throughout the chapter, we considered a high density RDL process where the resistivity of the RDLs is equal to the resistivity of copper. In this study, to reflect the change in RDL density, we sweep the RDL resistivity from the baseline value to 5 times the baseline value. Fig. 4.11(a) shows the results for this analysis. We present results from the steady state IR-drop perspective. Our analysis shows that if the RDL resistivity increases by 5x, there is a 24% increase in the maximum IR-drop

in FOWLP configuration under consideration. Fig. 4.11(b) shows similar analysis results for a 3-D FOWLP configuration. Our analysis shows that if the RDL resistivity increases by 5x, the IR-drop increases significantly in the Die-2. The solder bumps deliver power to the top die from the periphery. The RDL in the top package plays a significant role in the PDN path impedance. Hence, we see such impact of RDL resistivity on 3-D FOWLP PSN.

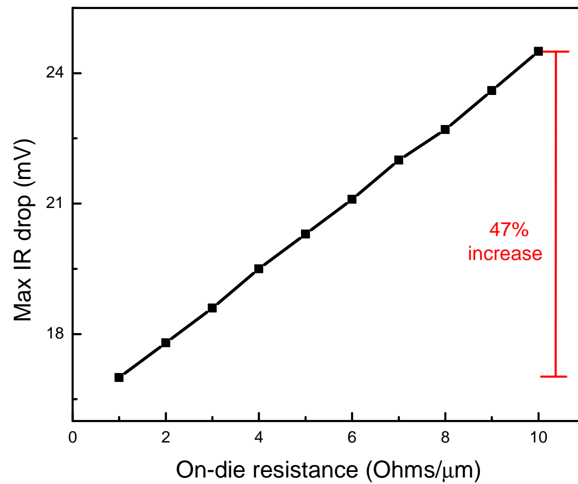


Figure 4.12: Impact of on-die PDN resistance on FOWLP supply noise

We also varied the on-die PDN resistance from our baseline value of 1 Ω/μm to 10 Ω/μm. We believe both multi-die and 3-D FOWLP configurations will follow a similar trend and hence, we only present the results for a multi-die FOWLP package. The results are summarized in Fig. 4.12. For a full range sweep, we observe a 47% variation in the maximum IR-drop seen in the dice.

4.3.3 Impact of Copper Pillar Pitch

In our initial design, we use 40 μm pitch for copper pillars connecting the package RDL and the on-die PDN. In this analysis, we sweep this parameter from 20 μm to 120μm. As we increase the copper pillar pitch, the more resistive on-die PDN contributes additional supply noise. As we increase this pitch similar to the flip-chip case, we see a similar DC IR-drop in the dice as we observed in the FC case. This proves the advantage of the dense

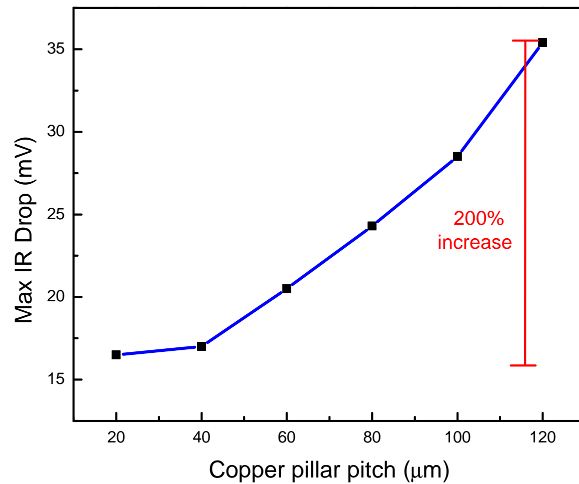


Figure 4.13: Impact of RDL to on-die PDN connectivity both FOWLP supply noise

copper pillars in the FOWLP technologies.

4.3.4 Double-sided RDL in 3-D FOWLP Technology

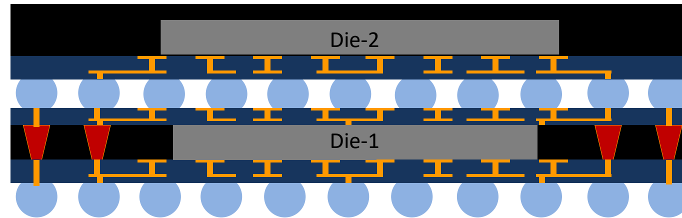


Figure 4.14: Double sided RDL in FOWLP POP structure

In a typical FOWLP POP structure, there are TMV + solder bumps connecting the top layer of the bottom package to the bottom layer of the top package. Hence, the solder bump distribution for the top package is an area-array distribution. The top package spreads the current towards the chip from these peripheral bumps. There have been research efforts [124] focused on making double sided RDLs for a FOWLP POP application. Fig. 4.14 presents such a configuration. In such configurations, the TMVs conduct the necessary current from the bottom package. These current spreads in the top layer of the bottom package located on top of the Die-1. Owing to the fine pitch RDLs, the solder bumps of the top package can be uniformly distributed. There are two advantages of a configuration of

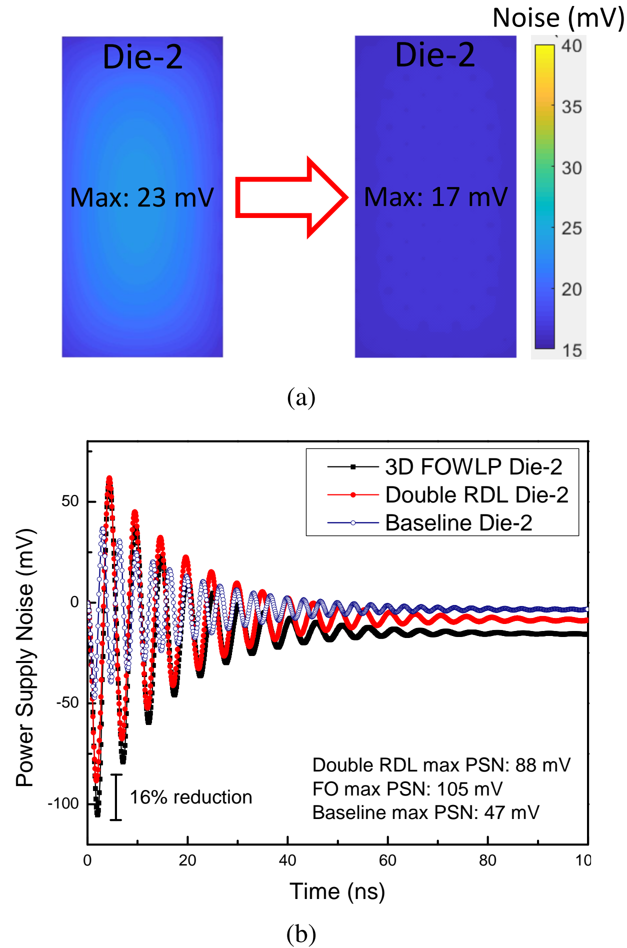


Figure 4.15: (a) IR-drop analysis and (b) transient analysis results for Double sided RDL 3-D FOWLP structures. The figures show a comparison between Die-2 PSNs for 3-D FOWLP and 3-D double sided RDL configurations, respectively.

this sort. First, there are more RDLs in parallel for the spread of current in the top package which reduces the effective resistance and inductance. Second, there are more number of bumps for the top package. This reduces the bump parasitics for the top package further. As a result of this reduction in parasitics, there is less IR-drop and transient noise induced in Die-2. Fig. 4.15 summarizes the results for this configuration. Compared to a typical FOWLP POP configuration discussed in this chapter, the IR-drop is reduced by 25% as shown in Fig. 4.15(a). In terms of transient noise, there is a 16% reduction if a FOWLP package has double RDLs in the bottom package. Fig. 4.15(b) presents this result. We assume that the additional RDL in the bottom package only changes the PDN configuration

of the top package. Hence, we have not seen any significant change in the PSN results of the bottom die (Die-1).

4.3.5 Through Mold Via Distribution

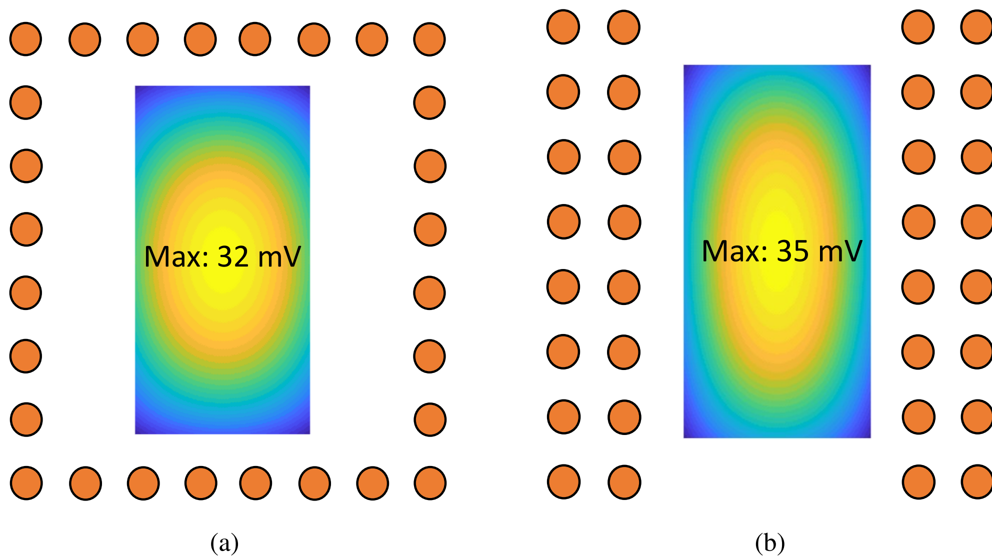


Figure 4.16: Different TMV-BGA distribution for the top die in 3-D FOWLP configurations: (a) Single line BGA+TMVs and (b) Dual-line BGA+TMVs

We studied the impact of different TMV distributions. For this study, we allocated 30% of the solder bumps to the top die for a FOWLP POP configuration. Each bump is connected to a bundle of TMVs for package-to-package interconnection. We looked at two specific solder bump+TMV distributions as shown in Fig. 4.16. In the first scenario, the interconnections are distributed along the periphery of the top package. In the second scenario, the bumps are distributed only at the two opposing sides. In each case, there is more than 50% increase in the PSN compared to the baseline design with uniform split of the bumps in the bottom package. Compared to the PSN in the first scenario as in Fig. 4.16(a), the PSN in the second scenario (Fig. 4.16(b)) is $\sim$10% higher. This can be attributed to the reduction of effective PDN path from the bumps to the active circuitry.

4.3.6 Comparison Between 3-D FOWLP, FC POP, and 3-D IC with FOWLP

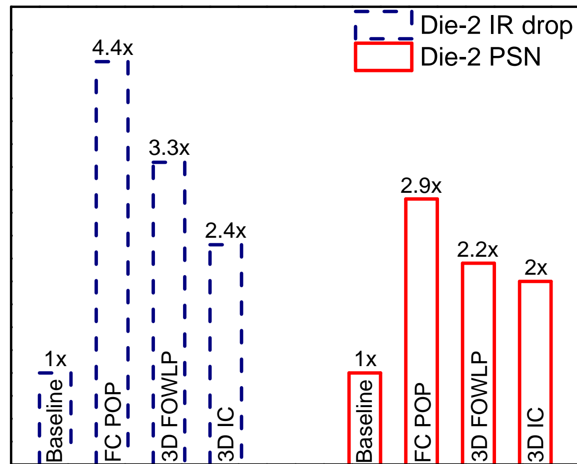


Figure 4.17: Power supply noise comparison for different integration technologies

In this segment, we analyze a thorugh silicon via (TSV) based 3-D IC stacking of the dice under consideration. We assume that the Die-1 has TSVs delivering power to Die-2. Moreover, we assume that the 3-D IC stack is bonded with a fan-out package, and the FOWLP based 3-D IC is balled to the PCB. Fig. 6.9 shows the PSN results for this analysis. We compare this 3-D IC configuration with FC POP, 3-D FOWLP, and baseline configurations. Recall, baseline configuration is a single die package with similar package size as 3-D FOWLP configuration. We normalize the results with respect to the baseline case. Since Die-2 has longer parasitic PDN path, we exclude Die-1 from this analysis. As shown in the figure, 3-D IC configuration provides 27 % and 45 % lower IR-drop compared to 3-D FOWLP and FC POP configurations, respectively. However, our analysis shows that from the transient SSN perspective, 3-D IC case provides slightly smaller (<10 %) first droop as we obtain from the 3-D FOWLP case. Hence for specific applications, it is up to the designers to decide whether to pursue 3-D IC stacking with its inherent manufacturing complexities, instead of 3-D FOWLP configurations.

4.4 Conclusion

In this chapter, we present a framework for analyzing power delivery networks in Fan-out Wafer Level Packages. Since the fan-out packages have fine pitch RDLs, we model these packages in an on-die PDN fashion. We analyze both conventional multi-die FOWLP and 3-D FOWLP package-on-package structures. For each FOWLP configuration, we compare the results with its flip chip based counterpart. We perform three different types of analysis: steady state IR-drop, frequency domain impedance, and transient analysis. Our results indicate that, owing to the shorter interconnection in the FOWLP configurations, the power supply noise decreases significantly. On average, we show close to a $\sim$20 % reduction for the conventional multi-die FOWLP packages and more than $\sim$30 % reduction for the 3-D FOWLP structures. We also perform sensitivity analysis of the FOWLP packages on different system level parameters. It is evident from our results that if a FOWLP package uses tighter pitch BGA, there will be significant reduction in PSN for both types of FOWLP packages. Having modified our framework to analyze a double sided RDL configuration for FOWLP POP packages, we present $\sim$20 % reduction in PSN for this kind of structures. The impact of the TMV distribution in the top package is noteworthy. Our results on different RDL resistivity indicates that 3-D FOWLP technologies are more sensitive to this parameter than the multi-die FOWLP configurations.

CHAPTER 5

POWER DELIVERY NETWORK (PDN) MODELING FOR BACKSIDE-PDN CONFIGURATIONS WITH BURIED POWER RAILS AND μTSVS

Backside PDN configuration attempts to tackle some of the PDN challenges by separating the on-die PDN from the conventional back-end-of-the-line (BEOL) [125]. This approach is a complete redesign of existing architectures in that both sides of the silicon have metallization layers. Moreover, alternative metallization is considered for the bottom-most metals to tackle the resistivity challenges. Therefore, in this chapter, based on our prior PDN modeling techniques [5, 96, 109], we develop a framework to analyze power supply noise (PSN) in a backside PDN configuration. Furthermore, using this framework, we benchmark the backside PDN configuration with respect to a conventional BEOL PDN to identify unique opportunities as well as limitations of this approach.

The chapter is organized as follows: in Section II, we introduce the differences between backside and conventional front-side PDN configurations. In Section III, we evaluate the power delivery performance of a backside PDN configuration. We present results for different power maps and compare modeling results with physical design results. In Section IV, we perform a design space exploration; we analyze impacts of package-to-die interconnection pitch, input pulse, capacitor density on PDN performance. Additionally, we investigate the thermal implication of dielectric bonding for a backside PDN configuration.

5.1 Modeling Framework and Specifications

5.1.1 Simulation Configurations

In a conventional PDN approach, as shown in Fig. 5.1, die VDD/VSS I/Os Interconnect to the global metal PDNs in the BEOL. Next, the global PDN connect to the local PDN

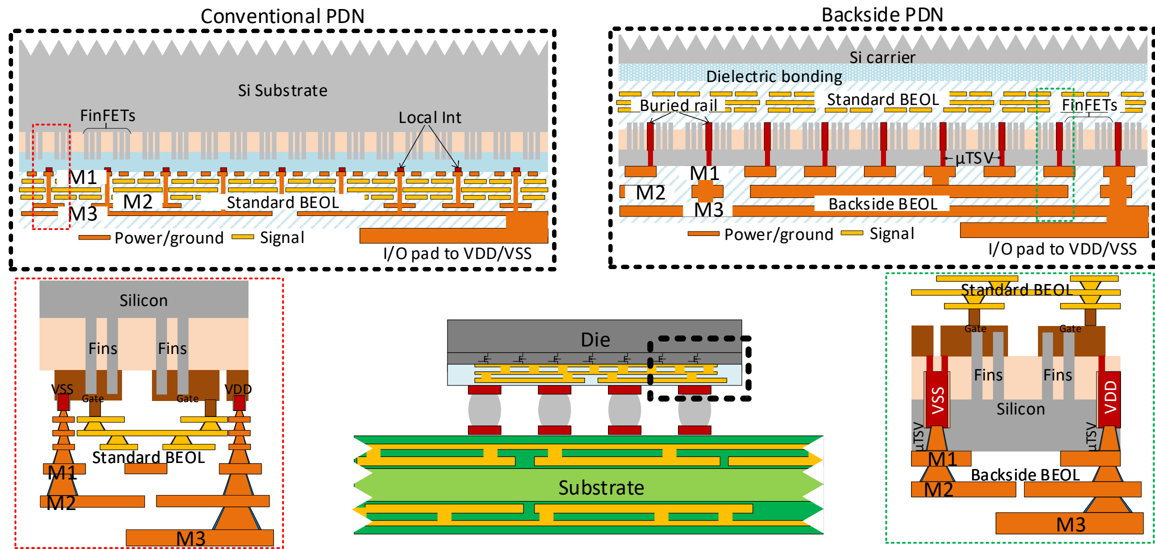


Figure 5.1: Die placement and metal configurations for a conventional front-side PDN configuration and backside PDN configuration

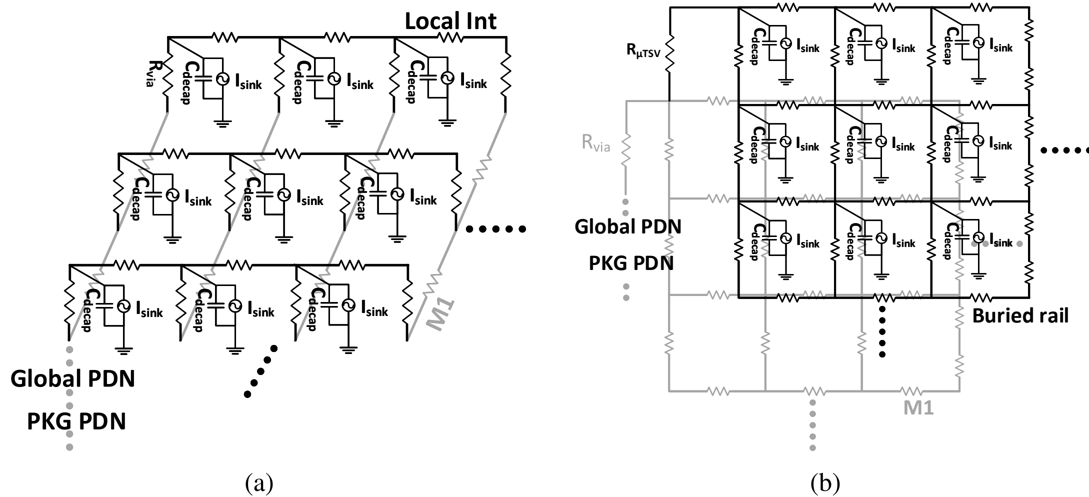


Figure 5.2: On-die PDN structure for (a) conventional interleaved BEOL PDN configuration and (b) meshed backside PDN configuration

(bottom metal layers and local interconnects) using inter-metal vias [126]. Fig. 5.1 also presents a backside PDN configuration that we explore in this chapter. The figure shows the die placement on the package and a detailed segment of the die highlighting the supply (VDD or VSS) I/Os and the PDN. There are two types of metallizations in this configuration. The conventional BEOL is located on the front side of the die and is directly con-

nected to the front-end-of-the-line (FEOL). This front-side BEOL interconnect network is primarily dominated by the signaling network. In the proposed approach, the backside metallization of the die is dedicated to the PDNs. Additionally, this configuration has a buried power rail (BPR) network within the FEOL to locally supply power to active circuits. BPR is interconnected to the backside PDN using μTSVs.

We consider three separate PDN domains in the modeling [5, 105]: on-die, package, and board level PDNs. In the backside PDN configuration, we assume a re-design of the on-die PDNs. Package and board domains remain similar in both configurations. Fig. 5.2 presents the structure of an on-die PDN. The conventional PDN (within BEOL) is an interleaved structure, as shown in Fig. 5.2(a), whereas the backside PDN is a mesh-like network. Fig. 5.2(b) shows the on-die PDN for a backside PDN configuration. The bottom two metal layers in each configuration are connected by different via resistances: dense inter-metal via stack in the conventional BEOL configurations and μTSVs in the backside configurations. Moreover, the top-most metal layer is connected to the package PDN by C4 bumps and copper pillars in the conventional BEOL PDN and backside PDN, respectively.

5.1.2 Specifications

In each configuration, we consider 3 metal layers (local interconnect/BPR, M1, and M2) for the PDN. The specifications are tabulated in Table 5.1. We consider high aspect ratio Ru based bottom-most metal layer for the backside PDN configurations [127, 128]. For advanced technology nodes, Ru provides a significant reduction in resistivity compared to other conventional metal options (Cu, Co, etc.) [74]. Moreover, in the backside PDN configuration, the package-to-die bumps are denser compared to conventional PDN; denser bumps improve the distribution of current. We use similar inductance values for package traces, microbumps, solder bumps, etc. as in our prior PDN modeling effort [5]. Additionally, the inter-metal via resistance for bottom metals is $\sim$7x smaller in the backside configuration compared to similar connections in conventional BEOL PDNs [129, 74].

Table 5.1: PDN specifications for different configurations

Parameters	Conventional PDN	Backside PDN
No. of metal layers	3	3
PDN metal	Local Int: Cu, M1: Cu, M2: Cu	BPR: Ru, M1: Cu, M2: Cu
PDN metal resistance (Ω/μm)	Local Int: 500, M1: 1, M2: 1	BPR: 50, M1: 1, M2: 1
μTSV Diameter/Height/Resistivity (nm/nm/nΩm)	N/A	50/500/80
Via resistance (Ω/via)	Local Int-M1: 160, M1-M2: 2	BPR-M1: 24, M1-M2: 2
Die-to-package bumps	Diameter: 70 μm, Height: 140 μm, Pitch: 140 μm	Diameter: 20 μm, Height: 40 μm, Pitch: 40 μm
On-die decoupling cap (nF/mm^2)	1.8	1.8
Package effective decap R/L/C (mΩ/pH/μF)	541.5/220.7/52	541.5/220.7/52
Package resistance/inductance (mΩ/mm/pH/mm)	1.2/24	1.2/24
PCB decap R/L/C ($\mu\Omega$/nH/μF)	166/19.54/240	166/19.54/240
PCB resistance/inductance ($\mu\Omega$/pH)	166/21	166/21

This reduction can be attributed to the via stack through the signaling network in a conventional BEOL [126]. We consider both uniform and hotspot based power maps for the PDN analysis. In uniform power map analysis, total die power is uniformly distributed across the die. In hotspot power density maps, some areas of the die consume significantly more power than the rest of the die. For the uniform power map case, we consider 74.49 W/cm^2 [58] power density. Unless otherwise specified, we use this power for uniform power map analysis throughout the chapter. We assume a 5 mm $\times$ 5 mm die with 0.9 V rail voltage

throughout the chapter. We consider two types of on-die decoupling capacitors: die-level MOS caps and metal-insulator-metal (MIM) capacitors (caps) connected to the top metal layers.

5.1.3 Adaptive Meshing

We use domain specific adaptive meshing for PDN modeling; we use different grid sizes for on-die, package, and board PDNs. For hotspot power density analysis, we use within-die adaptive meshing technique. In our prior PDN modeling, we have used package-to-die bump granularity for meshing [5]. In within-die adaptive meshing, we use denser grids in the hotspot regions while using coarse grids for the rest of the die. For example, if we have a 100 μm $\times$ 100 μm hotspot in a 5 mm $\times$ 5 mm diefor the front side PDN configuration, we use 1 μm grid for the hotspot and 140 μm grid for the rest of the die. Using a 1 μm grid for the whole die would yield 25 M nodes for a single layer of the PDN. On the contrary, using a 140 μm grid would compromise on accuracy of the PDN model. The adaptive meshing technique reduces the number of total mesh elements while performing a fine-grain PDN analysis for critical die blocks.

5.2 Power Delivery Network Benchmarking

In this section, we present the PSN results for both configurations. We exclude MIM caps from the analysis in this section.

5.2.1 Uniform Power Density Maps

We explore the step response based simultaneous switching noise for both configurations using a 400 ps rise time. Fig. 5.3 summarizes the results. We report the improvement in each noise droop for the backside PDN configuration with respect to its conventional front-side counterpart. From the DC IR-drop analysis, the backside PDN provides more than 4x reduction in PSN. This reduction can be attributed to the denser PDN and denser

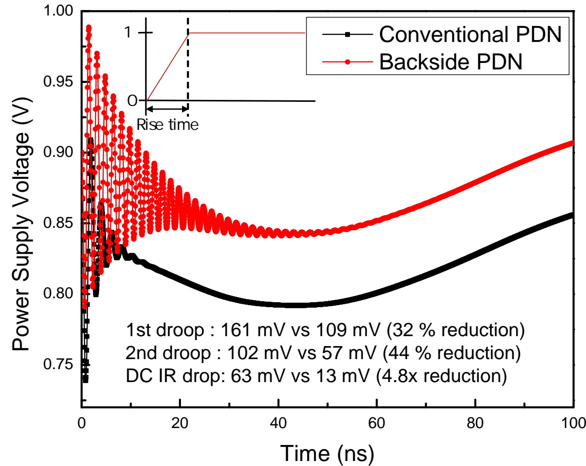
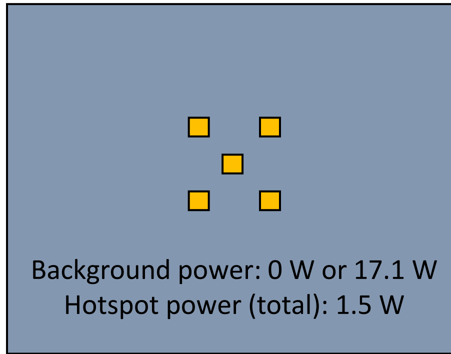


Figure 5.3: Power supply noise results for uniform power map

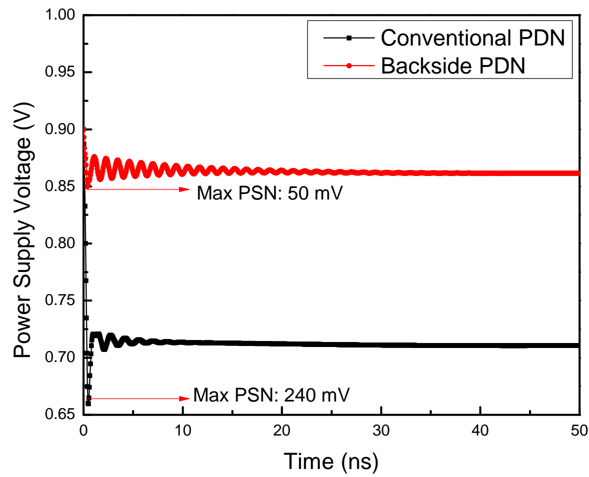
package-to-die bumps in such a configuration. Examining the inductive noise, we observe a significant 32% and a 44% reduction in first droop and second droop noise, respectively. Only the on-die PDNs are modified between two configurations. The unmodified package PDN is mostly inductive and hence, for this uniform power map case under consideration, we do not observe as much reduction in the first droop noise as the DC IR-drop noise.

5.2.2 Hotspot Power Densities

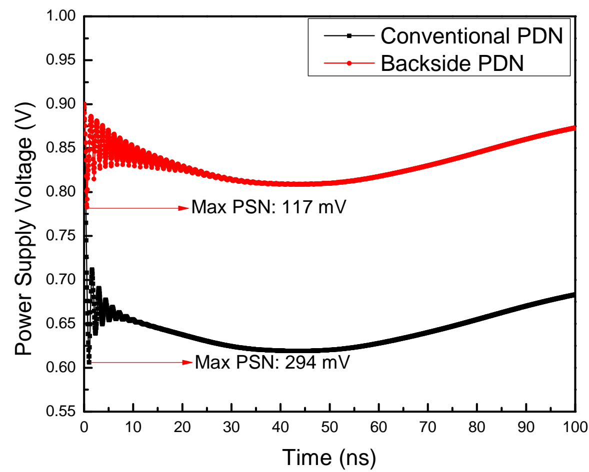
The uniform power map case emulates the worst case scenario where all the die nodes are switching simultaneously. However, this is an average power map across the die. The micro-scale circuit blocks have a higher power density compared to an average power map [5]. Moreover, we have stated previously that owing to higher densities, advanced technology nodes, such as 7 nm, 3 nm, etc., will consume significantly higher power [76]. We emulate such an aggressive micro-scale power map in Fig. 5.4(a). While we keep the die size unchanged as 5 mm $\times$ 5 mm, we assume that 0.3 W is being consumed in a 100 μm $\times$ 100 μm region. Moreover, we assume that the rest of the die is not consuming any power. For simplicity, we begin with one such hotspot and observe 32 mV and 130 mV peak IR-drop in the backside PDN and the conventional BEOL PDN configurations, respectively. This single hotspot simulation is an emulation of a standalone computing block [125, 76].



(a)



(b)



(c)

Figure 5.4: (a) Power map with five adjacent hotspots, (b) PSN results for the hotspot power map with zero background power, and (c) PSN results for the hotspot power map with 17.1 W uniform background power

To emulate a more realistic scenario, we add four additional such hotspots surrounding the first one. This equates to a total die power of 1.5 W. This case emulates multiple cores or computing blocks running in parallel. For the five hotspots case, we observe 50 mV and 240 mV peak IR-drop in the backside PDN and the conventional BEOL PDN configurations, respectively. Fig. 5.4(b) summarizes these results. For a backside PDN configuration compared to its conventional counterpart, the single hotspot and five hotspots cases provide a 4x and 5.4x reduction in PSN, respectively. The Ldi/dt noise reflects the impact of both inductive and resistive components of a network. Since hotspot power density maps create significantly higher DC IR-drop in the conventional BEOL PDN configurations, unlike the uniform power map case, we observe a similar improvement in both steady-state and transient analysis results for a backside PDN configuration. Fig. 5.4(c) shows the results for five hotspots case with a total power of 18.6 W. This case is a combination of the uniform power map and the high-density hotspot power map under consideration. In this analysis, the first droop noise is 2.5x lower in the backside PDN configuration compared to the front-side PDN case. However, similar to other case studies with different power maps, second and third droop show a similar greater than 4x reduction. In Fig. 5.4(b) and 5.4(c), the total power consumption is 1.5 W and 18.6 W, respectively. For higher power consumption, the package and board losses are significantly higher and hence, compared to Fig. 5.4(b), we observe an increase in second and third droop noises in Fig. 5.4(c). For a zero background power case, only a few on-die nodes are switching as opposed to all on-die nodes in a non-zero background power case. Hence, we do not observe greater than 4x reduction in the first droop noise as we observe in Fig. 5.4(b).

5.2.3 Physical Design Results

We perform physical design implementation for both configurations for a RISC-V architecture. Table 5.2 summarizes the results. For conventional BEOL PDNs, we implement PDNs of different densities. We use contacted poly pitch (CPP) to reflect the scaling of

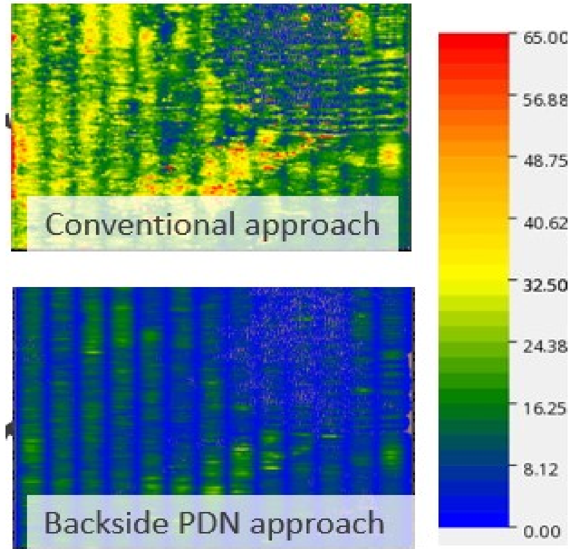


Figure 5.5: Physical design results for different PDN configurations

technology nodes [125] and hence, PDN density. Furthermore, we implement conventional BEOL PDN with BPR, backside PDN with standard power rails, and backside PDN with BPR. After placement and routing, we observe a 25%-30% area reduction in the backside PDN configurations. We normalize the peak IR-drop in each configuration to the IR-drop value for the backside PDN configuration with BPR. From physical implementations, we observe that each metric has a minimum 4x reduction for a backside PDN configuration with BPR. In Table 5.3, we compare the PDN modeling results with the physical implementation results. The conventional BEOL PDN is a dense 8 CPP design. The backside PDN configuration is an 8 CPP design with μTSVs. From physical design results, we

Table 5.2: Summary physical design results for a RISC-V architecture

Technology	PDN density	Area (μm^2)	Peak IR-drop
conventional BEOL PDN	8 CPP	8594	4x
	16 CPP	7365	7.5x
	24 CPP	6874	9x
conventional BEOL PDN + BPR	48 CPP	6874	4x
Backside PDN excluding BPR	32 CPP	6446	4x
Backside PDN + BPR	8 CPP μTSV	5926	1x

Table 5.3: Summary PSN results from physical design and PDN modeling

Configuration	PDN density	Target density (%)	**Peak IR-drop from physical design**	**Peak steady-state IR-drop from PDN modeling**	**Peak Ldi/dt noise from PDN modeling**
conventional BEOL PDN	8 CPP	65	**4x**	**4x**	**5x**
Backside PDN	8 CPP μTSV	87	**1x**	**1x**	**1x**

observe better core utilization in the backside PDN configurations. Fig. 5.5 shows the extracted IR-drop results from physical implementation for both configurations. Although the physical design is an exact architectural implementation whereas the PDN modeling framework in this chapter is abstract modeling from high-level PDN parameters, e.g., number of metal layers, dimensions of PDN metals, decoupling cap densities, etc., both in steady-state IR-drop and transient Ldi/dt noise analysis, we observe similar trends between physical implementation and PDN modeling results. Owing to the generalization in the PDN modeling with larger die size and inclusion of the package and the board PDN, we observe some discrepancy between results of the physical design and the PDN modeling framework.

5.3 Sensitivity Analysis

In this section, we show results from design space explorations to determine the limits and benefits of a backside PDN configuration.

5.3.1 Chip-to-Package Interconnection

Throughout the chapter, we assume 40 μm pitch die-to-package interconnections for backside PDNs and 140 μm pitch for conventional BEOL PDNs. In Fig. 5.6, we report results for varying pitches in each configuration. We report the maximum transient Ldi/dt noise

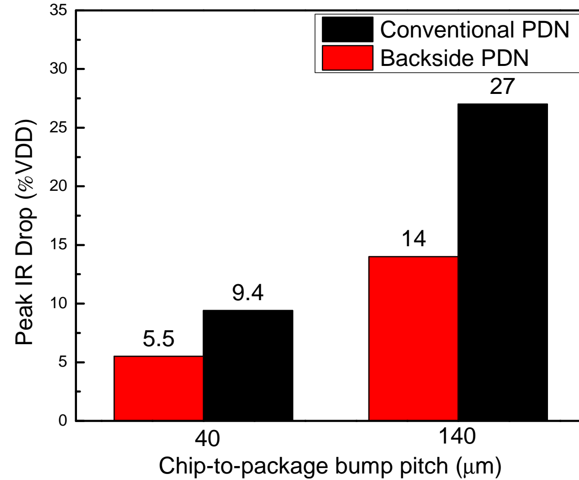


Figure 5.6: Peak IR-drop comparison for different package-to-chip bump pitches

for each variant. For each pitch value, backside PDNs provide almost 2x reduction in PSN compared to its conventional BEOL counterpart. This reduction can be attributed to the denser PDN along with lower via resistance for lowest metal layers. This essentially means better performance in the backside PDNs. Moreover, we observe that scaling I/O pitch can have a significant impact regardless of the configuration, and denser power/ground (P/G) I/Os are favorable in both cases. For both configurations, compared to 140 μm pitch, 40 μm pitch interconnections provide 2x improvement in PSN. An assembled die has successively higher resistance metal traces from board to die-level PDNs. As such, between the package PDN and the die PDN, package PDNs have lower resistance and hence, help spread the current. Denser P/G bumps enhance this spreading in the package level. Hence, we observe this improvement in PSN with respect to bump pitch reduction.

5.3.2 Impact of Input Pulse

In this analysis, we evaluate the step response by varying the rise time of the input current load. We assume the uniform power map discussed in this chapter. We sweep the rise time from 200 ps to 1 ns. Fig. 5.7 presents the results for these different rise times. For reference, we also report the step response for conventional BEOL PDN with 1 ns rise time. As expected, with increased rise time, the PSN reduces in the backside configuration. How-

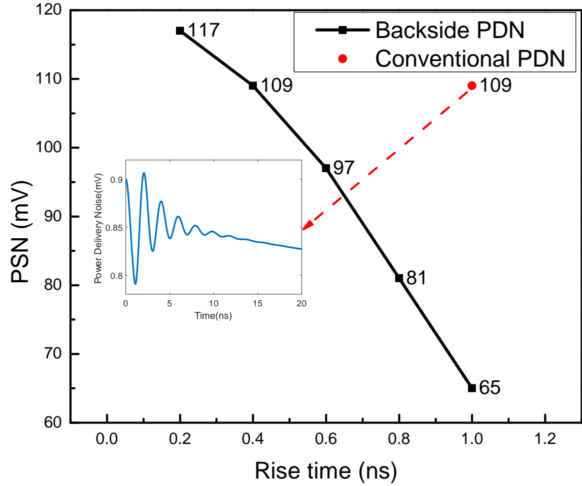


Figure 5.7: Impact of rise time variation on step response for backside PDN configuration. The red line shows the step response result for conventional BEOL PDN with 1 ns rise time

ever, between a backside PDN switching with 400 ps rise time and a conventional BEOL PDN switching with 1 ns rise time, we observe a similar supply noise. Hence, a backside PDN enables faster switching compared to a conventional front-side PDN configuration.

5.3.3 MIM Decoupling Cap Density

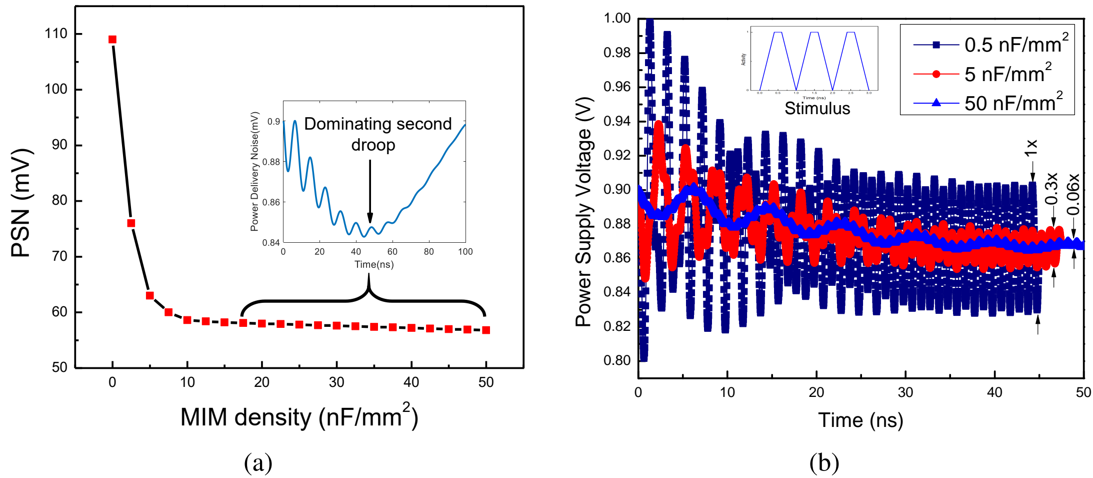


Figure 5.8: (a) Step response results for different MIM densities and (b) supply noise for 1 GHz pulse input

One additional advantage of the backside PDN configuration is to have denser MIM

caps connected to the top level metal layers. This is facilitated by the separation of signaling layers from the PDNs of the dice. Throughout the chapter, we exclude the MIM caps from the analysis. In this subsection, we investigate the impact of MIM cap density on PSN, as shown in Fig. 5.8(a). We use a uniform power map and step response based simultaneous switching noise analysis for this analysis. We vary the MIM density from 0 nF/mm^2 to 50 nF/mm^2. The 0 nF/mm^2 MIM density corresponds to the backside PDN analysis results shown in Fig. 5.3. As we increase the MIM cap density, the transient droop reduces. However, beyond a certain cap density, PSN does not improve any further. From the figure, we can see that beyond 10 nF/mm^2, the second droop begins to dominate the PSN. The inset of Fig. 5.8(a) shows the noise profile for a 50 nF/mm^2 MIM cap density. This is representative of noise profiles with MIM cap densities greater than 10 nF/mm^2.

In the second part of this analysis, we explore a different input excitation for PSN analysis. We analyze simultaneous switching noise for an input pulse with 1 GHz frequency. The input has 400 ps rise time, 200 ps conduction time, and 400 ps fall time, respectively. Fig. 5.8(b) summarizes the results for this analysis. We consider three different MIM cap densities (0.5 nF/mm^2, 5 nF/mm^2, and 50 nF/mm^2). Evidently, as we increase the MIM cap density across the die, both supply noise and supply noise fluctuation reduce significantly. As we increase MIM cap density from 0.5 nF/mm^2 to 5 nF/mm^2, the peak-to-peak fluctuation of the supply voltage reduces by more than 3x. A higher cap density may not improve the first droop noise, however, it helps reduce the high-frequency noise ripple.

5.3.4 Thermal Implications of a Backside PDN Configuration

The backside PDN configuration uses a dielectric bonding of the active layers with a carrier wafer. Compared to a conventional front-side PDN configuration, this bonding layer increases the junction-to-ambient thermal resistance. We use our thermal modeling framework [58] to evaluate the impact of this additional layer on the thermal performance of a backside PDN configuration. For this analysis, we assume a 1 cm $\times$ 1 cm die with 74.49 W

power. Although the dielectric bonding layer is typically a 1 μm$\sim$2 μm, we simulate for up to 20 μm dielectric bonding layer in the backside PDN configuration. We assume that the dielectric bonding layer has a thermal conductivity of 0.9 W/m-K which is similar to the thermal conductivity of SiO_2. Moreover, we assume an air-cooled heat sink with 0.218 ^{o}C/W case-to-ambient thermal resistance.

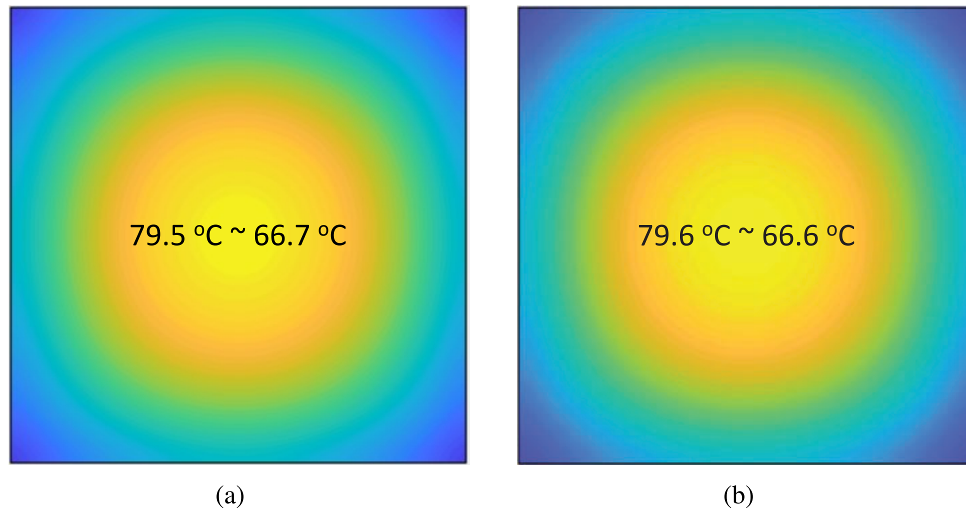


Figure 5.9: Temperature distribution for (a) conventional front-side PDN configuration and (b) backside PDN configuration

Fig. 5.9 shows the thermal results for both configurations. As evident from the figure, According to our analysis, the additional dielectric bonding layer in the backside PDN configuration has a negligible impact on temperature distribution of the dice. The conventional front-side PDN configuration and the backside PDN configuration have a maximum junction temperature of 79.5 ^{o}C and 79.6 ^{o}C, respectively.

5.4 Conclusion

In this chapter, we present a PDN modeling framework for backside PDN configurations. Backside PDN configurations are similar to double side processed dice with signaling network and power delivery network on either side of the FEOL. Owing to the denser PDN and new materials for bottom-most metal, this configuration provides significant improvement

in power supply noise reduction. We use a uniform power map to emulate a real computing block and a non-uniform power map to emulate futuristic computing blocks at advanced technology nodes. For both power maps, we observe greater than 4x reduction in power supply noise in the backside PDN configurations relative to conventional BEOL counterparts. We use physical implementation of a RISC-V architecture to validate our modeling results. Our physical design shows a 25%-30% area improvement in the backside PDN configuration compared to the conventional BEOL configurations. Our package-to-die bump pitch analysis shows at least 2x performance improvement for a backside PDN configuration over a conventional counterpart. We sweep the rise time of the input pulse. We observe that a backside PDN configuration with 400 ps rise time provides a similar noise profile of a conventional BEOL configuration with 1 ns rise time. Moreover, based on our assumptions, we observe that a MIM cap density greater than 5 nF/mm^2 does not improve the first droop noise further, however, more MIM caps can reduce high-frequency ripple for a given input pulse. Despite the inclusion of a thermally resistive bonding layer, thermal modeling results indicate that the backside PDN configuration has similar temperature distribution as a front-side PDN configuration.

CHAPTER 6

THERMAL- POWER DELIVERY NETWORK (PDN) CO-ANALYSIS OF 2.5-D INTEGRATION TECHNOLOGIES

In prior chapters, PDNs have been analyzed for different emerging heterogeneous integration technologies. However, PDN and temperature of a given configuration are interdependent. Fig. 6.1 shows the dependencies between power dissipation, temperature, and power delivery network (PDN). Without considering the interactions between each of the components in Fig. 6.1 for emerging architectures with increased power density, the results from the standalone or partially integrated models could be overestimated.

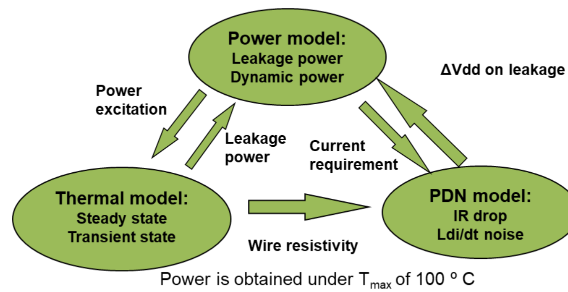


Figure 6.1: Thermal-PDN interaction models

In previous efforts [5, 7], we benchmarked our PDN and thermal models to open source IBM benchmarks and finite element based modeling using ANSYS, respectively. Moreover, we presented the PDN results for different 2.5-D integration technologies in [5] and thermal-PDN co-analysis results for 3-D stacked ICs in [7]. However, in [7], only the on-die PDN is considered. A detailed distributed package PDN model for different 2.5-D integration technologies is necessary to capture the unconventional PDN interfaces of these technologies, as shown in Chapter 3. Also, for better convergence of the simulation, the interconnect loss of a system needs to be fed back in consecutive iterations. Hence, In this chapter, we present a complete thermal-PDN co-analysis framework for multi-die pack-

ages and bridge-based technologies [6]. We also present the thermal modeling results for a bridge-chip based 2.5-D configuration. Moreover, we present the results from a thermal-PDN co-analysis perspective. We report results from both steady-state and transient-state analysis.

6.1 Thermal Evaluation of Bridge-Chip Based 2.5-D Configurations

6.1.1 Bridge-Chip Based 2.5-D Configuration

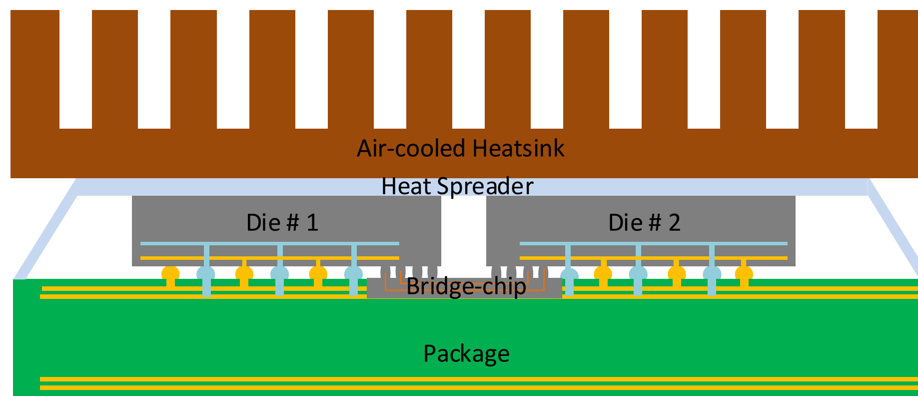


Figure 6.2: Bridge-chip based simulation configuration

Fig. 6.2 shows the bridge-chip based 2.5-D integration technology. Chip-to-chip interconnects are routed on the bridge-chip, and fine-pitch microbumps are used to connect the bridge-chip and the active dice. With this technology, 2.5-D heterogeneous integration of multifunctional chips can be realized. In this chapter, we focus on integration of processor-FPGA for high-performance computing. The FPGA and the processor are placed side-by-side on the same package with a bridge-chip underneath it.

6.1.2 Thermal Modeling Specifications

The thermal modeling framework used in this chapter is reported in [7]. The model is based on finite volume method and developed in MATLAB. The specifications for thermal modeling of a FPGA-processor integration are specified in Table I. The specifications include the thickness and thermal conductivity of different layers of the 2.5-D integration.

Table 6.1: Thermal simulation parameters

Layer	Conductivity (W/mK)		Thickness (μm)
	In-plane	Through-plane	
TIM	3		30
Heat spreader	400		1000
TIM 1	3		30
Chip-1 die	149		100
Chip-2 die	149		100
Bridge-chip	149		100
Underfill	3		N/A
Package	30.4	0.38	1000
Package-to-die bumps	60		70

The system is assumed to be air-cooled. The boundary conditions used are similar to the ones in [7]. The power maps of the emulated processor and the FPGA are given in Fig. 7, which are based on Intel Core i7 processor and Altera Stratix FPGAs [7]. The total power of the processor and the FPGA are 74.49W and 44.8W, respectively. The package-to-die connection is established using bumps. However, the die-to-bridge-chip connections are high microbumps.

6.1.3 Thermal Results

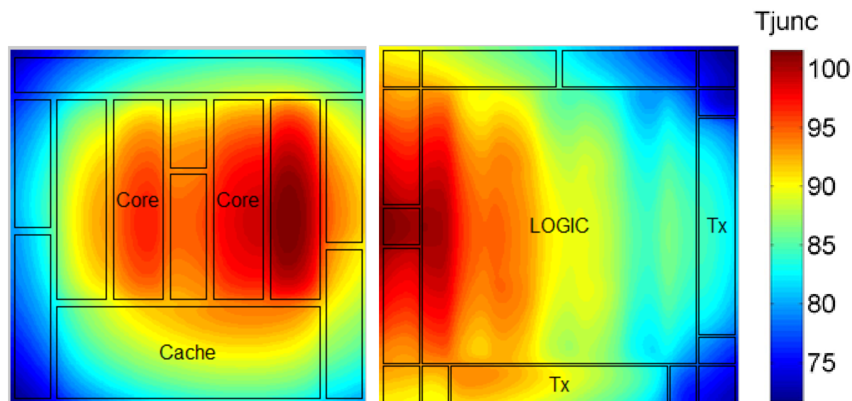


Figure 6.3: Top view of thermal profile of each die (processor-FPGA)

Fig. 6.3 shows the thermal profiles of each die from the thermal analysis. The processor

die and the FPGA die have a maximum temperature of 102°C and 89.5^{o}C, respectively. As evident from the figure, there is significant thermal coupling from the high power die to the low power die. There are two paths associated with the thermal coupling. The primary coupling path is the heat spreader atop. For the 2.5-D based integration case that we are investigating, there is also a secondary heat coupling path through the bridge-chip.

Thermal Coupling with Respect to Varying Power of The Low-Power Die

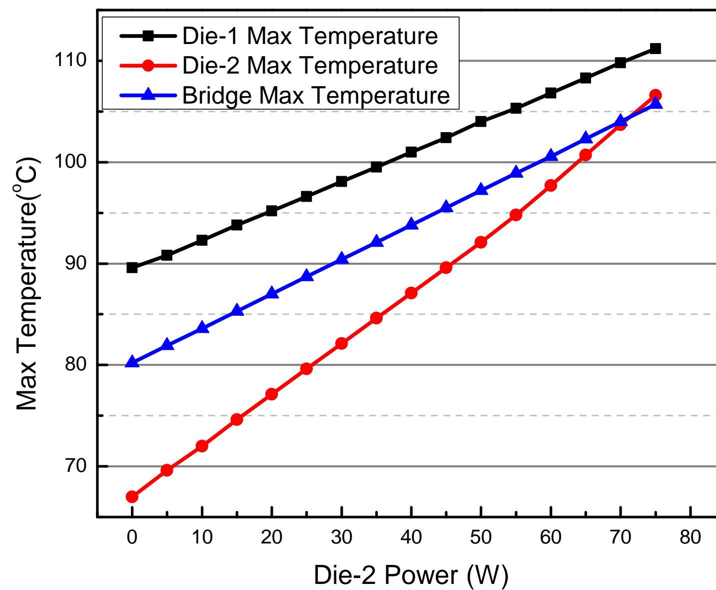
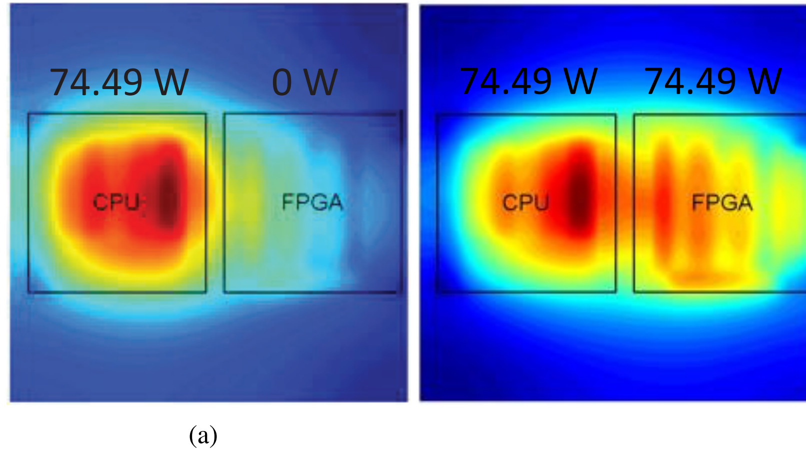


Figure 6.4: Maximum temperature of different layers with respect to the variation in low power die power

In this section, we sweep the power of Die-2 (FPGA) from 0 W to 74.5 W. Fig. 6.4 shows the maximum temperature in each die for different power levels in the FPGA die. The figure also shows the maximum temperature in the bridge-chip. With increased power, there is a linear increase in the temperature at both dice. A 0 W FPGA is an emulation of a processor-dummy die configuration. Hence, for this configuration, the temperature distribution in the FPGA is solely due to the thermal coupling through the heat spreader and the bridge-chip. Likewise, in the low power range, there is heat coupling from the processor to the FPGA. However, in the high power range, the bridge-chip helps spread power in

both directions. Especially, when both dice have similar power levels, hotspots in the die power map will dictate the spreading of power. We can see in the figure that the maximum temperature in the bridge-chip is rising above the temperature of Die-2 temperature after ∼70 W. Fig. 6.5 reports the package thermal profile of two different power processor-



(a)

Figure 6.5: Thermal profile of (a) 74 W processor – 44.8 W FPGA integration and (b) 74 W processor – 74 W FPGA integration

FPGA simulations. The bounding box defines the boundary of each die. The bridge-chip is located between these two bounding boxes. We can see in the figure that, depending on the FPGA power, there is significant thermal coupling and heat spreading in the bridge-chip.

6.2 Thermal PDN Co-Analysis for Bridge-Chip Based 2.5-D Configuration

6.2.1 Steady-state IR-drop Modeling Framework

In Fig. 6.6, we present the proposed modeling framework for steady-state analysis. We begin thermal and PDN simulations with a reference power for each die estimated from an architectural tool [118]. Moreover, we use HSPICE to estimate the temperature and supply voltage dependencies of the leakage power. In the subsequent iterations, the power dissipation is updated by the power models that use the updated temperature and supply voltage values. At the end of the simulations, the power dissipation, temperature distribution, and the supply noise of each die become consistent with each other within our interaction mod-

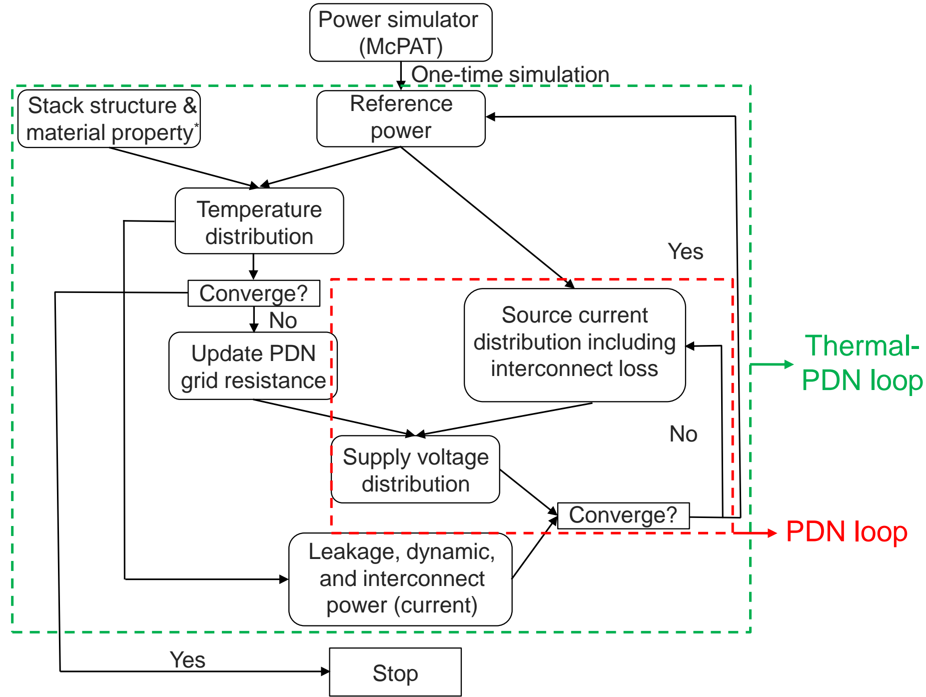


Figure 6.6: The flow chart for the thermal-PDN co-analysis

els [7]. We consider two different thermal effects, as shown in the figure. First, the power estimation of a die from an architectural tool or HSPICE simulations is temperature dependent; the outer path in Fig. 6.6 accounts for this effect. Second, there is self-heating of the PDN where temperature changes the PDN resistivity. Additionally, we included a distributed package model in our co-analysis framework to incorporate irregular packaging structures owing to emerging advanced packaging technologies. In our two die package, Die #1 and Die #2 emulate a 14 nm FPGA die with peak total power of 44.8 W [102] and a 22 nm processor die with peak total power of 74.49 W [5], respectively. We assume uniform power map for both dice with a supply voltage of 0.9 V. Both dice are assumed to be $1\ cm \times 1\ cm$ and are placed side-by-side with a die spacing of 0.5 mm. For the bridge-based configuration, we assume a $2.5\ mm \times 6\ mm$ bridge interconnecting the dice. The framework is implemented in MATLAB.

6.2.2 Steady-State Thermal-PDN Co-Analysis Results

In this section, we analyze the thermal-PDN interactions of different configurations. Table 6.1 summarizes the specifications of the thermal simulations. Similar to the thermal evaluation cases in the previous section of this chapter, we assume that the system uses air cooled heat sinks and the case-to-ambient thermal conductance is 0.218 W/K. The secondary heat path is through the PCB. We use an effective heat transfer coefficient of 311 W/m^{2}K as the boundary condition at this interface. The ambient temperature is assumed to be 38°C.

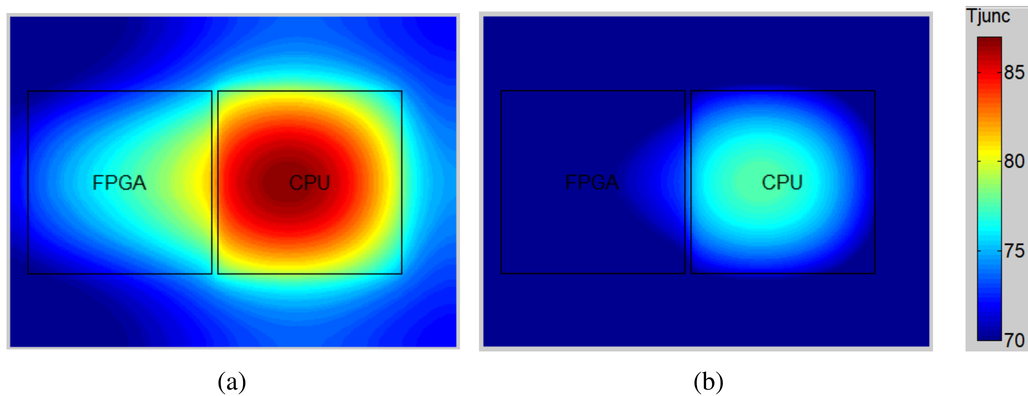


Figure 6.7: The temperature distribution for (a) standalone model, and (b) co-analysis model in multi-chip packages

Fig. 6.7 presents the temperature distribution from thermal-PDN co-analysis for a multi-die package for our two die system. Fig. 6.7(a) shows the thermal results from a standalone simulation assuming an ideal supply voltage. The maximum temperature of the CPU and the FPGA dice is 88°C and 81.3^{o}C, respectively. Likewise, Fig. 6.7(b) presents the temperature distribution accounting for all the interactions between the thermal and the PDN simulations. In this scenario, the maximum temperature of the CPU die and the FPGA die is 78.7^{o}C and 73.2^{o}C, respectively. Hence, we see that the standalone thermal simulation overestimates the maximum temperature by 11.3% and 10% for the CPU die and the FPGA die, respectively. Fig. 6.8 presents the results for a bridge-based configuration for our two die system. Since there is a silicon bridge interconnecting the dice on the package, there are two thermal coupling pathways from the 'hotter' die to the 'cooler' die

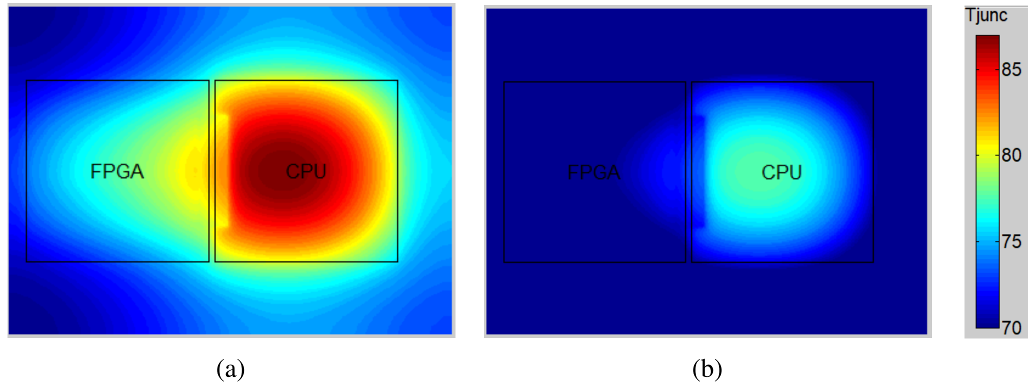


Figure 6.8: The temperature distribution for (a) standalone model, and (b) co-analysis model in bridge-based 2.5-D packages

(in this case, CPU die to FPGA die). However, since the heat sink is sitting atop the heat spreader, the primary thermal coupling path remains through the heat spreader. Hence, the temperature map is similar to the one observed for our multi-die package.

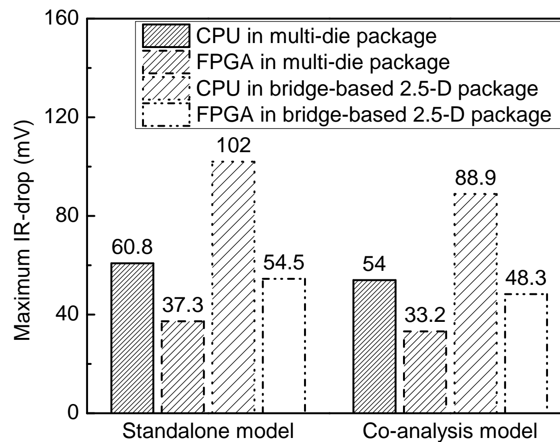


Figure 6.9: Steady-state IR-drop comparison for different configurations

Finally, in Fig. 6.9, we summarize the steady-state IR-drop results from these two configurations. The first half of the figure is the same as shown in Fig. 4.7 and is included for clarity. As stated previously, the leakage power is dependent on both the temperature of the die and the supply voltage. In each iteration of the analysis, we use a fitting function to determine the effective leakage power. We assume a worst case temperature as our initial condition (100°C). However, since the temperature is lower than the initial condition, the

estimated leakage power decreases. Likewise, our dynamic power estimation is based on a perfect supply voltage. When we incorporate the supply voltage fluctuations, the overall estimated power decreases. Moreover, the resistivity of the metal layers in the PDN is temperature dependent. Hence, in Fig. 6.9, we see that for both the multi-die package and the bridge-based package, there is a significant overestimation in the standalone model. For the multi-die package case, compared to the standalone modeling, both the CPU die and the FPGA die overestimate the maximum IR-drop by almost 11%. For the bridge-based case, the maximum IR-drop follows a similar trend where both dice overestimate the maximum IR-drop by approximately 12%. However, compared to the multi-die package configuration, the increase in IR-drop for the bridge-based configuration is 64% and 45% for the CPU die and the FPGA die, respectively. This increase in IR-drop is similar to what we observed in the standalone models.

6.2.3 Impact of Different Interaction Models and Number of Bridge-Chips

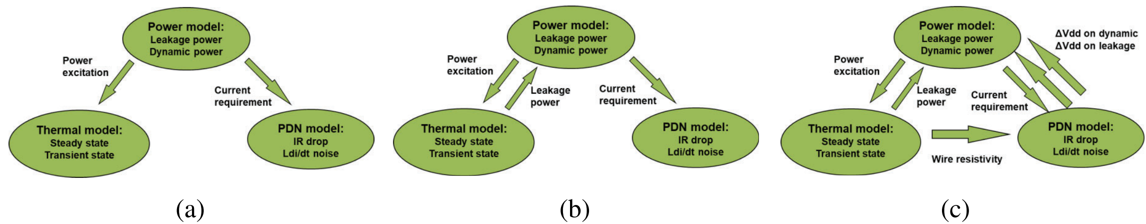


Figure 6.10: Different interaction models (a) standalone model, (b) thermal-leakage model, and (c) full model

In a thermal-PDN co-analysis environment, different interaction models contribute to the final self-consistent results. Yang et. al. [7] performed a comprehensive analysis on this for a 3-D integration technology with memory-on-CPU configuration. Fig. 6.10 shows three such cases. In the previous sections, we consider the standalone and full models. Fig. 6.10(a) and Fig. 6.10(c) show these interactions. Fig. 6.10(b) shows an intermediate model where the thermal impact on leakage power is considered. Table 6.2 summarizes the results for these three interaction models. We observe that the dominant contributor to the

Table 6.2: Comparison of different interaction models

Metric	Full model	Full model Vs. standalone	Thermal-leakage Vs. standalone	Full mode Vs. thermal-leakage
CPU temperature (C)	78.7	12%	10%	1.6%
FPGA temperature (C)	73.2	11%	10%	1%
CPU power (W): Dy-namic/Leakage	56.4/7.5	6%/98%	0%/52%	6%/31%
FPGA power (W): Dy-namic/Leakage	34.7/3.8	3%/135%	0%/99%	3%/18%

overestimation of results is thermal impact on leakage power. Between the full model and the thermal-leakage model, there is a 31% and 18% overestimation in leakage power for the CPU die and the FPGA die, respectively. Hence, with the increase in leakage power of a die, especially for circuits with HP models [69] will tend to overestimate the results more than their LP counterparts.

Table 6.3: Impact of bridge-chip splitting

	CPU max IR-drop (mV)		FPGA max IR-drop (mV)	
	Standalone	Co-analysis	Standalone	Co-analysis
Standalone	60.8	54 (11.2%)	37.3	33.2 (11%)
Single bridge-chip	102	88.9 (12.8%)	54.5	48.3 (11.4%)
Standalone	76.2	66.8 (12.4%)	44.1	39.1 (11.3%)

In Chapter 3, in order to reduce the IR-drop of a bridge-chip based configuration, we proposed to split a bridge-chip into multiple small bridge-chips with similar aggregate area. In this section, we use the thermal-PDN co-analysis framework to analyze such a case. Table 6.3 summarizes the results including five smaller bridge-chips instead of a single large bridge-chip. In the co-analysis model, the trends in PSN are similarly compared to the prior analysis. Across different number of bridge-chips, we observe a $\sim$12% overestimation in thermal and PSN results between the standalone model and the co-analysis model.

6.3 Transient-State Thermal-PDN Co-Analysis

In this section, we analyze the thermal-PDN framework for transient L$\frac{di}{dt}$ response.

6.3.1 Transient-state IR-drop Co-Analysis Modeling Framework

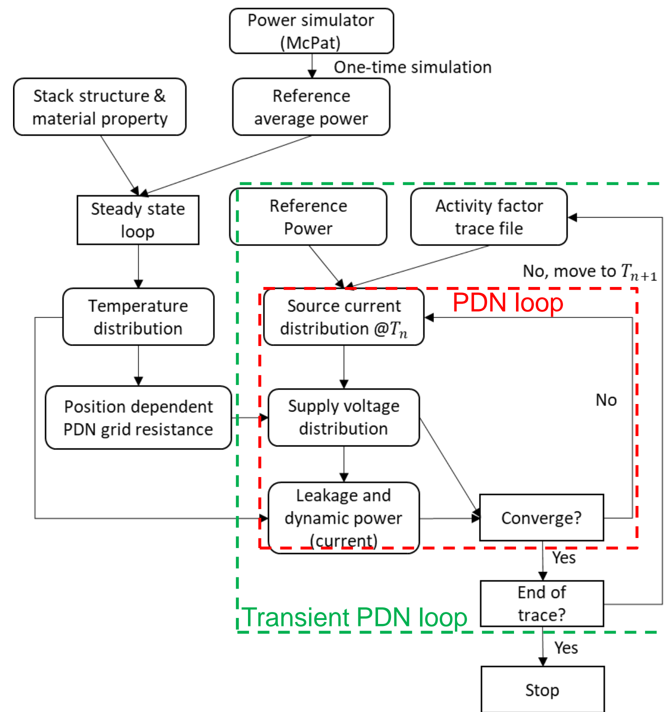


Figure 6.11: The flow chart for transient thermal-PDN co-analysis

Fig. 6.11 presents the analysis flow for transient simultaneous switching noise (SSN) analysis. Similar to the steady-state analysis, we start with one time McPAT simulation for reference power maps. We embed a switching activity with these power maps to calculate the average power maps for each die under consideration. For example, for step response based analysis, the average power is similar to the corresponding peak power of each die. However, for pulse based excitation, the average power is calculated based on the activity factor of the pulse and hence, the excitation is lower than the peak current excitation. The time-scale of transient thermal response is typically in the ms regime [90] whereas the time-scale for transient PDN is in nanoseconds. Since our PDN simulation time is ∼100 ns-200

ns, we assume that the thermal response of the system under consideration is invariant within this time-scale. We run one steady-state thermal framework with average power maps to get the temperature profile of each die. Beyond this step, the analysis flow has two explicit loops: one PDN loop for each time step and an external loop defined by the simulation time. After the completion of this self-consistent simulation loop, we achieve the final temperature and PDN analysis results.

6.3.2 Transient-State Co-Analysis Results

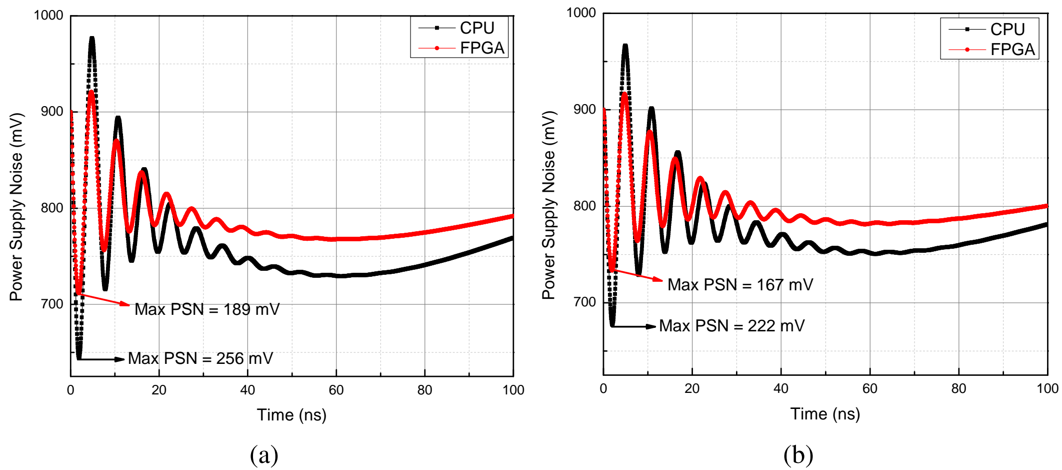


Figure 6.12: PDN step response results for (a) standalone model, and (b) co-analysis model in bridge-based 2.5-D packages

Fig. 6.12(a) and Fig. 6.12(b) provide the results for a bridge-chip based configuration with aforementioned CPU and FPGA dice. This analysis shows the SSN results for step response with 400 ps rise time. For the standalone analysis case, we observe 189 mV and 256 mV first droop noise for the FPGA die and the CPU die, respectively. However, from the co-analysis results, we observe that these values are over-estimated by 13.2% and 15.3%, respectively. These results are dependent on the specific power maps under consideration and the overlap region between the bridge-chip and the dice, as shown in Chapter 3. However, between a standalone model and a co-analysis model, we observe similar trends across different power maps and overlap regions.

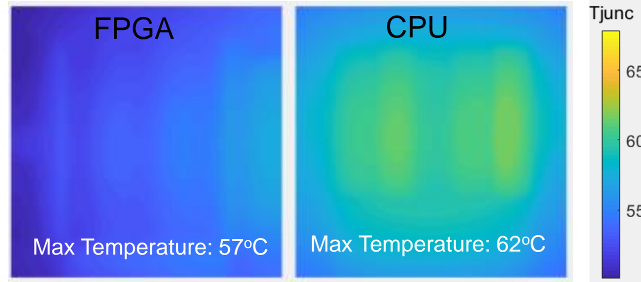


Figure 6.13: Average temperature profile for a 1 GHz on-die excitation

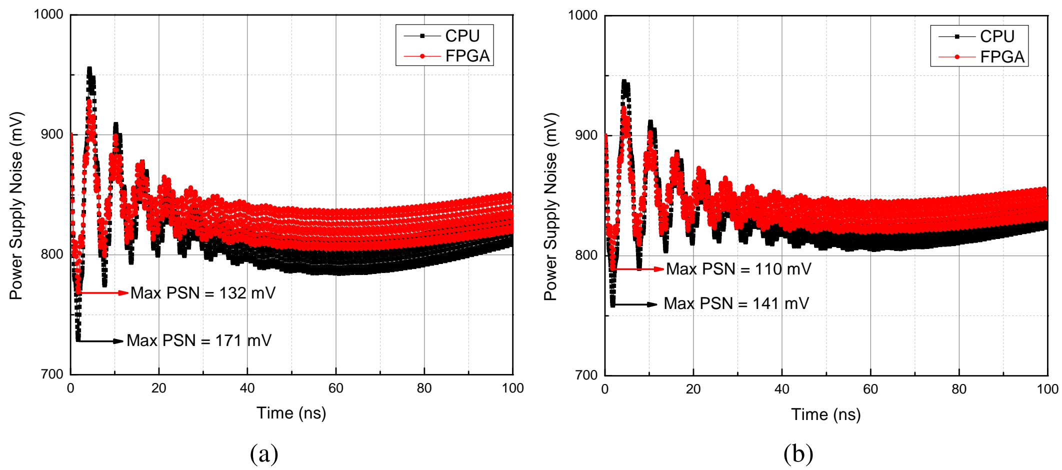


Figure 6.14: PDN results for a pulse excitation for (a) standalone model, and (b) co-analysis model in bridge-based 2.5-D packages

Step response analysis shows the worst-case current excitation scenario for an average power map analysis. However, a more realistic on-die excitation is a high frequency pulse. In Chapter 3, we analyzed the impact of different input pulse frequencies. Thermal-PDN co-analysis modeling vastly depends on the temperature gradient across a die. Since the average power map of a pulse excitation consumes less power compared to the worst-case step response based power maps, pulse response generates lesser peak temperature than an initial assumption of 100 ^oC. Fig. 6.13 presents the temperature profile for each die under the assumed current maps and input pulse. We use a 1 GHz input pulse with 1 ns period, 400 ps rise time, 400 ps fall time, respectively. Under these assumptions, CPU and FPGA temperatures are 62°C and 57°C (79°C and 71°C for step response), respectively. Attributed to the lower average power, the die temperature is 30% lower than the step

response scenario. Fig. 7.1(a) and Fig. 7.1(b) show the SSN results for a 1 GHz pulse excitation. For the standalone analysis case, we observe 132 mV and 171 mV first droop noise for the FPGA die and the CPU die, respectively. However, from the co-analysis results, we observe that these values are over-estimated by 20% and 21.2%, respectively. Since the power estimation is temperature dependent, the over estimation in the pulse SSN case is $\sim$21% compared to a 13-15% in the step response case.

6.4 Conclusions

In this chapter, we present a thermal-PDN co-analysis framework that incorporates impact of the thermal distribution of the dice on the supply voltage and vice versa. We incorporated a distributed package PDN model into our existing co-analysis framework to analyze different 2.5-D integration technologies. From steady-state co-analysis, we observe approximately 11% overestimation in the maximum temperature and 11-12% overestimation of the supply voltage for each die compared to the standalone models. We also perform thermal and transient PDN co-analysis. Our analysis shows that depending on the temperature gradient of a die, the standalone model can overestimate the thermal and PDN results by as much as $\sim$20%. While the standalone models can be adequate for pre-design exploration and mostly accurate for conventional packages, the co-analysis model provides added accuracy for 2.5-D/3-D architectures with increased power density and higher temperature gradients within and between dice. The worst-case pre-analysis results can be significantly different depending on the on-die stimulus we use. For example, a 1 GHz on-die stimulus results in a higher temperature gradient across a die under consideration. This leads to an overestimation of as much as 21% compared to a standalone PDN analysis. The leakage contribution to the total power is also an important factor since temperature gradient has the most significant impact on the leakage power of a die.

CHAPTER 7
SUMMARY AND FUTURE WORK

In this thesis, we evaluate thermal-mechanical and thermal-power delivery network (PDN) performance for emerging 2.5-D/3-D integration technologies.

7.1 Summary of the Presented Work

This thesis has five major contributions and are summarized below:

First, we perform a study that explores different means by which both interconnect reliability is improved and interposer warpage is decreased for an interposer-to-board integration platform using mechanically flexible interconnects (MFIs). Central to this exploration is the design and distribution/orientation of the MFIs on the interposer. Using Finite Element based tool ANSYS, different MFI distributions and configurations are investigated. A radially-oriented interconnect distribution in which the MFIs line up along the contours of thermal expansion/contraction is evaluated. Furthermore, a multi-objective genetic algorithm is employed to reduce max von-Mises stress in the MFIs and warpage in the interposer for each MFI distribution configuration. Impact of chip size and MFI pitch (from 400 μm to 1200 μm) on the mechanical integrity of the MFIs and interposer are also explored.

Second, a power delivery network (PDN) modeling framework for heterogeneous 2.5-D integration platforms is presented. The modeling framework, which includes both IR-drop and transient analyses, is first validated using IBM power grid benchmarks and the maximum relative errors are less than 7.3%. To evaluate both interposer and bridge-chip based 2.5-D integration platforms, we assume an FPGA-CPU 2.5-D integrated module in which the FPGA consumes 45 W and the CPU consumes 75 W. Modeling results show that an interposer with dense power/ground grids and microbumps can suppress power supply

noise (PSN) by a small margin with the requirement of high-density TSVs. For bridge-chip based 2.5-D integration, under the assumption that the active dice above the bridge-chips are not connected to package power/ground planes, some PDN challenges are highlighted and modeled. Using multiple bridge chips and smaller overlap areas between the bridge-chips and the active dice, the worst-case PSN in bridge-chip based 2.5-D integration is minimized. Next, we analyze the impact of including a PDN in the bridge-chip. We analyze a CPU-FPGA configurtion and a stacked-memory-FPGA configuration. For the latter configuration, although the base die suffers from high PDN noise for larger bridge-chip overlap area, the memory dice are invariant to this parameter. Moreover, a design space exploration of power delivery networks is performed for multi-chip 2.5-D and 3-D IC technologies. The focus of the study is the effective placement of the voltage regulator modules (VRMs) for power supply noise (PSN) suppression. Multiple on-package VRM configurations have been analyzed and compared. Additionally, 3D IC chip-on-VRM and backside-of-the-package VRM configurations are studied. From the PSN perspective, the 3D IC chip-on-VRM case suppresses the PSN the most even with high current density hotspots. The thesis also studies the impact of different parameters such as VRM-chip distance on the package, on-chip decoupling capacitor density, etc. on the PSN.

Third, we present a power delivery network (PDN) modeling framework for Fan-out Wafer Level Packaging (FOWLP) technologies with focus on multi-die heterogeneous integration. Results are compared to conventional multi-die packaging and 3D package-on-package technologies. Owing to the shorter interconnections enabled by thinner packages and elimination of large C4 bumps with copper pillars, the package contributes less parasitics to the PDN path. Hence, the IR-drop, transient droop, and impedance are reduced in the evaluated FOWLP technologies. The simultaneous switching noise in the evaluated multi-die FOWLP configuration is more than 20% lower than its flip-chip package counterpart. Likewise, similar improvement trends are seen for 3D stacked configurations. Specifically, if a double sided RDL is utilized in a 3D FOWLP, PSN can be reduced by

20% on average.

Fourth, a power delivery network (PDN) modeling framework for backside PDN configurations is presented. A backside PDN configuration contains dense micro-through silicon vias (μTSVs) and power/ground metal stack on the backside of the die. This approach separates the PDN from a conventional signaling network of the back-end-of-the-line (BEOL) and improves power integrity and core utilization. We benchmark this technology with conventional front-side BEOL PDN configurations. Owing to the lower resistivity compared to Cu metal lines for advanced technology nodes, we use Ruthenium (Ru) based buried power rail for PDN modeling. Our analysis shows that the steady-state IR-drop reduces by more than 4x in the backside PDN configuration, and a simultaneous switching noise analysis shows a significant reduction in transient droops. The framework results are validated with a place-and-route (P&R) based physical implementation flow. We quantify the area improvement in the actual flow and observe 25%-30% improvement in the backside PDN configuration. From PDN modeling framework, PDN results follow a trend similar to the ones obtained from block-level P&R of the given configurations. Moreover, we investigate the impact of package-to-die interconnect pitch, metal-insulator-metal cap density, and input pulse on PDN performance. Additionally, we perform thermal modeling to analyze thermal implications of a backside PDN configuration. From a thermal modeling perspective, there is negligible influence from dielectric bonding layer in a backside PDN configuration.

Fifth, we present a thermal-power delivery network (PDN) co-analysis framework to analyze various multi-die integration schemes. In the proposed approach, we capture the inter-dependencies between temperature distribution of the dice in a package and the supply voltage noise. We use standalone thermal and PDN analyses as references to compare our co-analysis results. Using a multi-die package and a bridge-based 2.5-D package case studies, our analysis shows a 10-12% overestimation in steady-state temperature and power supply noise. We also developed a framework to analyze the transient analysis of the

system. From this framework, we observe as much as $\sim$20% overestimation in the results compared to a standalone configuration. This is very much dependent on the activity factor of the on-die PDN. For a 1 GHz stimulus, owing to the lower average power consumption, we observe that the temperature can be $\sim$30% lower than the peak power case. This, consequently, impacts how the standalone analysis overestimates the system performance.

7.2 Future Research Extensions

For the five different tasks performed in this thesis, each part can be extended to better serve the scientific community.

7.2.1 Thermo-Mechanical Analysis for Emerging Technologies

The thermo-mechanical analysis can be extended to analyze a number of other emerging technologies. One key technology is compressible micro-interconnect based heterogeneous interconnect stitching technology [60]. Moreover, since this work is structural optimization of interconnects, it can be applied to non-flexible interconnect optimization scenarios as well. Fan-out wafer level packaging, 3-D stacking, etc. are a few candidates. From the algorithm perspective, several other optimization algorithms can be explored to reduce the run-time complexity.

7.2.2 Power Delivery Network and Thermal-PDN Co-Analysis

The PDN framework that we presented in this thesis can be extended to analyze several other configurations. We have analyzed fan-out based packages, backside PDN configurations, and bridge-chip based 2.5-D configurations. These technologies are mutually exclusive and hence, several combination of these different technologies can be analyzed to leverage the best performance of each technology. Poppod et. al. [6] shows a combination of bridging technology and fan-out packaging technology. This can be analyzed from power delivery perspective. Moreover, 3-D stacking of dice with backside PDN configura-

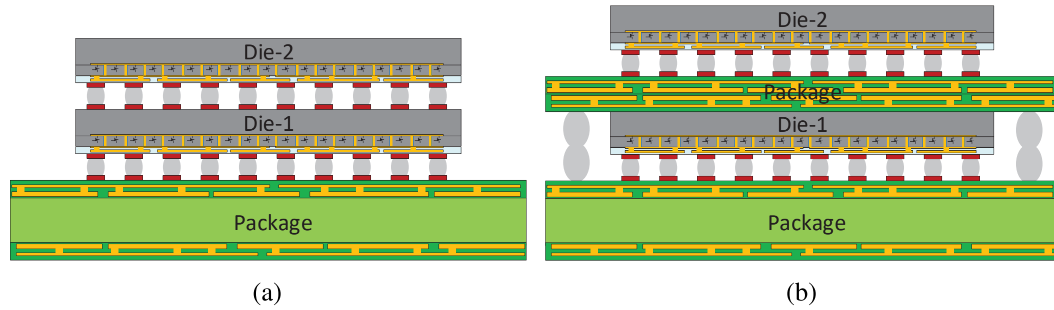


Figure 7.1: 3-D stacking with backside PDN for (a) face-to-face bonding, and (b) fan-out wafer level packaging based package-on-package

tion can be analyzed. Specifically, there are two different directions we can follow for this configuration. First, we can deliver power to the top die through the signaling network of the bottom die. Second, we can use a 3-D FOWLP configuration to deliver power using TMVs. Fig. 7.1 presents these two cases. Thermal-PDN co-analysis can be performed for this configuration as well.

7.2.3 PDN-Signaling Co-Analysis with Backside PDN

The backside PDN configurations analyzed in this thesis require some analysis from a signaling perspective. This configuration partially separates PDN from the signaling network. However, signaling network is farther from the I/Os. Full-wave analysis can be performed in Ansys electromagnetic suite to characterize the behavior of such channels. Moreover, power supply induced characteristics of such a signaling network can be characterized using HSPICE and the PDN modeling framework discussed in this thesis.

7.2.4 Impact of Emerging Heterogeneous Integration Technologies on Network-on-Chip for Applied Machine Learning Algorithms

Emerging packaging technologies can be analyzed from an architectural perspective. The ever-growing demand for large-scale data analytics rejuvenated the idea of near-memory computing. There are two fundamental memory bottlenecks: limited off-chip bandwidth and long access latency. Moreover, processors alone cannot meet the demand of these

power hungry computations. Alternative hardwares such as Graphics Processing Unit (GPUs) and Field Programmable Gate Arrays (FPGAs) are getting increasingly popular as accelerator fabrics. For a scale-out architecture for deep neural network training/inference, different packaging technologies can be benchmarked.

REFERENCES

[1] Seagate, *The Digitization of the World From Edge to Core*, https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf.

[2] L. B. N. L. US Department of Energy, *United States Data Center Energy Usage Report*, https://datacenters.lbl.gov/sites/default/files/EnergyUsageWebinar12062016.pdf.

[3] S. Pelley, D. Meisner, T. F. Wenisch, and J. W. VanGilder, "Understanding and abstracting total data center power," *Proc. Workshop Energy-Efficient Des.*, pp. 1–6, 2009.

[4] Microsoft, *PROJECT NATICK: 50% OF US LIVE NEAR THE COAST. WHY DOESN'T OUR DATA?* https://natick.research.microsoft.com/.

[5] Y. Zhang, M. O. Hossen, and M. S. Bakir, "Power delivery network benchmarking for interposer and bridge-chip-based 2.5-d integration," *IEEE Electron Device Letters*, vol. 39, no. 1, pp. 99–102, 2018.

[6] A. Podpod, J. Slabbekoorn, A. Phommahaxay, F. Duval, A. Salahouedlhadj, M. Gonzalez, K. Rebibis, R. Miller, G. Beyer, and E. Beyne, "A novel fan-out concept for ultra-high chip-to-chip interconnect density with 20-μm pitch," in *2018 IEEE 68th Electronic Components and Technology Conference*, 2018.

[7] Y. Zhang and M. S. Bakir, "Integrated thermal and power delivery network co-simulation framework for single-die and multi-die assemblies," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 7, no. 3, pp. 434–443, 2017.

[8] C. F. Tseng, C. S. Liu, C. H. Wu, and D. Yu, "Info (wafer level integrated fan-out) technology," in *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, 2016, pp. 1–6.

[9] D. Greenhill, R. Ho, D. Lewis, H. Schmit, K. H. Chan, A. Tong, S. Atsatt, D. How, P. McElheny, K. Duwel, J. Schulz, D. Faulkner, G. Iyer, G. Chen, H. K. Phoon, H. W. Lim, W. Koay, and T. Garibay, "3.3 a 14nm 1ghz fpga with 2.5d transceiver integration," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 54–55.

[10] Microsoft, *Project Catapult*, https://www.microsoft.com/en-us/research/project/project-catapult/.

[11] Xilinx, *Versal: The First Adaptive Compute Acceleration Platform (ACAP)*, https://www.xilinx.com/support/documentation/white_papers/wp505-versal-acap.pdf.

[12] Cerebras, *The WSE is the largest chip ever built*, https://www.cerebras.net/.

[13] Intel, *Fact Sheet: New Intel Architectures and Technologies Target Expanded Market Opportunities*, https://newsroom.intel.com/articles/new-intel-architectures-technologies-target-expanded-market-opportunities/#gs.101ekvr.

[14] ITRS, *International Technology Roadmap for Secmiconductors, 2013*, http://www.itrs.net/.

[15] T. E. Sarvey, Y. Zhang, L. Zheng, P. Thadesar, R. Gutala, C. Cheung, A. Rahman, and M. S. Bakir, "Embedded cooling technologies for densely integrated electronic systems," in *Custom Integrated Circuits Conference (CICC), 2015 IEEE*, 2015, pp. 1–8.

[16] H. Su, F. Liu, A. Devgan, E. Acar, and S. Nassif, "Full chip leakage-estimation considering power supply and temperature variations," in *Proc. Int. Symp. on Low Power Electronics and Design*, 2003, pp. 78–83.

[17] H. Oprins and E. Beyne, "Generic thermal modeling study of the impact of 3d -interposer material and thickness options on the thermal performance and die-to-die thermal coupling," in *Proc. IEEE Intersociety Conf on Thermal and Thermomechanical Phenomena in Electronic Systems.*, 2014, pp. 72–78.

[18] Intel®, *Embedded Multi-die Interconnect Bridge, A breakthrough in advanced packaging technology*, http://www.intel.com/content/www/us/en/foundry/emib.html.

[19] NVIDIA, *NVIDIA Tesla P100*, https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf.

[20] B. Banijamali, S. Ramalingam, K. Nagarajan, and R. Chaware, "Advanced reliability study of tsv interposers and interconnects for the 28nm technology fpga," in *2011 IEEE 61st Electronic Components and Technology Conference (ECTC)*, 2011, pp. 285–290.

[21] R. Chaware, K. Nagarajan, and S. Ramalingam, "Assembly and reliability challenges in 3d integration of 28nm fpga die on a large high density 65nm passive interposer," in *2012 IEEE 62nd Electronic Components and Technology Conference*, 2012, pp. 279–283.

[22] C. Lee, C. Hung, C. Cheung, P. Yang, C. Kao, D. Chen, M. Shih, C. C. Chien, Y. Hsiao, L. Chen, M. Su, M. Alfano, J. Siegel, J. Din, and B. Black, "An overview of the development of a gpu with integrated hbm on silicon interposer," in *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, 2016, pp. 1439–1444.

[23] P. T. Huang, L. C. Chou, T. C. Huang, S. L. Wu, T. S. Wang, Y. R. Lin, C. A. Cheng, W. W. Shen, K. N. Chen, J. C. Chiou, C. T. Chuang, W. Hwang, K. H. Chen, C. T. Chiu, M. H. Cheng, Y. L. Lin, and H. M. Tong, "18.6 2.5d heterogeneously integrated bio-sensing microsystem for multi-channel neural-sensing applications," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 320–321.

[24] U. Kang, H.-J. Chung, S. Heo, S.-H. Ahn, H. Lee, S.-H. Cha, J. Ahn, D. Kwon, J. H. Kim, J.-W. Lee, H.-S. Joo, W.-S. Kim, H.-K. Kim, E.-M. Lee, S.-R. Kim, K.-H. Ma, D.-H. Jang, N.-S. Kim, M.-S. Choi, S.-J. Oh, J.-B. Lee, T.-K. Jung, J.-H. Yoo, and C. Kim, "8gb 3d ddr3 dram using through-silicon-via technology," in *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, 2009, 130–131,131a.

[25] D. Foley and J. Danskin, "Ultra-performance pascal gpu and nvlink interconnect," *IEEE Micro*, vol. 37, no. 2, pp. 7–17, 2017.

[26] R. Mahajan, R. Sankman, N. Patel, D. W. Kim, K. Aygun, Z. Qian, Y. Mekonnen, I. Salama, S. Sharan, D. Iyengar, and D. Mallik, "Embedded multi-die interconnect bridge (emib) – a high density, high bandwidth packaging interconnect," in *IEEE Electronic Components and Technology Conf.*, 2016, pp. 557–565.

[27] X. Zhang, P. K. Jo, M. Zia, G. S. May, and M. S. Bakir, "Heterogeneous interconnect stitching technology with compressible microinterconnects for dense multi-die integration," *IEEE Electron Device Letters*, vol. 38, no. 2, pp. 255–257, 2017.

[28] C. C. Liu, S. M. Chen, F. W. Kuo, H. N. Chen, E. H. Yeh, C. C. Hsieh, L. H. Huang, M. Y. Chiu, J. Yeh, T. S. Lin, T. J. Yeh, S. Y. Hou, J. P. Hung, J. C. Lin, C. P. Jou, C. T. Wang, S. P. Jeng, and D. C. H. Yu, "High-performance integrated fan-out wafer level packaging (info-wlp): Technology and system integration," in *2012 International Electron Devices Meeting*, 2012, pp. 14.1.1–14.1.4.

[29] C. T. Wang and D. Yu, "Signal and power integrity analysis on integrated fan-out pop (info_pop) technology for next generation mobile applications," in *2016 IEEE*

66th Electronic Components and Technology Conference (ECTC), 2016, pp. 380–385.

[30] Y. H. Chou, P. C. Pan, C. Y. Huang, M. F. Jhong, and C. C. Wang, "The comparison of package design and electrical analysis in mobile application," in *2017 IEEE 67th Electronic Components and Technology Conference (ECTC)*, 2017, pp. 1855–1860.

[31] T. Meyer, G. Ofner, S. Bradl, M. Brunnbauer, and R. Hagen, "Embedded wafer level ball grid array (ewlb)," in *2008 10th Electronics Packaging Technology Conference*, 2008, pp. 994–998.

[32] B. Keser, C. Amrine, T. Duong, O. Fay, S. Hayes, G. Leal, W. Lytle, D. Mitchell, and R. Wenzel, "The redistributed chip package: A breakthrough for advanced packaging," in *2007 Proceedings 57th Electronic Components and Technology Conference*, 2007, pp. 286–291.

[33] G. Dickerson, *NAVIGATING THE PERFECT STORM ENABLING THE A.I. ERA*, `http://www.semiconwest.org/programs-catalog/`.

[34] E. A. Burton, G. Schrom, F. Paillet, J. Douglas, W. J. Lambert, K. Radhakrishnan, and M. J. Hill, "Fivr - fully integrated voltage regulators on 4th generation intel®core™socs," in *Applied Power Electronics Conference and Exposition, 2014. APEC 2014. IEEE International*, 2014, pp. 432–439.

[35] N. Sturcken, E. O'Sullivan, N. Wang, P. Herget, B. Webb, L. Romankiw, M. Petracca1, R. Davies, R. Fontana, G. Decad, I. Kymissis, A. Peterchev, L. Carloni, W. Gallagher, and K. Shepard, "A 2.5d integrated voltage regulator using coupled-magnetic-core inductors on silicon interposer delivering 10.8a/mm^2," in *Solid-State Circuits Conference, 2012. ISSCC 2012. IEEE International*, 2012, pp. 400–402.

[36] E. A. Burton, G. Schrom, F. Paillet, J. Douglas, W. J. Lambert, K. Radhakrishnan, and M. J. Hill, "Fivr - fully integrated voltage regulators on 4th generation intel®core™socs," in *2014 IEEE Applied Power Electronics Conference and Exposition - APEC 2014*, 2014, pp. 432–439.

[37] R. Devaney, *Failure Analysis of Solder Joints and Circuit Boards*, `https://www.smta.org/chapters/files/Oregon_Failure_Analysis_of_Solder_Joints__PCAs_October_2012.ppt`.

[38] X. Liu, Q. Chen, V. Sundaram, R. R. Tummala, and S. K. Sitaraman, "Failure analysis of through-silicon vias in free-standing wafer under thermal-shock test," *Microelectronics Reliability*, vol. 53, no. 1, pp. 70 –78, 2013, Reliability of Micro-Interconnects in 3D IC Packages.

[39] C. Wan, T. K. Gaylord, and M. S. Bakir, “Rigorous coupled-wave analysis equivalent-index-slab method for analyzing 3d angular misalignment in interlayer grating couplers,” *Appl. Opt.*, vol. 55, no. 35, pp. 10 006–10 015, 2016.

[40] N. Kim, D. Wu, D. Kim, A. Rahman, and P. Wu, “Interposer design optimization for high frequency signal transmission in passive and active interposer using through silicon via (tsv),” in *2011 IEEE 61st Electronic Components and Technology Conference (ECTC)*, 2011, pp. 1160–1167.

[41] J. U. Knickerbocker, C. S. Patel, P. S. Andry, C. K. Tsang, L. P. Buchwalter, E. J. Sprogis, H. Gan, R. R. Horton, R. J. Polastre, S. L. Wright, and J. M. Cotte, “3-d silicon integration and silicon packaging technology using silicon through-vias,” *IEEE Journal of Solid-State Circuits*, vol. 41, no. 8, pp. 1718–1725, 2006.

[42] M. Amagai, “Mechanical reliability in electronic packaging,” *Microelectronics Reliability*, vol. 42, no. 4, pp. 607 –627, 2002.

[43] P Hassell, “Advanced warpage characterization: Location and type of displacement can be equally as important as magnitude,” in *Proc of Pan Pacific Microelectronics Symposium Conference*, 2001.

[44] K. Y. Au, S. L. Kriangsak, X. R. Zhang, W. H. Zhu, and C. H. Toh, “3d chip stacking amp; reliability using tsv-micro c4 solder interconnection,” in *2010 Proceedings 60th Electronic Components and Technology Conference (ECTC)*, 2010, pp. 1376–1384.

[45] V. H. O. S. H. L. Susant K. Patra Jian Ma, “Alignment issues in packaging for free-space optical interconnects,” *Optical Engineering*, vol. 33, pp. 33 –33 –10, 1994.

[46] M. Cabezón, I. Garcés, A. Villafranca, J. Pozo, P. Kumar, and A. Kaźmierczak, “Silicon-on-insulator chip-to-chip coupling via out-of-plane or vertical grating couplers,” *Appl. Opt.*, vol. 51, no. 34, pp. 8090–8094, 2012.

[47] S. Bernabé, C. Kopp, M. Volpert, J. Harduin, J.-M. Fédéli, and H. Ribot, “Chip-to-chip optical interconnections between stacked self-aligned soi photonic chips,” *Opt. Express*, vol. 20, no. 7, pp. 7886–7894, 2012.

[48] J. Yao, X. Zheng, G. Li, I. Shubin, H. Thacker, Y. Luo, K. Raj, J. E. Cunningham, and A. V. Krishnamoorthy, “Grating-coupler based low-loss optical interlayer coupling,” in *8th IEEE International Conference on Group IV Photonics*, 2011, pp. 383–385.

[49] S. Y. Hou, W. C. Chen, C. Hu, C. Chiu, K. C. Ting, T. S. Lin, W. H. Wei, W. C. Chiou, V. J. C. Lin, V. C. Y. Chang, C. T. Wang, C. H. Wu, and D. Yu, “Wafer-level integration of an advanced logic-memory system through the second-generation

cowos technology," *IEEE Transactions on Electron Devices*, vol. 64, no. 10, pp. 4071–4077, 2017.

[50] D. C. H. Yu, "Advanced packaging with greater simplicity," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 3.6.1–3.6.4.

[51] S. Raghavan, K. Klein, S. Yoon, J. Kim, K. Moon, C. P. Wong, and S. K. Sitaraman, "Methodology to predict substrate warpage and different techniques to achieve substrate warpage targets," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, no. 7, pp. 1064–1074, 2011.

[52] M. Detalle, B. Vandevelde, P. Nolmans, J. D. Messemaeker, M. Gonzalez, A. Miller, A. L. Manna, G. Beyer, and E. Beyne, "Minimizing interposer warpage by process control and design optimization," in *2014 IEEE 64th Electronic Components and Technology Conference (ECTC)*, 2014, pp. 33–40.

[53] K. Murayama, M. Aizawa, K. Hara, M. Sunohara, K. Miyairi, K. Mori, J. Charbonnier, M. Assous, J. Bally, G. Simon, and M. Higashi, "Warpage control of silicon interposer for 2.5d package application," in *2013 IEEE 63rd Electronic Components and Technology Conference*, 2013, pp. 879–884.

[54] M. S. Bakir, H. A. Reed, H. D. Thacker, C. S. Patel, P. A. Kohl, K. P. Martin, and J. D. Meindl, "Sea of leads (sol) ultrahigh density wafer-level chip input/output interconnections for gigascale integration (gsi)," *IEEE Transactions on Electron Devices*, vol. 50, no. 10, pp. 2039–2048, 2003.

[55] L. Ma, Q. Zhu, T. Hantschel, D. K. Fork, and S. K. Sitaraman, "J-springs - innovative compliant interconnects for next-generation packaging," in *52nd Electronic Components and Technology Conference 2002. (Cat. No.02CH37345)*, 2002, pp. 1359–1365.

[56] S. Muthukumar, C. D. Hill, S. Ford, W. Worwag, T. Dambrauskas, P. C. Challela, T. S. Dory, N. M. Patel, E. L. Ramsay, and D. S. Chau, "High-density compliant die-package interconnects," in *56th Electronic Components and Technology Conference 2006*, 2006, 6 pp.–.

[57] R. Okereke and S. K. Sitaraman, "Three-path electroplated copper compliant interconnects — fabrication and modeling studies," in *2013 IEEE 63rd Electronic Components and Technology Conference*, 2013, pp. 129–135.

[58] P. Jo, M. O. Hossen, X. Zhang, Y. Zhang, and M. Bakir, "Heterogeneous multi-die stitching: Technology demonstration and design considerations," in *2018 IEEE 68th Electronic Components and Technology Conference (ECTC)*, doi:10.1109/ECTC.2018.00230, 2018, pp. 1512–1518.

[59] C. Zhang, H. S. Yang, and M. S. Bakir, "Highly elastic gold passivated mechanically flexible interconnects," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 3, no. 10, pp. 1632–1639, 2013.

[60] P. K. Jo, X. Zhang, J. L. Gonzalez, G. S. May, and M. S. Bakir, "Heterogeneous multi-die stitching enabled by fine-pitch and multi-height compressible microinterconnects (cmis)," *IEEE Transactions on Electron Devices*, vol. 65, no. 7, pp. 2957–2963, 2018.

[61] J. L. Gonzalez, P. K. Jo, R. Abbaspour, and M. S. Bakir, "Flexible interconnect design using a mechanically-focused, multi-objective genetic algorithm," *Journal of Microelectromechanical Systems*, vol. 27, no. 4, pp. 677–685, 2018.

[62] W. Chen and S. K. Sitaraman, "Response surface and multiobjective optimization methodology for the design of compliant interconnects," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 4, no. 11, pp. 1769–1777, 2014.

[63] Q. Zhu, L. Ma, and S. K. Sitaraman, "Design optimization of one-turn helix - a novel compliant off-chip interconnect," in *ITherm 2002. Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (Cat. No.02CH37258)*, 2002, pp. 833–839.

[64] N. Kurd, M. Chowdhury, E. Burton, T. P. Thomas, C. Mozak, B. Boswell, M. Lal, A. Deval, J. Douglas, M. Elassal, A. Nalamalpu, T. M. Wilson, M. Merten, S. Chennupaty, W. Gomes, and R. Kumar, "5.9 haswell: A family of ia 22nm processors," in *Solid-State Circuits Conference, 2014. ISSCC 2014. IEEE International*, 2014, pp. 112–113.

[65] K. Gillespie, H. R. F. III, C. Henrion, R. Jotwani, S. Kosonocky, R. S. Orefice, D. A. Priore1, J. White, and K. Wilcox, "5.5 steamroller: An x86-64 core implemented in 28nm bulk cmos," in *Solid-State Circuits Conference, 2014. ISSCC 2014. IEEE International*, 2014, pp. 104–105.

[66] A. Iyer and D. Marculescu, "Power efficiency of voltage scaling in multiple clock multiple voltage cores," in *Conference on Computer Aided Design, 2002. ICCAD 2002. IEEE/ACM International*, 2002, pp. 379–386.

[67] A. Putnam, A. Caulfield, E. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J.-Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Xiao, and D. Burger, "A reconfigurable fabric for accelerating large-scale datacenter services," in *Proc. IEEE Int. Symp. on Computer Architecture*, 2014, pp. 13–24.

[68] J. Power, A. Basu, J. Gu, S. Puthoor, B. M. Beckmann, M. D. Hill, S. K. Reinhardt, and D. A. Wood, "Heterogeneous system coherence for integrated cpu-gpu systems," in *Proc. Annual Int. Symp. Microarchitecture*, Davis, California, 2013, pp. 457–467.

[69] ASU, *Predictive Technology Model*, http://ptm.asu.edu/.

[70] M. S. Gupta, J. L. Oatley, R. Joseph, G.-Y. Wei, and D. M. Brooks, "Understanding voltage variations in chip multiprocessors using a distributed power-delivery network," in *Proc. Design, Automation and Test in Europe*, Nice, France, 2007, pp. 624–629.

[71] W. Kim, M. S. Gupta, G. Y. Wei, and D. Brooks, "System level analysis of fast, per-core dvfs using on-chip switching regulators," in *14th IEEE International Symposium on High Performance Computer Architecture, Salt Lake City, UT, 2008*, 2008, pp. 123–134.

[72] H. K. Krishnamurthy, V. Vaidya, S. Weng, K. Ravichandran, P. Kumar, S. Kim, R. Jain, G. Matthew, J. Tschanz, and V. De, "20.1 a digitally controlled fully integrated voltage regulator with on-die solenoid inductor with planar magnetic core in 14nm tri-gate cmos," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 336–337.

[73] M. Mayberry, "What lies ahead for interconnects and devices," in *2012 IEEE International Interconnect Technology Conference*, doi:10.1109/IITC.2012.6251563, 2012, pp. 1–3.

[74] K. Croes, C. Adelmann, C. J. Wilson, H. Zahedmanesh, O. V. Pedreira, C. Wu, A. Leśniewska, H. Oprins, S. Beyne, I. Ciofi, D. Kocaay, M. Stucchi, and Z. Tőkei, "Interconnect metals beyond copper: Reliability challenges and opportunities," in *2018 IEEE International Electron Devices Meeting (IEDM)*, doi:10.1109/IEDM.2018.8614695, 2018, pp. 531–534.

[75] Altera®, *Power Delivery Network (PDN) Tool User Guide*, https://www.altera.com/en_US/pdfs/literature/ug/ug_pdn.pdf.

[76] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Power challenges may end the multicore era," *Commun. ACM*, vol. 56, no. 2, pp. 93–102, Feb. 2013.

[77] D. Oh, "System level jitter characterization of high speed i/o systems," in *2012 IEEE International Symposium on Electromagnetic Compatibility*, doi:10.1109/ISEMC.2012.6351789, 2012, pp. 173–178.

[78] C. Erdmann, D. Lowney, A. Lynam, A. Keady, J. McGrath, E. Cullen, D. Breathnach, D. Keane, P. Lynch, M. D. L. Torre, R. D. L. Torre, P. Lim, A. Collins, B. Farley, and L. Madden, "A heterogeneous 3d-ic consisting of two 28 nm fpga die and 32 reconfigurable high-performance data converters," *Solid-State Circuits, IEEE Journal of*, vol. 50, no. 1, pp. 258–269, 2015.

[79] N. H. K. S. M. Alam and S. Hassoun, "System-level comparison of power delivery design for 2d and 3d ics," in *2009 IEEE International Conference on 3D System Integration*, 2009, pp. 1–7.

[80] J. Xie and M. Swaminathan, "Electrical-thermal co-simulation of 3d integrated systems with micro-fluidic cooling and joule heating effects," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, no. 2, pp. 234–246, 2011.

[81] R. Zhang, K. Wang, B. Meyer, M. Stan, and K. Skadron, "Architecture implications of pads as a scarce resource," in *Proc. IEEE Int. Symp. on Computer Architecture*, 2014, pp. 373–384.

[82] S. J. Park and M. Swaminathan, "Temperature-aware power distribution network designs for 3d ics and systems," in *IEEE Electronic Components and Technology Conf.*, 2015, pp. 732–737.

[83] X. Zhang, T. Tong, S. Kanev, S. K. Lee, G.-Y. Wei, and D. Brooks, "Characterizing and evaluating voltage noise in multi-core near-threshold processors," in *Proc. Int. Symp. on Low Power Electronics and Design*, 2013, pp. 82–87.

[84] H. He and J.-Q. Lu, "Modeling and analysis of pdn impedance and switching noise in tsv-based 3-d integration," *Electron Devices, IEEE Transactions on*, vol. 62, no. 4, pp. 1241–1247, 2015.

[85] Y. Shao, Z. Peng, and J.-F. Lee, "Thermal-aware dc ir-drop co-analysis using non-conformal domain decomposition methods," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 468, no. 2142, pp. 1652–1675, 2012.

[86] C. Pan, S. Mukhopadhyay, and A. Naeemi, "System-level chip/package co-design for multi-core processors implemented with power-gating technique," in *epeps*, 2014, pp. 11–14.

[87] J. Xie and M. Swaminathan, "Electrical and thermal cosimulation with nonconformal domain decomposition method for multiscale 3-d integrated systems," *Components, Packaging and Manufacturing Technology, IEEE Transactions on*, vol. 4, no. 4, pp. 588–601, 2014.

[88] Y. Zhang, Y. Zhang, and M. S. Bakir, “Thermal design and constraints for heterogeneous integrated chip stacks and isolation technology using air gap and thermal bridge,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 4, no. 12, pp. 1914–1924, 2014.

[89] Y. Liu, R. Dick, L. Shang, and H. Yang, “Accurate temperature-dependent integrated circuit leakage power estimation is easy,” in *Proc. Design, Automation and Test in Europe*, 2007, pp. 1–6.

[90] Y. Zhang, T. E. Sarvey, and M. S. Bakir, “Thermal evaluation of 2.5-d integration using bridge-chip technology: Challenges and opportunities,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 7, no. 7, pp. 1101–1110, 2017.

[91] E. Slavcheva, W. Mokwa, and U. Schnakenberg, “Electrodeposition and properties of niw films for mems application,” *Electrochimica Acta*, vol. 50, no. 28, pp. 5573 –5580, 2005.

[92] H. S. Yang and M. S. Bakir, “Design, fabrication, and characterization of freestanding mechanically flexible interconnects using curved sacrificial layer,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 2, no. 4, pp. 561–568, 2012.

[93] J. Wu, P. Zheng, C. Lee, S. Hung, and J. Lee, “A study in flip-chip ubm/bump reliability with effects of snpb solder composition,” *Microelectronics Reliability*, vol. 46, no. 1, pp. 41 –52, 2006.

[94] M.-Y. Tsai, C. H. J. Hsu, and C. T. O. Wang, “Investigation of thermomechanical behaviors of flip chip bga packages during manufacturing process and thermal cycling,” *IEEE Transactions on Components and Packaging Technologies*, vol. 27, no. 3, pp. 568–576, 2004.

[95] M. O. Hossen, J. L. Gonzalez, and M. S. Bakir, “Thermomechanical analysis and package-level optimization of mechanically flexible interconnects for interposer-on- motherboard assembly,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 8, no. 12, pp. 2081–2089, 2018.

[96] Y. Zhang, M. O. Hossen, and M. S. Bakir, “Power delivery network modeling and benchmarking for emerging heterogeneous integration technologies,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, pp. 1–1, 2019.

[97] Y. Peng, T. Song, D. Petranovic, and S. K. Lim, “Silicon effect-aware full-chip extraction and mitigation of tsv-to-tsv coupling,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 12, pp. 1900–1913, 2014.

[98] S. R. Nassif, "Power grid analysis benchmarks," in *Proc. IEEE Asia and South Pacific Design Automation Conf.*, Seoul, Korea, 2008, pp. 376–381.

[99] K. Cho, Y. Kim, S. Kim, H. Lee, S. Choi, H. Kim, and J. Kim, "Power distribution network (pdn) design and analysis of a single and double-sided high bandwidth memory (hbm) interposer for 2.5d terabtye/s bandwidth system," in *2016 IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 2016, pp. 96–99.

[100] Altera®, *Enabling Next-Generation Platforms Using Altera's 3D System-in-Package Technology*, https://www.altera.com/content/dam/.

[101] Altera, *PowerPlay Early Power Estimators (EPE) and Power Analyzer (Stratix IV and Stratix V)*, https://www.altera.com/support/support-resources/operation-and-testing/power/pow-powerplay.html.

[102] Altera®, *Leveraging HyperFlex Architecture in Stratix 10 Devices to Achieve Maximum Power Reduction*, https://www.altera.com/products/fpga/stratix-series/stratix-10/overview.html.

[103] C. Yoon, G. Chen, D. Greenhill, H. Kang, A. Hashemi, and W. Beyene, "Analysis of noise coupling and timing error in silicon bridge application," in *2018 IEEE 27th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, 2018, pp. 279–281.

[104] H. Jun, J. Cho, K. Lee, H. Son, K. Kim, H. Jin, and K. Kim, "Hbm (high bandwidth memory) dram technology and architecture," in *2017 IEEE International Memory Workshop (IMW)*, 2017, pp. 1–4.

[105] M. O. Hossen, Y. Zhang, and M. S. Bakir, "Thermal-power delivery network co-analysis for multi-die integration," in *2018 IEEE 27th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, 2018, pp. 155–157.

[106] M.-J. Wang, C.-Y. H. C.-L. Kao, P.-N. Lee, C.-H. Chen, C.-P. Hung, and H.-M. Tong, "Tsv technology for 2.5d ic solution," in *2012 IEEE 62nd Electronic Components and Technology Conference, San Diego, CA, 2012*, 2012, pp. 284–288.

[107] P. Enquist, G. Fountain, C. Petteway, A. Hollingsworth, and H. Grady, "Low cost of ownership scalable copper direct bond interconnect 3d ic technology for three dimensional integrated circuit application," in *2009 IEEE International Conference on 3D System Integration, San Francisco, CA, 2009*, 2009, pp. 1–6.

[108] Altera, *Device-Specific Power Delivery Network (PDN) Tool 2.0 User Guide*, https://www.altera.com/content/dam/altera-www/global/.

[109] L. Zheng, Y. Zhang, and M. S. Bakir, “Full-chip power supply noise time-domain numerical modeling and analysis for single and stacked ics,” *IEEE Transactions on Electron Devices*, vol. 63, no. 3, pp. 1225–1231, 2016.

[110] O. Semiconductor, *Methods to Characterize Parasitic Inductance and Resistance of Modern VRM*, `http://www.onsemi.com/pub/Collateral/`.

[111] R. Kumar, K. I. Farkas, N. P. Jouppi, P. Ranganathan, and D. M. Tullsen, “Single-isa heterogeneous multi-core architectures: The potential for processor power reduction,” in *Proceedings. 36th Annual IEEE/ACM International Symposium on Microarchitecture, 2003. MICRO-36.*, 2003, pp. 81–92.

[112] X. Zhang, V. Kumar, H. Oh, L. Zheng, G. S. May, A. Naeemi, and M. S. Bakir, *Impact of On-Chip Interconnect on the Performance of 3-D Integrated Circuits With Through-Silicon Vias: Part II*, vol. 63, no. 6, pp. 2510–2516, 2016.

[113] K. Tu, “Reliability challenges in 3d ic packaging technology,” vol. 51, no. 3, pp. 517–523, 2011.

[114] J. R. Black, “Electromigration - a brief survey and some recent results,” *IEEE Trans. Electron Devices*, vol. 16, no. 4, pp. 338–347, 1969.

[115] B. Lee and Y. Lee, “On-chip decouplig capacitor preplacement for power integrity enhancement,” in *2013 IEEE Electrical Design of Advanced Packaging Systems Symposium (EDAPS), Nara, 2013*, 2013, pp. 48–51.

[116] S. Zhao, K. Roy, and C. Koh, “Decoupling capacitance allocation and its application to power-supply noise-aware floorplanning,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 1, pp. 81–92, 2002.

[117] K. Matsumoto, S. Ibaraki, K. Sueoka, K. Sakuma, H. Kikuchi, Y. Orii, and F. Yamada, “Experimental thermal resistance evaluation of a three-dimensional (3d) chip stack,” in *2011 27th Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, 2011, pp. 125–130.

[118] S. Li, J. H. Ahn, R. Strong, J. Brockman, D. Tullsen, and N. Jouppi, “Mcpat: An integrated power, area, and timing modeling framework for multicore and manycore architectures,” in *Proc. Annual Int. Symp. Microarchitecture*, 2009, pp. 469–480.

[119] V. J. Reddi, H. Yoon, and A. Knies, “Two billion devices and counting,” *IEEE Micro*, vol. 38, no. 1, pp. 6–21, 2018.

[120] JEDEC, *LPDDR3 and LPDDR4:How Low-Power DRAM Can Be Used in high-Bandwidth Applications*, `http://www.jedec.org/https://www.`

jedec.org/sites/default/files/M_Greenberg_Mobile%20Forum_May_%202013_Final.pdf.

[121] F. W. Grover, *Inductance Calculations: Working Formulas and Tables*. Dover Publications, New York, 1946.

[122] K. Gala, V. Zolotov, R. Panda, B. Young, J. Wang, and D. Blaauw, "On-chip inductance modeling and analysis," in *Proceedings 37th Design Automation Conference*, 2000, pp. 63–68.

[123] I. Ciofi, A. Contino, P. J. Roussel, R. Baert, V. Vega-Gonzalez, K. Croes, M. Badaroglu, C. J. Wilson, P. Raghavan, A. Mercha, D. Verkest, G. Groeseneken, D. Mocuta, and A. Thean, "Impact of wire geometry on interconnect rc and circuit delay," *IEEE Transactions on Electron Devices*, vol. 63, no. 6, pp. 2488–2496, 2016.

[124] S. W. Ho, F. M. Daniel, L. Y. Siow, W. H. SeeToh, W. S. Lee, S. C. Chong, and S. R. Vempati, "Double side redistribution layer process on embedded wafer level package for package on package (pop) applications," in *2010 12th Electronics Packaging Technology Conference*, 2010, pp. 383–387.

[125] B. Chava, K. A. Shaik, A. Jourdain, S. Guissi, J. Ryckaert, G. V. D. Plaas, A. Spessot, E. Beyne, and A. Mocuta, "Backside power delivery as a scaling knob for future systems," in *SPIE Advanced Lithography*, doi:10.1117/12.2514942, vol. 10962, 2019.

[126] C. Auth, A. Aliyarukunju, M. Asoro, D. Bergstrom, V. Bhagwat, J. Birdsall, N. Bisnik, M. Buehler, V. Chikarmane, G. Ding, Q. Fu, H. Gomez, W. Han, D. Hanken, M. Haran, M. Hattendorf, R. Heussner, H. Hiramatsu, B. Ho, S. Jaloviar, I. Jin, S. Joshi, S. Kirby, S. Kosaraju, H. Kothari, G. Leatherman, K. Lee, J. Leib, A. Madhavan, K. Marla, H. Meyer, T. Mule, C. Parker, S. Parthasarathy, C. Pelto, L. Pipes, I. Post, M. Prince, A. Rahman, S. Rajamani, A. Saha, J. D. Santos, M. Sharma, V. Sharma, J. Shin, P. Sinha, P. Smith, M. Sprinkle, A. S. Amour, C. Staus, R. Suri, D. Towner, A. Tripathi, A. Tura, C. Ward, and A. Yeoh, "A 10nm high performance and low-power cmos technology featuring 3rd generation finfet transistors, self-aligned quad patterning, contact over active gate and cobalt local interconnects," in *2017 IEEE International Electron Devices Meeting (IEDM)*, doi:10.1109/IEDM.2017.8268472, 2017, pp. 2911–2914.

[127] A. Gupta, S. Kundu, L. Teugels, J. Bommels, C. Adelmann, N. Heylen, G. Jamieson, O. V. Pedreira, I. Ciofi, B. Chava, C. J. Wilson, and Z. Tokei, "High-aspect-ratio ruthenium lines for buried power rail," in *2018 IEEE International Interconnect Technology Conference (IITC)*, doi:10.1109/IITC.2018.8430415, 2018, pp. 4–6.

[128] B. Chava, J. Ryckaert, L. Mattii, S. M. Y. Sherazi, P. Debacker, A. Spessot, and D. Verkest, *Dtco exploration for efficient standard cell power rails*, 2018.

[129] I. Ciofi, P. J. Roussel, Y. Saad, V. Moroz, C. Hu, R. Baert, K. Croes, A. Contino, K. Vandersmissen, W. Gao, P. Matagne, M. Badaroglu, C. J. Wilson, D. Mocuta, and Z. Tőkei, "Modeling of via resistance for advanced technology nodes," *IEEE Transactions on Electron Devices*, vol. 64, no. 5, pp. 2306–2313, 2017, doi:10.1109/TED.2017.2687524.

VITA

Md Obaidul Hossen was born in Chittagong, Bangladesh, in February 1989. He received his B.S. degree in electrical and electronic engineering from Bangladesh University of Engineering and Technology, Bangladesh, in 2013. He is currently pursuing the Ph.D. degree in electrical engineering at Georgia Institute of Technology, Atlanta, GA, USA.

He joined the PhD program at Georgia Tech in 2014. In August 2014, he joined Integrated 3-D System Group supervised by Dr. Muhannad S. Bakir. His promary research is in the area of 2.5-D and 3-D IC design with a focus on thermally aware power delivery network design. His other research interests include network-on-chip modeling and algorithm development.