

**COMPACT PHYSICAL MODELS FOR POWER SUPPLY NOISE  
AND CHIP/PACKAGE CO-DESIGN IN GIGASCALE  
INTEGRATION (GSI) AND THREE-DIMENSIONAL (3-D)  
INTEGRATION SYSTEMS**

A Thesis  
Presented to  
The Academic Faculty

by

Gang Huang

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, 2008

**COMPACT PHYSICAL MODELS FOR POWER SUPPLY NOISE  
AND CHIP/PACKAGE CO-DESIGN IN GIGASCALE  
INTEGRATION (GSI) AND THREE-DIMENSIONAL (3-D)  
INTEGRATION SYSTEMS**

Approved by:

Dr. James D. Meindl, Advisor  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Jeffery A. Davis  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

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

Dr. Thomas K. Gaylord  
School of Electrical and Computer  
Engineering  
*Georgia Institute of Technology*

Dr. Paul A. Kohl  
School of Chemical and Biomolecular  
Engineering  
*Georgia Institute of Technology*

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

Date Approved: September 23<sup>rd</sup>,  
2008

To My Beloved Parents and Wife

## **ACKNOWLEDGEMENTS**

I would like express my thanks to my advisor, Dr. James D Meindl for his guidance and support through the course of my Ph.D. study. It is him whose professionalism, kindness, and wisdom teach me and influence me into who I am today. I feel very fortunate to have a close relationship and interaction with Dr. Azad Naeemi and Dr. Muhannad Bakir, and working with them makes my Georgia Tech experience so enjoyable. I also would like to thank Dr. Thomas Gaylord for serving as my proposal committee chair and dissertation reading committee member. He gave me lots of suggestions about how to improve my professional skills. Dr. Jeff Davis has been kind enough to serve both at my proposal and dissertation committees. I would like to thank him as well. I also want to express my thanks to Dr. Paul Kohl for his attendance in my committee and his leadership in Interconnect Focus Center (IFC). Dr. Ronald Harley is acknowledged for serving on my proposal committee.

I gratefully acknowledge the support of the Pakcage and I/O team, IBM, Poughkeepsie, NY. They provided me an internship which allowed me to validate my models using industry designs.

I express my deepest gratitude to my beloved parents who are always proud of me. Most important of all, I would like to thank my wife, Peijuan, for her understanding and endless devotion. This thesis concludes my life as a pure student. Friends certainly play a very important role during the more than 20 years of study. I want to thank them for their supports and encouragements.

# TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
SUMMARY	xiii
CHAPTER 1: Introduction and Background	1
1.1 Introduction	1
1.2 Compact Physical Models for Power Supply Noise	5
1.3 Chip/Package Co-Design	8
1.4 Power Integrity Issues Casued by Power Gating and Clock Gating Techniques	9
1.5 Power Integrity Issues Casued by Hot Spots	10
1.6 Power Integrity Issues Casued by 3-D Chip Stacks	10
1.7 Impact of Noise on On-Board Transmission Lines	11
1.8 Conclusion	11
CHAPTER 2: Blockwise Compact Physical Models for Power Supply Noise and Chip/Package Co-Design	13
2.1 Introduction	13
2.2 On-Chip Power Distribution Network	14
2.3 Partial differential equation for power distribution networks	16
2.4 Simplified Transfer Impedance Function	24
2.5 Analytical Solution for Noise Transients	26
2.6 Analytical solution of peak noise	29
2.7 Technology trends of power supply noise	34
2.8 Conclusion	36

CHAPTER 3: Compact Physical Models for Power Supply Noise with the Consideration of Hot Spots	37
3.1 Introduction	37
3.2 Mathematical Modeling for Hot Spots	38
3.3 Case study	42
3.4 Chip/Package Co-Design and Solutions to Suppress Noise	46
3.5 Conclusion	48
CHAPTER 4: Compact Physical Modeling and Design Implications of Power Distribution Networks for Three-Dimensional (3-D) Chip Stacks	49
4.1 Introduction	49
4.2 Mathematical Modeling for 3-D Chip Stacks	50
4.3 Model Validation	54
4.4 Design Implication for 3-D Integration	57
4.5 Solutions for Suppressing Noise Level in 3-D systems	61
4.6 Conclusion	66
CHAPTER 5: Compact Physical Models for Chip and Package Power and Ground Distribution Networks	68
5.1 Introduction	68
5.2 Mathematical Modeling for Chip and Package Power Distribution Networks	69
5.3 Model Validation by SPICE	76
5.4 Model Validation by SPEED2000	78
5.5 Conclusion	82
CHAPTER 6: Compact Physical Modeling to Minimize Energy-Per-Bit for On-Board LC Transmission Lines under Noise Conditions	83
6.1 Introduction	83
6.2 Noise Aware Bit-rate Limit	84
6.3 Minimized Energy-Per-Bit	86

6.4 Wiring Area Overhead	88
6.5 Conclusion	94
CHAPTER 7: Conclusions and Future WorkS	95
7.1 Conclusions of Dissertation	95
7.2 Power Supply Noise Analysis for Multicore Microprocessors	97
7.3 Optimizations of the Fluidic I/O and TSV Networks in 3-D Chip Stacks	98
Appendix A: Derivation for Partial Differential Equation (2.1)	99
Appendix B: Derivation to Obtain the Solution for (2.5)	104
REFERENCES	109
VITA	116

## LIST OF TABLES

	Page
Table 2.1 Table of pad shape parameters	20
Table 2.2 Range of main variable values	25
Table 5.1 Parameters for the comparison between SPICE and compact physical models	77



## LIST OF FIGURES

	Page
Figure 1.1 Simulated noise droops of Intel microprocessors [8]	2
Figure 1.2 Impact of compact physical models on power integrity aware design flow; (a) without compact physical models; (b) with compact physical models	7
Figure 2.1 Current transient pattern when a circuit block is powering up	13
Figure 2.2 On-chip power/ground grids and I/O pads in flip-chip technology	15
Figure 2.3 Division of power grid into independent cells	16
Figure 2.4 Simplified circuit model for GSI power distribution systems	18
Figure 2.5 Differential model of a node for single power/ground grid	18
Figure 2.6 Representation of the current flowing through the pad as a Dirac Delta function	20
Figure 2.7 Transfer impedance for three corner points; (a) Locations of three corner points; (b) Comparison between (2.8) and the results of SPICE simulations	23
Figure 2.8 Transfer impedance for three corner points: Comparison between (2.8), (2.9) and the results of SPICE simulations	26
Figure 2.9 Two parts for the power noise waveform	27
Figure 2.10 Power supply noise waveforms for three corner points: Comparison between (2.23) and the results of SPICE simulations	29
Figure 2.11 Noise values vs. location of nodes for IR-drop and peak noise. $L_p=0.5$ nH, 10% chip area is occupied by decaps, and $R_s=1.1$ $\Omega$	30
Figure 2.12 The worst case peak noise as a function of the chip area occupied by decaps: Comparison between (2.29) and the results of SPICE simulations for a pair of grids	31
Figure 2.13 The worst case peak noise as a function of $L_p$ : Comparison between (2.29) and the results of SPICE simulations for a pair of grids	32

Figure 2.14 The worst case peak noise as a function of grid fineness (the number of power (ground) wires between two power (ground) pads) : Comparison between (2.29) and the results of SPICE simulations for a pair of grids	33
Figure 2.15 The worst case peak noise as a function of segment resistance for different pad sizes: Comparison between (2.29) and the results of SPICE simulations	33
Figure 2.16 The worst case peak noise as a function of the number of pads: Comparison between (2.29) and the results of SPICE simulations	34
Figure 2.17 Technology trends of the worst case peak noise	35
Figure 3.1 Current map of Intel Itanium® Processor [2]	37
Figure 3.2 Simplified circuit model for GSI power distribution system with a hot spot	38
Figure 3.3 Circuit model for single grid structure and the square region allocated for the analysis	39
Figure 3.4 Illustration of the switching block, the hot spot and the 6x6 pad regions allocated for the analysis	42
Figure 3.5 Frequency domain noise response for the center point of the hot spot. (a) Magnitude response, (b) Phase response	44
Figure 3.6 Transient noise waveforms using SPICE simulations and different models	45
Figure 3.7 (a) Configurations of added pads within the hot spot and peak noise for different pad allocation schemes. (b) Noise waveforms for different pad allocation schemes with a hot spot current density of 400 A/cm <sup>2</sup>	47
Figure 4.1 Power integrity problem of 3-D chip stack	50
Figure 4.2 Division of the 3-D stack footprint	51
Figure 4.3 Simplified circuit model for 3-D stacked system	52
Figure 4.4 Circuit model for single grid structure	53
Figure 4.5 Five chip stacking for model validation: All dice are switching, and Die 3 is examined	55
Figure 4.6 Frequency response of the worst case noise for the third die: (a) Magnitude; (b) Phase.	56
Figure 4.7 Time domain response of the worst case noise for the third die	57

Figure 4.8 Single die switching, changing the switching die	58
Figure 4.9 Single die switching, increasing total number of dice	59
Figure 4.10 Making shorter interconnects between communicated blocks by using 3-D integration	60
Figure 4.11 All dice switching, increasing total number of dice	61
Figure 4.12 Effect of adding one “decap” die. (a) Single die switching; (b) Four switching dice without “decap” die; (c) Four switching dice with “decap” die at the bottom; (d) Four switching dice with “decap” die on the top	62
Figure 4.13 Effect of adding two “decap” dice. (a) Single die switching; (b) One “decap” die at the bottom and the other in the middle; (c) One “decap” die in the middle and the other on the top; (d) Both “decap” dice on the top	63
Figure 4.14 Effect of adding more TSVs: fixing the number of power/ground I/Os	64
Figure 4.15 Effect of adding more TSVs and power/ground I/Os	65
Figure 5.1 Power delivery system of GSI systems	70
Figure 5.2 Simplified circuit model of on-chip power/ground grids	71
Figure 5.3 Simplified circuit model of package power/ground planes	72
Figure 5.4 The setup of the boundary conditions for PDEs	74
Figure 5.5 Setup for the comparison between SPICE simulations and compact physical models	76
Figure 5.6 Frequency domain response at the center of the chip: (a) Magnitude; (b) Phase	78
Figure 5.7 Configuration of the IBM ceramic package module	79
Figure 5.8 Schematic view of the package module in SPEED2000	79
Figure 5.9 Divisions of the chip and package	81
Figure 5.10 Waveforms for the two locations as shown in Figure 5.9, the center of the chip and a decap location in the package	81
Figure 6.1 Power supply scheme for on-board interconnection	87

Figure 6.2 Normalized energy-per-bit vs. signal swing in different noise conditions	88
Figure 6.3 Illustration of data flux density, $\Phi_D$ (Error! Objects cannot be created from editing field codes.), where chip edge length $D_{chip}$ and wire width $w$ both have dimensions of cm	89
Figure 6.4 Data flux density vs. wire length marked by critical length of each technology generation	91
Figure 6.5 Data flux density vs. wire length marked by the critical length in different noise conditions based on the ITRS projections for the 45 nm node	92
Figure 6.6 Energy benefit factor and Area overhead factor vs. signal swing when $K_N=0.25$ , $V_{IN}=0.05V_{DD}$ .	93
Figure A.1 Circuit model for a node for a single grid	99

## SUMMARY

The objective of this dissertation is to derive a set of compact physical models addressing power integrity issues in high performance gigascale integration (GSI) systems and three-dimensional (3-D) systems. The aggressive scaling of CMOS integrated circuits makes the design of power distribution networks a serious challenge. This is because the supply current and clock frequency are increasing, which increases the power supply noise. The scaling of the supply voltage slowed down in recent years, but the logic on the integrated circuit (IC) still becomes more sensitive to any supply voltage change because of the decreasing clock cycle and therefore noise margin. Excessive power supply noise can lead to severe degradation of chip performance and even logic failure. Therefore, power supply noise modeling and power integrity validation are of great significance in GSI systems and 3-D systems.

Compact physical models enable quick recognition of the power supply noise without doing dedicated simulations. In this dissertation, accurate and compact physical models for the power supply noise are derived for power hungry blocks, hot spots, 3-D chip stacks, and chip/package co-design. The impacts of noise on transmission line performance are also investigated using compact physical modeling schemes. The models can help designers gain sufficient physical insights into the complicated power delivery system and tradeoff various important chip and package design parameters during the early stages of design. The models are compared with commercial tools and display high accuracy.

# CHAPTER 1: INTRODUCTION AND BACKGROUND

## 1.1 Introduction

Moore's law states that the number of transistors on a chip doubles about every two years [1], and semiconductor industry has kept this pace for nearly 40 years. New generation Intel® Itanium® processor Tukwila has been recently reported to have 2 billion transistors [2]. This realization of gigascale integration (GSI) along with the adoption of three-dimensional (3-D) integration has enhanced the performance of integrated systems to an unprecedented level [3, 4]. However, in GSI era, the power consumption of GSI chips is increasing at an alarming rate [5]. The increasingly faster devices packed at unprecedented densities result in higher current densities. The scaling of the supply voltage slowed down in recent years, but the logic on the integrated circuit (IC) still becomes more sensitive to any supply voltage change because of the decreasing clock cycle and therefore noise margin. With this trend, power supply noise, the voltage fluctuation on power delivery networks, has become a significant factor that can substantially influence the overall system performance. The design of power delivery systems becomes a very important and challenging task. Therefore, understanding complicated power delivery networks and supplying clean power to microprocessors is of great significance [6, 7].

IR-drop and  $\Delta I$  noise are the two main components of the power supply noise. IR-drop results from the supply current passing through the parasitic resistance of power distribution networks.  $\Delta I$  noise is caused by the inductance of the power delivery system, and becomes important when a group of circuits switch simultaneously. Power supply

noise consists of three distinct voltage droops [8], and they come from the interactions between the chip, package, and board. The three droops are illustrated as shown in Figure 1.1 [8].

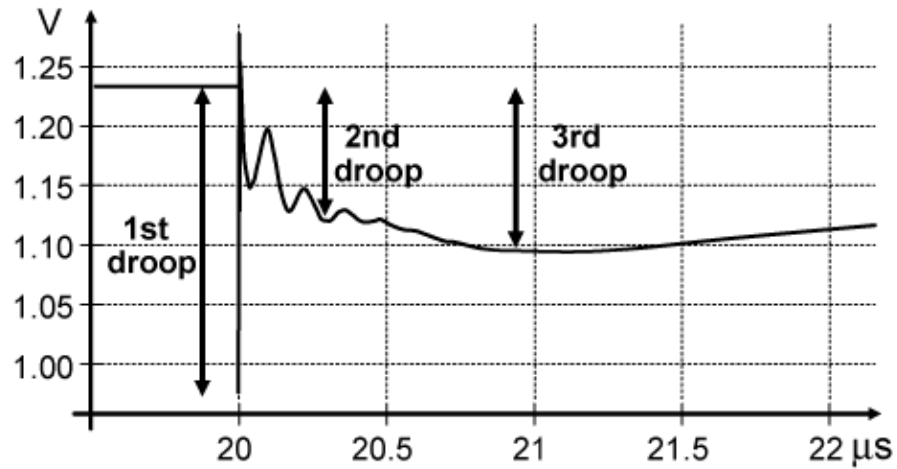


Figure 1.1 Simulated noise droops of Intel microprocessors [8]

The third droop is related to the bulk capacitors at the board level, and has time duration of a few microseconds. The third droop influences all critical paths but can be readily minimized by using more board space for bulk capacitors [6]. The second droop is caused by the resonance between the inductive traces on the motherboard and the decoupling capacitors (decaps) in the package. The second droop has time duration of a few hundred nanoseconds and impacts a significant number of critical paths. The first droop is caused by the package inductance and on-die capacitance. Because the resonance frequency of the first droop is in the range of tens of MHz to a few hundred of MHz depending on the sizes of package level components and on-chip decaps, the first droop noise is also called mid-frequency noise [9, 10]. Because putting additional on-chip decaps is very costly, among the three droops, the first droop is the most difficult one to suppress. The first droop noise has the largest magnitude. Even though the first droop has

the smallest time of occurrence it can adversely affect GSI circuits as its duration can be tens of nano seconds (ns). Chip performance can be severely degraded when the first droop affects some critical paths. Because of its severe impact on high-performance chips, the first droop is thus the main focus of this research.

Excessive power supply noise can lead to severe performance degradation of on-chip circuitry and off-chip high speed data links, and even result in logic failures [6]. Thus it is vitally important to model and predict the performance of power delivery networks with the objective of minimizing supply noise.

The modeling methods of power supply noise can be categorized into three types: partial element equivalent circuit (PEEC) method, electromagnetic solver, and lumped circuit model.

Power supply noise has been originally analyzed by extracting the circuit networks and later simulating the netlists with circuit simulators [11] since early 1990s. The PEEC method is the foundation of this type of modeling schemes [12]. A complete power supply distribution system includes package-level power distribution networks, on-chip power grids, and the equivalent circuits to represent switching functional blocks. Among the three major components, the package-level model is dominated by the inductance, the on-chip networks are dominated by the resistance, and the switching circuit blocks determine the current patterns throughout the chip. These circuit element values can be extracted from physical designs using extraction tools [13, 14, 15], and then the extracted RLC networks together with the switching circuit models can be simulated by circuit simulators such as SPICE [16]. A significant advantage of using the PEEC method is that the self and mutual inductances can be extracted without knowing



the return paths beforehand. However, because mutual inductance exists between any two segments in the circuit, the inductance matrix is large and dense. This makes the simulations of large circuits with millions of wire segments almost impossible. To simulate large circuits like power distribution networks, circuit matrix sparsification techniques [17, 18] and model order reduction methods [19, 20] have been developed for reducing matrix sizes.

With the increase of clock frequencies, the frequency dependent characteristics of power delivery networks become more and more significant, especially for the package level power and ground planes; hence, distributed modeling work based on the discretization of Maxwell's equations becomes necessary. Methods based on electromagnetic solutions can be formulated in the time or frequency domains by solving differential or integral equations. The work presented in [21], which is based on the finite difference time domain (FDTD) method, has been transferred into commercial tool SPEED2000 [22]. SPEED2000 is the first commercialized tool available for performing the noise transient electromagnetic simulation at package and board levels. Since the frequency domain methods, such as transmission matrix method (TMM) [23] and integral equation method [24], are able to capture different resonances in frequency domain, the frequency domain methods are more accurate than the time domain methods. However, the time domain methods are preferred for larger problem, because they need comparably less computational resources and can run much faster.

Before the detailed physical design is started, it is of great importance to have a quick snapshot on the power supply noise of various parts of the power delivery system. During early design stages, a good use of simple lumped models can avoid lots of

redundant dedicated simulations at later stages of design. In [25], an approach based on target impedance is proposed to estimate decaps needed for a given power distribution networks. It is a frequency domain method using the lumped package model to maintain the power distribution network impedance less than a target impedance value. This methodology provides a closed-form expression and is useful for fast “what-if” analysis. The lumped circuit model is also used to perform the post-design validation of Pentium<sup>®</sup> III and Pentium<sup>®</sup> 4 microprocessors in [26]. The model predictions are compared with measured data and provide useful insights in investigating the model regions of validity. The model can predict the impact of the second and third droops on chip performance, but the first droop is much harder to predict because the lump model can not incorporate the propagation and locality of the noise throughout the chip and package. This work indicates the drawback of lumped circuit models.

To tackle the drawback of lumped models and have an accurate and quick recognition of the first droop power supply noise in early design stages while avoiding dedicated simulations, compact physical models are developed in this research.

## **1.2 Compact Physical Models for Power Supply Noise**

To design an optimum system, there are major decisions that must be made at early design stages even before detailed physical layout designs start. If problems associated with the design and implementation of a power distribution network are undetected early in the design cycle, they can become very costly to fix in later design stages [27, 28]. An over-designed power distribution system would result in an expensive package and waste of the silicon and interconnect resources. An under-designed system (even partly) can lead to noise problems and difficulties regarding wire routing. As a

result, to gain sufficient physical insight, compact and accurate physical models are needed to model the complicated power distribution network. Such models would be critical in the early stages of design and can estimate the on-chip and off-chip resources needed for the power distribution network. Compact physical models can help designers perform quick assessment of the power integrity issues in GSI and 3-D systems.

In a power integrity aware design flow, as shown in Figure 1.2 (a), lumped models are used for prep-physical-design (Pre-PD) validation to help validate the power supply noise level and make key design decisions such as the type of package, number of grid and plane levels, and tentative size/number of decaps. Pre-PD validation enables detailed physical design to start from a reasonable point. During physical design phases, circuit simulators (PEEC method based) and 3D solvers (electromagnetic solver based) are used to perform dedicated simulations and to help refine the physical designs. At least several design iterations are needed to make the design converge and meet design requirements. However, package and chip power delivery network models can be very large, and manipulating large networks by simulation is time-consuming. Each of the design iterations may take days. It is noted that the initial design point is of great importance for later design stages, and an accurate early stage prediction can significantly decrease the number of design iterations. This effort can be served by compact physical models. Compact physical models are derived from the fundamental physical basis and are able to incorporate the distributed nature of power/ground grids and planes [27]. Compact physical models are more accurate than lumped models while with similarly simple form. Replacing lumped circuit models with compact physical models in Pre-PD validation, as shown in Figure 1.2 (b), designers can be allowed to greatly reduce the

number of iterations and therefore the length of design cycle. Compact physical models consider the propagation and locality of the noise, and thus are able to lead to an initial physical design closer to the target optimum design.

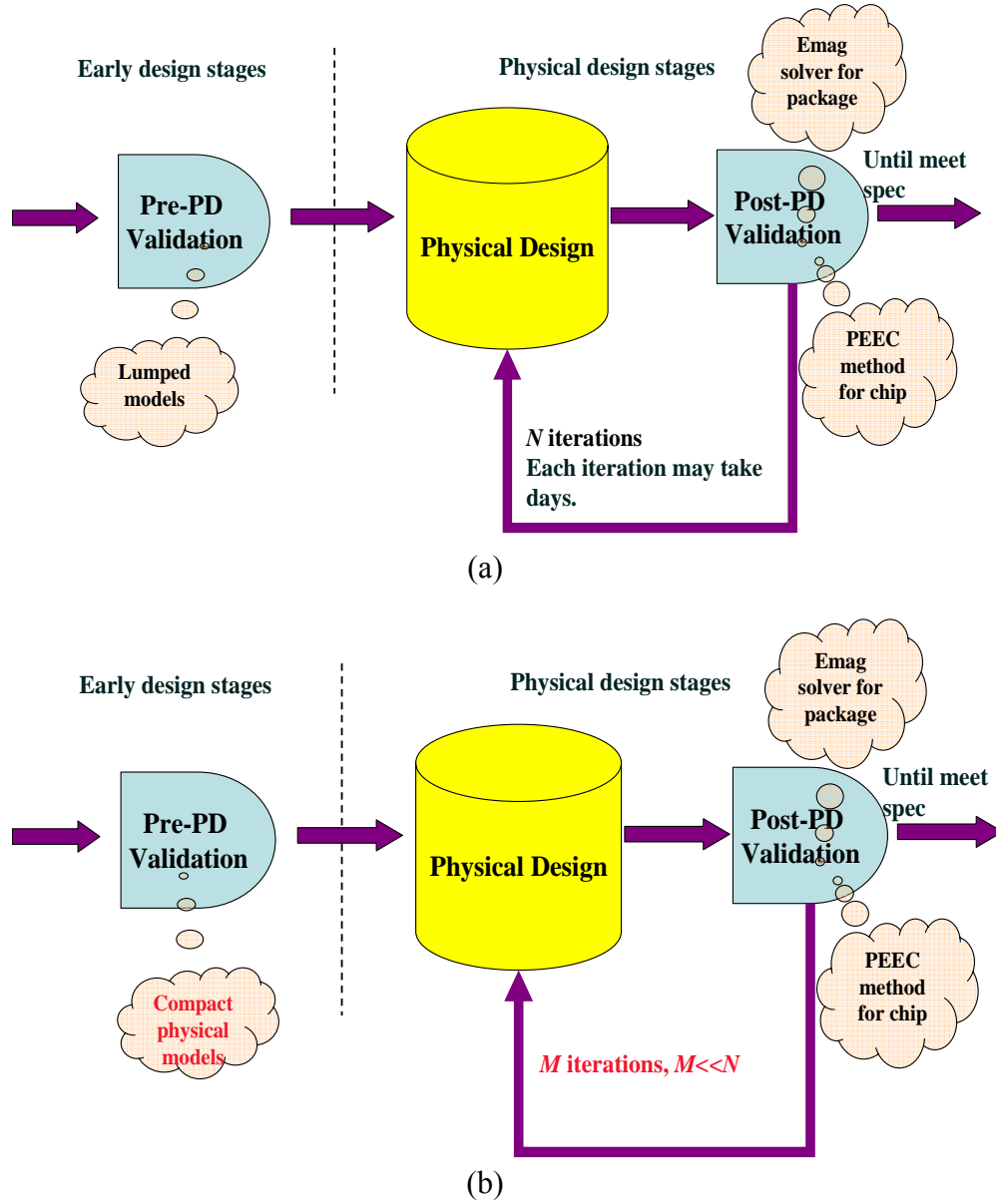


Figure 1.2 Impact of compact physical models on power integrity aware design flow; (a) without compact physical models; (b) with compact physical models

With technology scaling, power integrity is a big challenge when noise margins have been largely reduced because of the decreasing clock cycle. Therefore, it is important to quantify the relationship between the noise and various technology parameters. Compact physical models can be used to predict the power noise trends of different generations of technology from mathematical and physical bases. The predictions can provide meaningful design implications and help examine potential solutions [27, 28].

Compact physical models for IR-drop have been recently proposed in [28]. These models embody the distributed nature of on-chip power grids and display high accuracy when predicting the IR-drop of GSI power distribution networks. Power supply noise is a dynamic effect changing with time, and IR-drop is the case when the noise goes to steady state. The transient part of the power supply noise,  $\Delta I$  noise, is more significant in determining the timing budget of a system. In this dissertation, for the first time, compact physical models are developed for the first droop  $\Delta I$  noise in GSI and 3-D systems.

### **1.3 Chip/Package Co-Design**

Flip-chip packages introduce significant chip and package design complexities [29]. This chip/package combination must be viewed throughout various design phases especially for power integrity aware designs. The first droop noise is caused by the chip/package resonance (on-chip decaps and package inductance), and thus optimum designs require integrated chip/package co-design efforts to quickly evaluate the tradeoffs available in silicon chips and packages. It is necessary to build up a seamless linkage between chip and package designs, and this requires unified tool platforms that can make the real co-designs happen.

However, in reality, two sets of tools (PEEC method and SPICE simulations for chips and electromagnetic solvers for packages) are used for chip and package modeling. Those tools are based on different methodologies and bridging them becomes challenging. For flip-chip technology, on the one hand, some information of packages is needed for chip designers, such that some key design concerns, such as the amount of decaps to suppress chip/package resonance, can be considered in the early stages of chip design (decap allocation is considered even in the floorplanning phase of chip design [30]); on the other hand, thousands of inputs/outputs (I/Os) connect a chip to a package, and package designers must have the information about the chip to specifically handle critical I/Os.

Another important goal of this research is to use compact physical models to fulfill chip/package co-design of power distribution networks and to balance multiple design considerations, even when detailed physical designs have not been started yet during early design stages.

#### **1.4 Power Integrity Issues Casued by Power Gating and Clock Gating Techniques**

To address the issues of the excessive power dissipation in GSI systems, power management techniques, such as power gating [31] and clock gating [32], are widely adopted in nowadays chip designs. The basic ideas of these techniques are to dynamically switch off the funtional blocks that are in the idle state. However, when the blocks wake up, large current transients are induced from the power supply. Therefore these current changes lead to time-varying  $\Delta I$  noise and IR-drop when flowing through the inductive and resistive components on power distribution networks [33]. The noise seriously impacts the timing of the circuits in the functional blocks, and consequently slows down

the blocks. Especially power hungry blocks in today's chips can induce tens of and even over a hundred ampere current within in a very short time period (less than 1 ns). Thus, modeling the power distribution network with the objective of minimizing the supply noise resulting from power hungry blocks becomes highly necessary.

### **1.5 Power Integrity Issues Casued by Hot Spots**

The non-uniformity of the power density distribution throughout the die arises with the increasing functional complexities of microprocessors. It is very common to see local power densities greater than  $300 \text{ W/cm}^2$  for today's high-performance chips [34]. These extremely high power density regions are also called hot spots. Hot spots require advanced thermal solutions, but also challenge the design of power delivery systems. The non-uniformity of the power density distribution leads to high noise level at hot spot locations. Thus, it is meaningful to investigate this non-uniformity problem and to deliver compact physical models to address the power integrity issues casued by hot spots.

### **1.6 Power Integrity Issues Casued by 3-D Chip Stacks**

Three-dimentional (3-D) integration technology enables a new regime of design in terms of improving multi-functional integration, improving system speed and reducing power consumption [35]. At the high performance end, industry already started to develop 3-D microprocessor stacking and microprocessor-memory stacking [36, 37]. However, stacking multiple high-performance dice may result in severe power integrity problem. Several hundred amperes of current need to be delivered to a limited footprint area, and the supply current flows through the micro-bumps and narrow through-silicon-vias (TSVs) that may have large parasitics. These may potentially lead to a large  $\Delta I$  noise if stacked chips switch simultaneously. It is of great importance to obtain physical insight

into the complicated 3-D power delivery networks. Thus, the power distribution networks in 3-D systems need to be accurately and efficiently modeled to support 3-D system designs.

### **1.7 Impact of Noise on On-Board Transmission Lines**

Chips become more and more power-limited in GSI era [5, 38, 39]. At the same time, the demand for I/O bandwidth increases because of the increases in the speed and integration level of chips, and hence the power dissipation in I/O drivers becomes a major issue. Low-swing signaling is suggested as a viable technique for lowering the power dissipation [40, 41]. It is desirable to send a certain amount of information dissipating minimum possible energy. Because of the receiver sensitivity and some noise sources independent to the signal swing, lowering the signal swing requires larger bit duration such that a signal can reach a level large enough to fight against the noise. This increase in bit duration, however, increases the energy to transfer one bit of information, namely energy-per-bit. Thus, it is valuable to find the tradeoff between the bit duration and signal swing and to obtain the minimum energy-per-bit.

### **1.8 Conclusion**

Understanding the complicated power delivery networks is necessary to ensure power integrity for GSI and 3-D systems. Noise margin is shrinking with technologies because of the increase of supply current and the decrease of clock cycle time, and thus delivering clean power can greatly help enhance system performance and extend Moore's law. This chapter introduces the origin of the power supply noise modeling and the past works pertinent to this research. The goal of this research is to derive a set of compact physical models for the first droop power supply noise and chip/package co-design.



The outline of this dissertation is as follows. In Chapter 2, blockwise compact physical models are derived for the first droop power supply noise for power hungry blocks assuming uniform switching conditions. Analytical models are then introduced in Chapter 3 to extend blockwise models to general non-uniform switching conditions such as the hot spot case. To help identify the challenges brought by 3-D integration, models derived in Chapter 2 are also adapted in Chapter 4 to consider 3-D chip stacks. In Chapter 5, compact physical models with detailed package descriptions are derived. The models are validated by commercial power integrity tools using IBM package designs. The models enable real chip/package co-design in early design stages. The noise results in performance degradation of off-chip high speed links. In Chapters 6, the impact of noise on board-level transmission lines is investigated. Finally, conclusions and future work are portrayed in Chapter 7.

.

## CHAPTER 2: BLOCKWISE COMPACT PHYSICAL MODELS FOR POWER SUPPLY NOISE AND CHIP/PACKAGE CO-DESIGN

### 2.1 Introduction

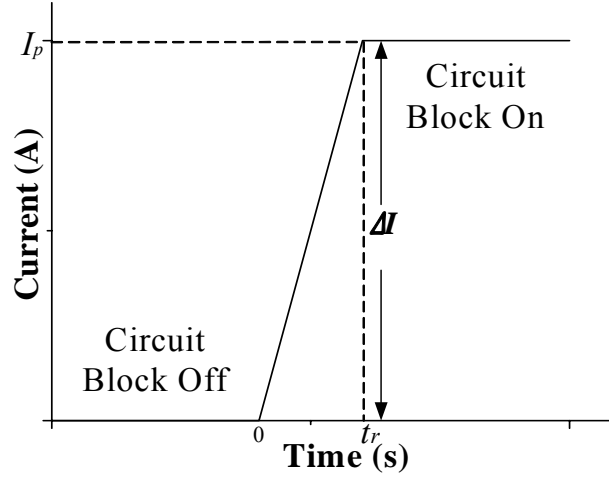


Figure 2.1 Current transient pattern when a circuit block is powering up

High performance GSI systems incur high power dissipation. Therefore there are several runtime power management techniques, such as power gating and clock gating, which dynamically switch off idle circuit blocks. Power gating technique disconnects idle blocks from the power distribution network, reducing leakage power [31]. Clock gating disables the clock signal and saves dynamic power [32]. However, when a functional block is switching from off state to on state (or from on to off), large current transients, as shown in Figure 2.1, occur on the power distribution network to power on (or power off) the block. These sharp current changes lead to the time-varying  $\Delta I$  noise and IR-drop when the current flows through the inductive and resistive components on the power distribution network [33]. The noise in turn causes timing divergence of the circuits in the

functional block, and consequently slows down the block. Power gating and clock gating are widely used nowadays in high performance IC products, and demonstrate power saving up to 60% [33]. These power management techniques inevitably result in difficulties for power delivery. Especially power hungry blocks in today's chips can induce tens of and even over a hundred ampere current within a very short time period (less than 1 ns). Satisfying given noise margin (such as 15% for IBM Power5 design [42]) requires great discretion in power delivery system designs, and modeling the power distribution network with the objective of minimizing the supply noise resulting from power hungry blocks becomes highly necessary.

In this chapter, a set of blockwise compact physical models for the first droop supply noise are derived. These models can be applied to a functional block with a large number of power and ground I/O pads and can give a quick snapshot of the power supply noise for power hungry blocks. These models are able to accurately capture the impact of package parameters as well as the distributed nature of power grids and decaps.

## **2.2 On-Chip Power Distribution Network**

On-chip power distribution networks consist of global and local networks. Global power distribution networks carry the supply current and distribute power across the chip. Local networks deliver the supply current from global networks to the active devices. Global networks contribute most of the parasitics, and thus are the main concern of this chapter. For global distribution networks, the most common way is to use a grid made of orthogonal interconnects routed on separate metal levels connected through vias [43]. Another method is to dedicate a whole metal level to power and another level to ground. The advantage of this method is the small parasitics and as a result small voltage drop.

But it is relatively expensive and has been reported only in the Alpha 21264 microprocessor [44]. This research will mainly focus on the grid structure.

Wire-bond and flip-chip technologies are the two most commonly used chip-to-package interconnects [29]. Wire-bond is cheaper than flip-chip interconnect; however, peripheral wire-bond interconnect causes higher power supply noise level because of larger parasitics. In flip-chip technology, the parasitics are reduced by spreading I/O pads over the surface area of the chip, therefore reducing the noise. The development of GSI systems is not only driven by more efficient silicon real estate usage but also by more I/O counts. Hence most of today's high performance designs are using flip-chip interconnect and area-array I/Os to provide larger bandwidth for chip to the next level interconnections. As such, the focus of this section is the power supply noise in flip-chip technology.

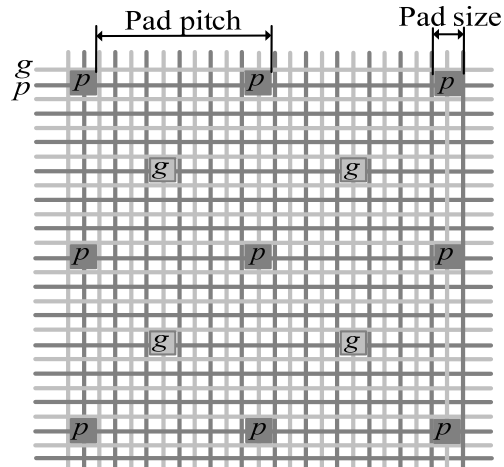


Figure 2.2 On-chip power/ground grids and I/O pads in flip-chip technology

As shown in Figure 2.2, the grid structure and area-array I/O pad allocation are adopted in this research [27, 28]. Power is fed through the power pads from the package.

The current flows through power wires and on-chip circuits, and returns to the package through ground wires and ground pads.

### 2.3 Partial differential equation for power distribution networks

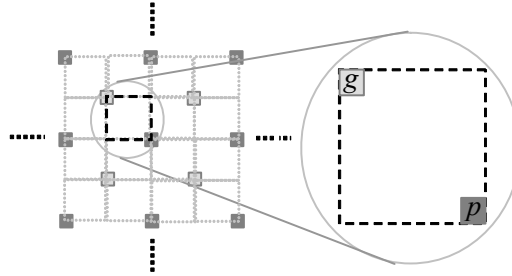


Figure 2.3 Division of power grid into independent cells

A microprocessor chip is composed of various functional blocks such as ALUs, caches, etc. Power supply noise is modeled assuming the switching current and decap distribution within a given functional block are uniform. For a functional block with a large number of power and ground pads, the block can be divided into cells, and each cell is the identical square region between a pair of adjacent quarter power and ground pads. It can be considered that no current passes normally through cell borders. One cell is thus enough for the power supply noise analysis, as shown in Figure 2.3.

For today's microprocessors, for example, Intel Itanium<sup>®</sup> [2], the power (current) densities of most of the functional blocks are approximately uniform, and the sizes of the these blocks are large enough to be covered by a large number of power and ground pads. For these blocks, this idealized assumption can be well applied and the fast analysis can be performed for most parts of chip. For the blocks with non-uniform power (current) density distribution and the blocks covered by a small number of pads, the assumption is no longer valid. More complicated modeling schemes to consider the current flowing

through cell borders need to be adopted, and Chapter 3 will specifically address this non-uniformity problem.

Recent study demonstrates that the package inductance is still overwhelming the on-die grid inductance [45], and the impact of the on-chip inductive effect remains insignificant when the clock frequency is less than 5 GHz. Because of the power issue, increasing the number of processor cores becomes the driving force to enhance chip performance instead of increasing clock frequencies. Most of the microprocessors will work at frequencies less than 5 GHz in the near future, and thus the on-chip inductive effect will be still a secondary effect. Only resistance is considered to model the power/ground grids in this research.

The simplified circuit model of the power distribution network associated with a cell is shown in Figure 2.4. The segment resistance of the grid is represented by  $R_s$ . Switching current between a power grid node and the adjacent ground grid node is modeled as a current source, and  $J(s)$  represents the switching current density in the Laplace domain (to enable the analysis of the current transients in a wide frequency range). Symbol  $C_d$  denotes the decoupling capacitance (including both the intentionally added decaps and the equivalent capacitance of the non-switching transistors) per unit area. Symbols  $\Delta x$  and  $\Delta y$  represent the distances between two adjacent power (or ground) nodes at the same wiring level for  $x$  and  $y$  directions, respectively. Symbol  $L_p$  ( $4L_p$  for quarter pad) represents the per pad loop inductance of the package.

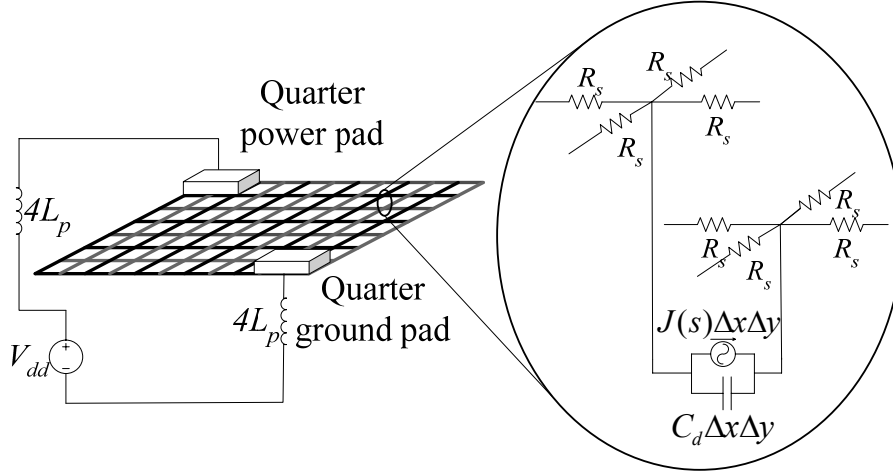


Figure 2.4 Simplified circuit model for GSI power distribution systems

Because the on-chip inductive coupling between the power and ground grids is neglected, the double grid structure can be decoupled into two individual grids, as shown in Figure 2.5. Assuming the voltage of a given point  $(x,y)$  in a grid is  $V(x,y,s)$ , the voltage of this point can be calculated from the following partial differential equation by using Kirchhoff's current law [28].

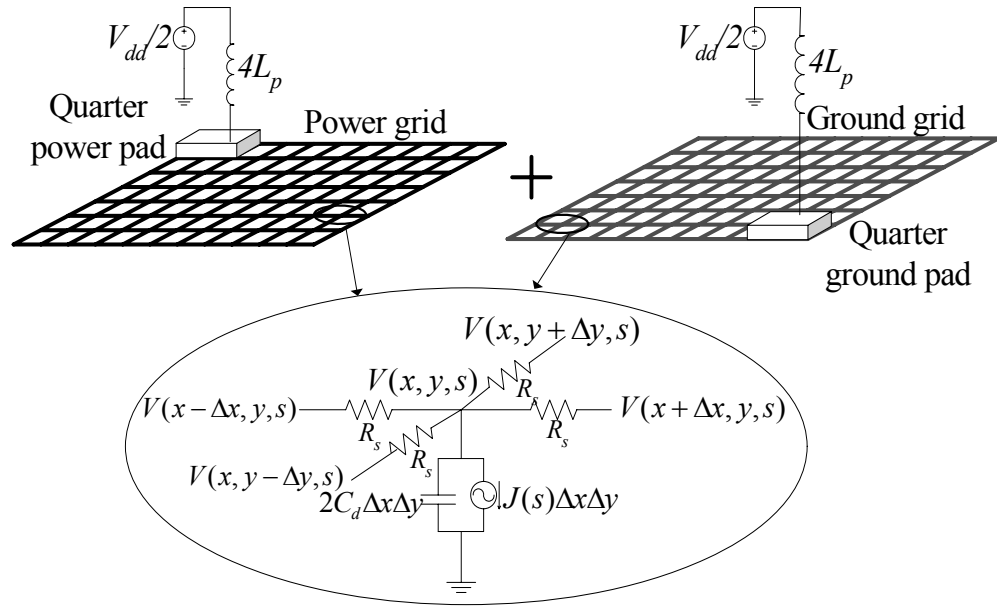


Figure 2.5 Differential model of a node for single power/ground grid

$$\nabla^2 V(x, y, s) = R_s J(s) + 2V(x, y, s) \cdot s R_s C_d + \Phi(x, y, s) . \quad (2.1)$$

In (2.1),  $\Phi(x,y,s)$  is the source function of this differential equation and is added to represent the voltage drop on  $L_p$ . As there is no current flowing through the cell boundaries, equation (2.2) should satisfy the following boundary conditions [27]:

$$\frac{\partial V(x,y,s)}{\partial y}\bigg|_{x=0}=0, \quad \frac{\partial V(x,y,s)}{\partial y}\bigg|_{x=a}=0, \quad \frac{\partial V(x,y,s)}{\partial x}\bigg|_{y=0}=0, \quad \frac{\partial V(x,y,s)}{\partial x}\bigg|_{y=a}=0, \quad (2.2)$$

where  $a$  denotes the size of the square cell.

When high frequency analysis is performed, for the circuit models shown in Figure 2.4 and 2.5, DC power supply source  $V_{dd}$  can be removed and the pad is directly connected to the ground through the package inductance (represented by  $4L_p$ ). It is necessary to know the amount of current  $I_{pad}(s)$  flowing through the quarter pad, since mathematically it determines the source function of the partial differential equation described by (2.1). Symbol  $I_{pad}(s)$  denotes the current delivered to this cell. This current is not equal to the total switching current within a cell, since at high frequencies, on-die decaps, as charge reservoirs, will supply some of the current. To accurately model  $I_{pad}(s)$ , we need to calculate the voltage on the pad, and the idea of equivalent pad radius can be used to calculate this voltage [28]. The concept of equivalent pad radius is as follows. By multiplying the edge length of the pad with a pad shape parameter  $\alpha$ , as shown in Table 2.1, the voltage contours resulting from a square pad are the same as the voltage contours from an equivalent circular pad [28]. Assuming the resistance of the pad is negligible  $I_{pad}(s)$  can be written as

$$I_{pad}(s) = -\frac{V(\alpha D_{pad}, 0, s)}{4sL_p}, \quad (2.3)$$

where  $D_{pad}$  denotes the edge length of the quarter square pad.



From a mathematical point of view, the effect of current flowing through a quarter pad is equivalent to putting a Dirac delta source function  $\Phi(x,y,s)$  at the corner point  $(0,0)$ , as shown in Figure 2.6.  $\Phi(x,y,s)$  has a form of

$$\Phi(x, y, s) = -R_s \frac{V(\alpha D_{pad}, 0, s)}{4sL_p} \delta(x)\delta(y), \quad (2.4)$$

where  $\delta(x)$  and  $\delta(y)$  are unit Dirac Delta functions in x and y directions, respectively.

Table 2.1 Table of pad shape parameters

<i>Kind of pad</i>	$\alpha$
Square Pad connected to multiple nodes of the grid	0.59
Square Pad connected to a single node of the grid	0.2

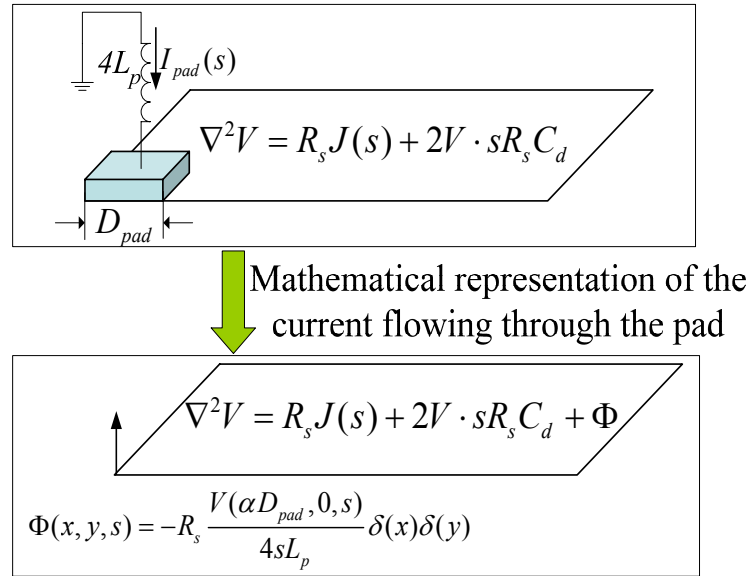


Figure 2.6 Representation of the current flowing through the pad as a Dirac Delta function

Using (2.4) as the source function for (2.1), the equation describing the voltage across a cell becomes

$$\nabla^2 V(x, y, s) = R_s J(s) + 2V(x, y, s) \cdot sR_s C_d - R_s \frac{V(\alpha D_{pad}, 0, s)}{4sL_p} \delta(x)\delta(y). \quad (2.5)$$

Equation (2.5) can be transformed into a pure Hemholtz equation and be solved analytically by putting in the boundary condition of the second kind as described by (2.2) [46]. The detailed derivation of (2.5) is demonstrated in Appendix B. The solution of  $V(x, y, s)$  is

$$V(x, y, s) = - \frac{s \cdot \frac{J(s)}{2C_d} - \frac{J(s) \cdot R_s}{2C_d \cdot 4L_p} [G(x, y, 0, 0, s) - G(\alpha D_{pad}, 0, 0, 0, s)]}{s^2 + s \cdot \frac{R_s}{4L_p} G(\alpha D_{pad}, 0, 0, 0, s)}, \quad (2.6)$$

where  $G(x, y, \xi, \eta, s)$  is the Green's function of a Helmholtz equation with the boundary condition of the second kind. This function can be written as a double series [46].

$$G(x, y, \xi, \eta) = \frac{1}{a^2} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{\varepsilon_n \varepsilon_m \cos(p_n x) \cos(q_m y) \cos(p_n \xi) \cos(q_m \eta)}{p_n^2 + q_m^2 + 2sR_s C_d} \quad (2.7)$$

$$p_n = \frac{\pi n}{a}, \quad q_m = \frac{\pi m}{a}$$

This Green's function gives the interaction between two points  $(x, y)$  and  $(\xi, \eta)$ , which is defined by the PDE.

In (2.6),  $V(x, y, s)$  can determine frequency characteristics of power noise at any location within a cell. Dividing  $V(x, y, s)$  by the total switching current within a cell  $J(s)a^2$ , the transfer impedance of the power distribution network  $Z(x, y, s)$  can be obtained.

$$Z(x, y, s) = - \frac{s \cdot \frac{1}{2C_d a^2} - \frac{R_s}{2C_d a^2 \cdot 4L_p} [G(x, y, 0, 0, s) - G(\alpha D_{pad}, 0, 0, 0, s)]}{s^2 + s \cdot \frac{R_s}{4L_p} G(\alpha D_{pad}, 0, 0, 0, s)}. \quad (2.8)$$

As the current source term is eliminated,  $Z(x,y,s)$  incorporates the intrinsic impedance of a GSI power distribution network.

A comparison between (2.8) and the results of SPICE simulation is shown in Figure 2.7. This analysis is performed for a grid on the top two metal levels of a chip designed at the 65 nm technology node. The chip area is assumed to be  $310 \text{ mm}^2$  according to the ITRS prediction [47]. The total number of power and ground pads is 2048, which accounts for 2/3 of the total pad number [47]. If 2/3 of total chip area is assumed to allocate for power and ground pads, the pad pitch, which is the distance between two power pads or two ground pads, can be calculated as  $375 \text{ }\mu\text{m}$ . The average metal thickness for the top two wiring levels of Intel 65nm generation microprocessors is  $0.77 \text{ }\mu\text{m}$  and the aspect ratio of a signal wire is around 2 [48]. If three signal wires are sandwiched between a pair of power and ground wires, the width of a power/ground wire should be 3 times the width of a signal wire [49]. Also, the average signal wire pitch for the top two wiring levels of a 65 nm generation Intel microprocessor is  $0.9 \text{ }\mu\text{m}$  [48]. As a result, the power/ground wire pitch, which is the distance between two power wires or two ground wires, can be calculated as  $8.74 \text{ }\mu\text{m}$ . The fineness, namely, the number of power (ground) wires between two power (ground) pads, can be calculated as 43 by using the pad pitch and power/ground wire pitch (we can call this grid a 43x43 grid). The segment resistance  $R_s$  can be calculated as  $0.22 \text{ }\Omega$ . The pad size is set as 1/6-1/5 of the pad pitch in [28], and here a pad is assumed to connect to 49 nodes on the grid (pad size is 1/6 pad pitch). In this analysis, 10% of the chip area is allocated as decaps [50]. The value of the Effective Oxide Thickness (EOT) is taken to be  $1.2 \text{ nm}$  according to the ITRS projection for the 65 nm node[47], and the EOT value shows the thickness of  $\text{SiO}_2$

gate oxide needed to obtain the same gate capacitance as the one obtained with other types of gate oxide materials. Typically, on-chip decaps are made of the gate oxide in CMOS technology [11], and the EOT value is used to calculate the decoupling capacitance. The typical value of the package inductance per I/O is less than 1 nH for flip-chip technology [51]. In this calculation, it is assumed a pair of power/ground I/Os have 1 nH inductance.

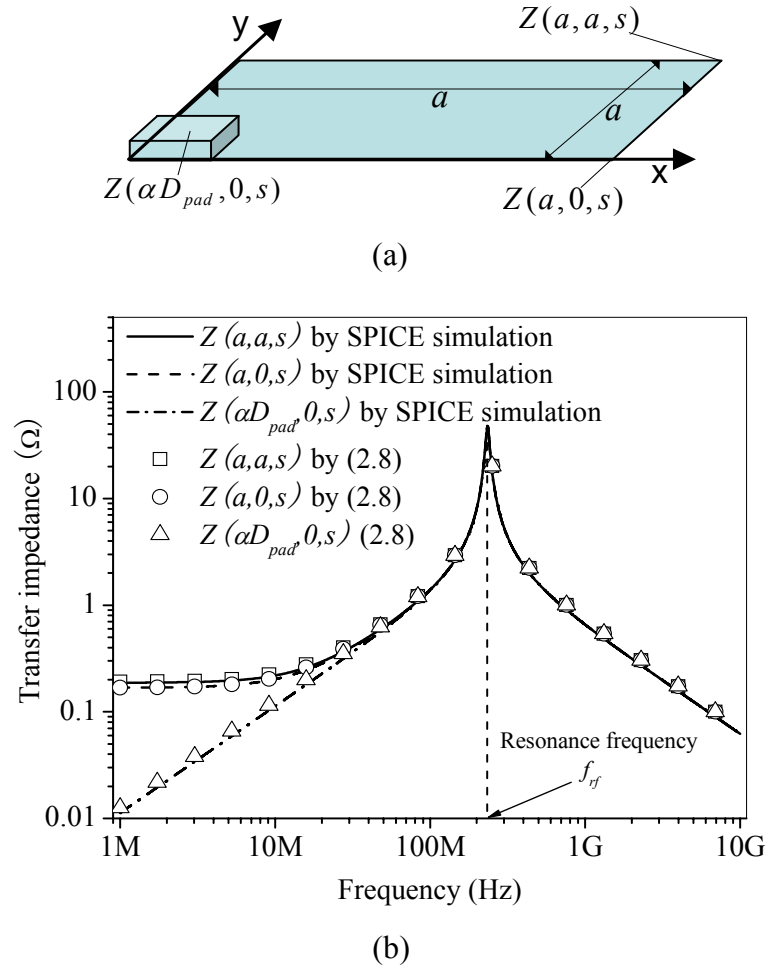


Figure 2.7 Transfer impedance for three corner points; (a) Locations of three corner points; (b) Comparison between (2.8) and the results of SPICE simulations

The transfer impedances for three typical corner points are shown in Figure 2.7, and it is noted that the transfer impedance has a low-pass characteristic with only one peak resonance frequency. The difference in DC values between these three points is due to different IR-drop values at various locations. The comparison in Figure 2.7 shows that (2.8) maintains high accuracy throughout the entire frequency range with less than 4% error.

## 2.4 Simplified Transfer Impedance Function

Observing (2.7) and (2.8), it is noted that the Green's function is an infinite series, and thus both the numerator and denominator have higher order terms of  $s$ . As a result, the calculation of noise transients in the time domain involves sophisticated numerical solutions. To be able to solve for noise transients analytically, a simpler transfer impedance function is needed. It can be observed from Figure 2.7 that (2.8) has low-pass characteristics with only one peak resonance frequency. This enables a simplified second-order transfer impedance function  $Z_s(x,y,s)$  to approximate (2.8).

$$Z_s(x,y,s) = -\frac{s \cdot \frac{1}{2C_d a^2} + \frac{Z(x,y,0)}{2C_d a^2 \cdot 4L_p}}{s^2 + s \cdot k \cdot \frac{Z(x,y,0)}{4L_p} + \frac{1}{2C_d a^2 \cdot 4L_p}}. \quad (2.9)$$

In the DC case,  $s=0$ , and

$$Z_s(x,y,0) = Z(x,y,0) = R_{IR}(x,y,0). \quad (2.10)$$

Transfer impedance  $Z_s(x,y,0)$  can be calculated from (2.8), and determines the effective IR-drop resistance for point  $(x,y)$ . Coefficient  $k$  is the other unknown quantity which can be calculated from (2.8) as follows. The resonance frequency  $f_{rf}$  for the power distribution

network, as shown in Figure 2.7, is determined by the total amount of decaps within a cell and the total amount of package inductance connected to a cell,

$$f_{rf} = \frac{1}{2\pi\sqrt{2C_d a^2 \cdot 4L_p}}. \quad (2.11)$$

If we let

$$Z_s(x, y, j2\pi f_{rf}) = Z(x, y, j2\pi f_{rf}), \quad (2.12)$$

The value of  $k$  can be obtained as

$$k = \frac{\sqrt{\left(\frac{4L_p}{2C_d a^2}\right)^2 \left(\frac{1}{Z(x, y, 0)}\right)^2 + \left(\frac{4L_p}{2C_d a^2}\right)^2}}{|Z(x, y, j2\pi f_{rf})|}. \quad (2.13)$$

Table 2.2 Range of main variable values

<i>Variables</i>	<i>min</i>	<i>max</i>	<i>unit</i>
$R_s$	0.01	2	$\Omega$
Decaps insertion	0	40% of chip area	----
$L_p$	0.01	2	nH
Pad area	Single node	30% of chip surface	----
Grid fineness	5	$\infty$	----

The intuition behind the derivation for (2.9) is that to get the simplified transfer impedance function, two particular points need to be calculated in the frequency domain from (2.8). This greatly reduces computational complexity and enables analytical derivation of noise transients in the time domain. The accuracy of (2.9) is verified in Figure 2.8. Simulation parameters are the same as those used in Section 2.3. There is almost no difference between (2.8) and (2.9), and (2.9) also has less than 4% error compared to the results of SPICE simulations. Both (2.8) and (2.9) maintain less than 4%

error in a large range when the main physical variables change. The range of the main physical variables is shown in Table 2.2. To ensure the models in this chapter can cover a large design space, the maximum value of each variable is at least 4 times the empirical values which are chosen in the simulations.

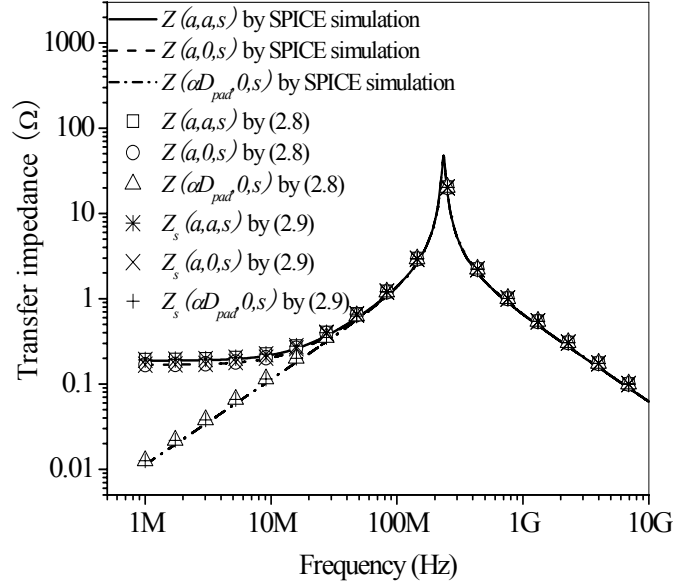


Figure 2.8 Transfer impedance for three corner points: Comparison between (2.8), (2.9) and the results of SPICE simulations

## 2.5 Analytical Solution for Noise Transients

As discussed in Section 2.1, the current waveform induced by a function block is approximated by a ramp function.

$$i(t) = \frac{I_p}{t_r} [t \cdot u(t) - (t - t_r)u(t - t_r)]. \quad (2.14)$$

$I_p$  represents the peak current and  $t_r$  is the rise time of the ramp. The Laplace transform of (2.14) is

$$I(s) = \frac{I_p}{t_r} \left( \frac{1}{s^2} - \frac{e^{-t_r \cdot s}}{s^2} \right). \quad (2.15)$$

Equation (2.9) can be also rewritten as

$$Z_s(s) = \frac{K_1 s + K_0}{s^2 + 2Bs + \omega_{rf}^2} = K_1 \cdot \frac{s + \frac{K_0}{K_1}}{(s + B)^2 + (\omega_{rf}^2 - B^2)}, \quad (2.16)$$

where

$$K_1 \equiv -\frac{1}{2C_d a^2}, \quad K_0 \equiv -\frac{Z(x, y, 0)}{2C_d a^2 \cdot 4L_p}, \quad B \equiv \frac{k}{2} \cdot \frac{Z(x, y, 0)}{4L_p}, \quad \omega_{rf} \equiv \sqrt{\frac{1}{2C_d a^2 \cdot 4L_p}}. \quad (2.17)$$

Using (2.15) and (2.16),

$$V(s) = I(s) \cdot Z_s(s) = K_1 \frac{I_p}{t_r} \cdot \frac{s + \frac{K_0}{K_1}}{s^2 \cdot [(s + B)^2 + (\omega_{rf}^2 - B^2)]} \cdot (1 - e^{-t_r \cdot s}). \quad (2.18)$$

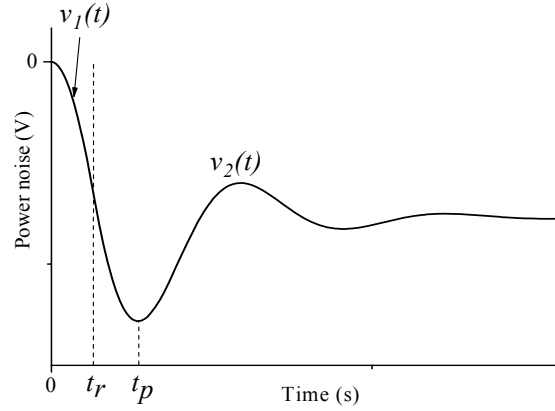


Figure 2.9 Two parts for the power noise waveform

The inverse Laplace transform of (2.18) represents the time domain response of power supply noise, and the transients can be divided into two parts, as shown in Figure 2.9.

From  $t=0^+$  to  $t=t_r$ , the power noise transients can be written as  $v_1(t)$ ,



$$v_1(t) = \frac{I_p}{t_r \cdot \omega_{rf}^2} \left( K_1 - \frac{2 \cdot K_0 \cdot B}{\omega_{rf}^2} \right) + \frac{I_p \cdot K_1 \sqrt{\left( \frac{K_0}{K_1} - B \right)^2 + \omega_{rf}^2 - B^2}}{t_r \cdot \omega_{rf}^2 \sqrt{\omega_{rf}^2 - B^2}} \cdot e^{-Bt} \cdot \sin \left( \frac{\sqrt{\omega_{rf}^2 - B^2}}{+ \phi} \cdot t \right), \quad (2.19)$$

where

$$\phi = \tan^{-1} \left( \frac{\sqrt{\omega_{rf}^2 - B^2}}{\frac{K_0}{K_1} - B} \right) + 2 \tan^{-1} \left( \frac{\sqrt{\omega_{rf}^2 - B^2}}{B} \right). \quad (2.20)$$

(2.19) is composed of a linear term and a sinusoidal term with exponential decay.

When  $t > t_r$ , the power supply noise transients can be written as  $v_2(t)$

$$v_2(t) = -I_p \cdot Z(x, y, 0) + \frac{I_p \cdot K_1 \sqrt{\left( \frac{K_0}{K_1} - B \right)^2 + \omega_{rf}^2 - B^2}}{t_r \cdot \omega_{rf}^2 \sqrt{\omega_{rf}^2 - B^2}}, \quad (2.21)$$

$$\times \sqrt{1 - 2 \cdot e^{Bt_r} \cos(\sqrt{\omega_{rf}^2 - B^2} \cdot t_r)} + e^{2Bt_r} \cdot e^{-Bt} \cdot \sin \left( \sqrt{\omega_{rf}^2 - B^2} \cdot t + \phi + \phi_0 \right)$$

where

$$\phi_0 = \tan^{-1} \left( \frac{e^{Bt_r} \cdot \sin(\sqrt{\omega_{rf}^2 - B^2} \cdot t_r)}{1 - e^{Bt_r} \cdot \cos(\sqrt{\omega_{rf}^2 - B^2} \cdot t_r)} \right). \quad (2.22)$$

The first term in (2.21) is a constant DC value, which denotes the steady state IR-drop.

The second term is a sinusoidal function with exponential decay.

As a result, the noise transients  $v(t)$  can be written as the sum of  $v_1(t)$  and  $v_2(t)$ .

$$v(t) = v_1(t) \cdot [u(t) - u(t - t_r)] + v_2(t) \cdot u(t - t_r). \quad (2.23)$$

Figure 2.10 illustrates the power noise transients of three corner points for the same grid as mentioned in previous Sections. The allowable maximum current for a high performance MPU implemented at the 65 nm technology node is 172 A [47]. If we divide this current by the total chip area, the applied peak current density in this analysis

can be calculated as  $0.55 \text{ A/mm}^2$ . The rise time of the current is calculated as 5 times the on-chip clock cycle [52], which is 1 ns for a 5 GHz chip. Equation (2.23) matches the results of SPICE simulations well and has less than 5% error for the entire range of variables shown in Table 2.2.

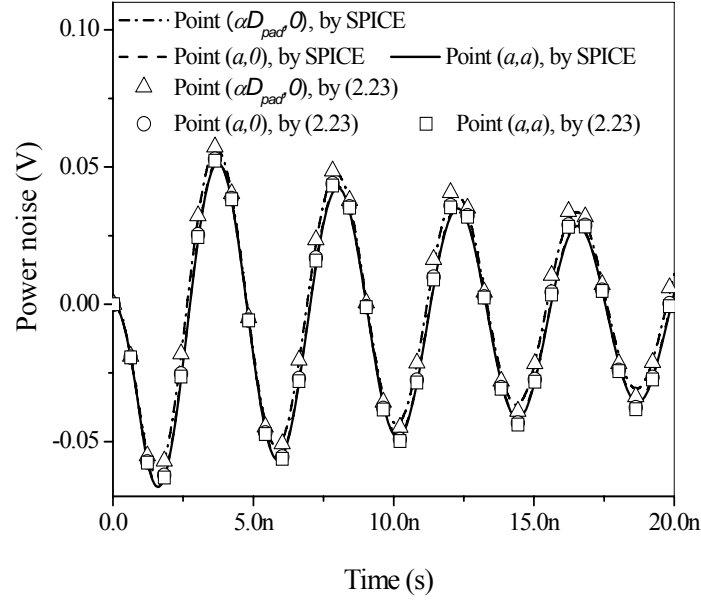


Figure 2.10 Power supply noise waveforms for three corner points: Comparison between (2.23) and the results of SPICE simulations

## 2.6 Analytical solution of peak noise

In previous sections, a single power or ground grid is modeled. The total transfer impedance of the whole power distribution network is equal to the sum of the transfer impedances of power and ground grids [7]. If the two grids have symmetrical structures, as shown in Figure 2.5, the total impedance can be calculated as

$$Z_{total}(x, y, s) = Z_s(x, y, s) + Z_s(a - x, a - y, s). \quad (2.24)$$

Similarly, the total noise  $v_{total}(x,y,t)$  is equal to the sum of the noise produced by power and ground grids.

$$v_{total}(x,y,t) = v(x,y,t) + v(a-x,a-y,t). \quad (2.25)$$

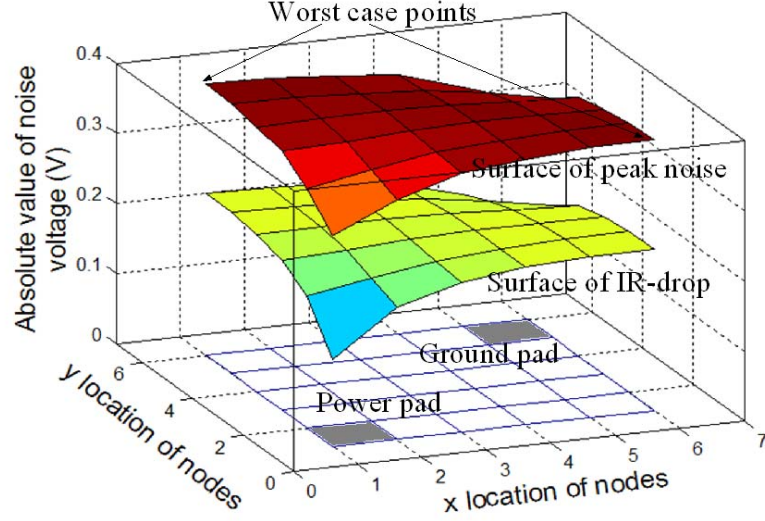


Figure 2.11 Noise values vs. location of nodes for IR-drop and peak noise.  $L_p=0.5$  nH, 10% chip area is occupied by decaps, and  $R_s=1.1 \Omega$

To identify the points with the worst case noise, SPICE simulation is performed on a pair of 11x11 power and ground grids simultaneously by using the simplified circuit model, as shown in Figure 2.4. Figure 2.11 gives a spatial view of the IR-drop value and peak noise value for each node. The minimum noise always occurs at the corner points  $(0,0)$  and  $(a,a)$ , which is where pads are located, and the worst case noise occurs at the two remaining corner points, or  $(0,a)$  and  $(a,0)$ . For a single grid network of two metal levels, the peak noise occurs when the sinusoidal function in (2.19) reaches its first peak value. The time at which this occurs, or the peak time  $t_p$ , can be solved by

$$\sin\left(\sqrt{\omega_{rf}^2 - B^2} \cdot t_p + \phi + \phi_0\right) = 1 \Rightarrow t_p = \frac{\frac{5}{2}\pi - \phi - \phi_0}{\sqrt{\omega_{rf}^2 - B^2}}. \quad (2.26)$$

Consequently putting this peak time  $t_p$  into (2.21) the peak noise value of a single grid network is

$$V_{peak} = -I_p \cdot Z(x, y, 0) + \frac{I_p \cdot K_1 \sqrt{\left(\frac{K_0}{K_1} - B\right)^2 + \omega_{rf}^2 - B^2}}{t_r \cdot \omega_{rf}^2 \sqrt{\omega_{rf}^2 - B^2}} \cdot \sqrt{1 - 2 \cdot e^{Bt_r} \cdot \cos(\sqrt{\omega_{rf}^2 - B^2} \cdot t_r) + e^{2Bt_r} \cdot e^{-Bt_p}}. \quad (2.27)$$

The total worst case noise always occurs at points  $(a, 0)$  and  $(0, a)$ . The total noise at  $(a, 0)$  can be written as

$$v_{total}(a, 0, t) = v(a, 0, t) + v(0, a, t) = 2 \cdot v(a, 0, t), \quad (2.28)$$

and the worst case peak noise of the double grid network becomes equal to

$$V_{total-worstcase-peak} = 2 \cdot V_{peak}(a, 0). \quad (2.29)$$

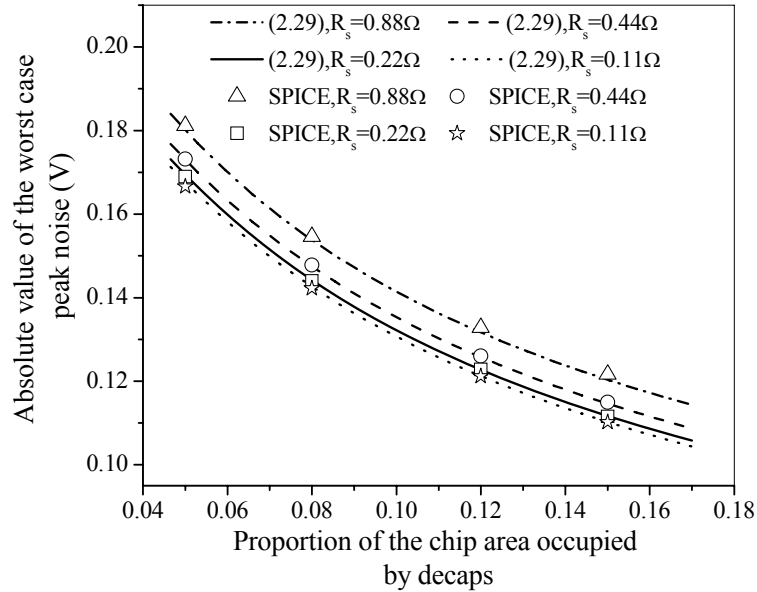


Figure 2.12 The worst case peak noise as a function of the chip area occupied by decaps: Comparison between (2.29) and the results of SPICE simulations for a pair of grids

SPICE simulations are performed on a pair of 43x43 power and ground grids, and each grid is similar to the grid used in previous sections except for changing some of the physical parameters. These comparisons are illustrated in Figures 2.12-2.16 using the same switching current density of 0.55 A/mm<sup>2</sup>.

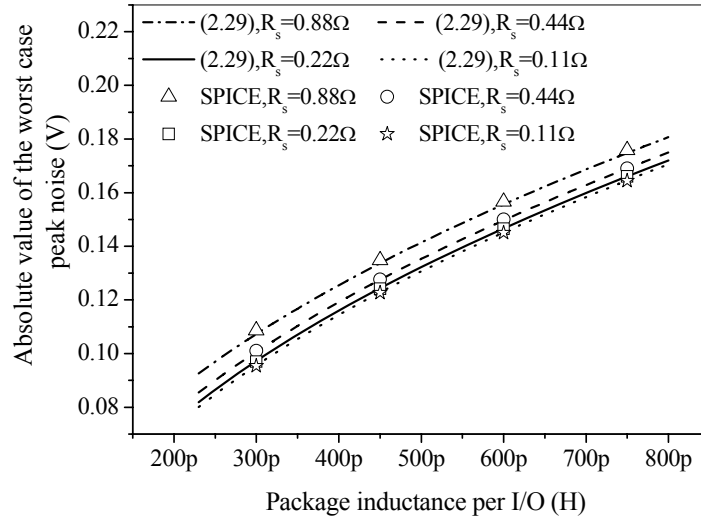


Figure 2.13 The worst case peak noise as a function of  $L_p$ : Comparison between (2.29) and the results of SPICE simulations for a pair of grids

It is observed from Figures 2.12, 2.13, and 2.16 that  $\Delta I$  noise is sensitive to the amount of decaps, package level inductance, and the number of I/O pads. Decap insertion is an effective way to reduce the noise level. However, the on-die area budget for decoupling capacitors can be limited. In this situation, package-level high density I/O solutions, such as sea of leads (SoL) [53], can be used to suppress the power supply noise. High density chip I/Os can greatly reduce the loop inductance of power distribution networks, resulting in smaller noise. Larger numbers of I/Os can also reduce the IR-drop.

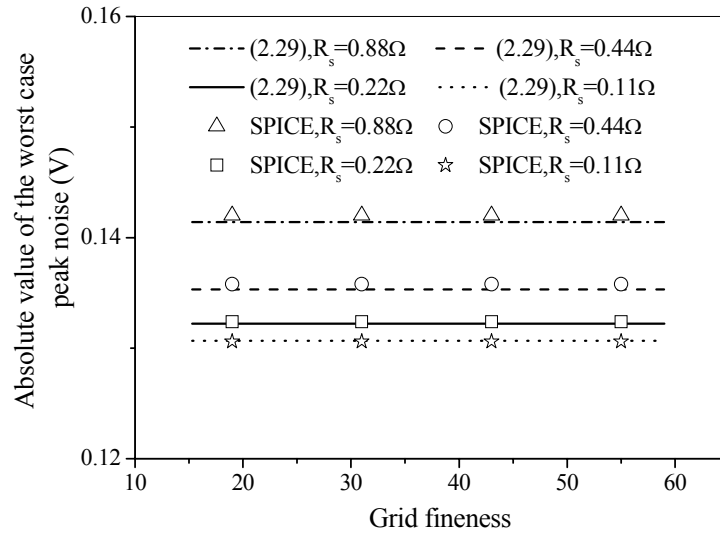


Figure 2.14 The worst case peak noise as a function of grid fineness (the number of power (ground) wires between two power (ground) pads) : Comparison between (2.29) and the results of SPICE simulations for a pair of grids

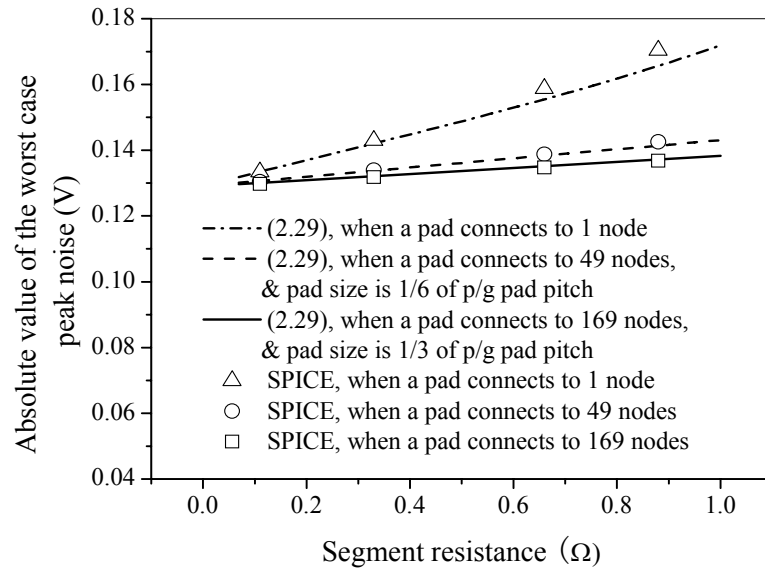


Figure 2.15 The worst case peak noise as a function of segment resistance for different pad sizes: Comparison between (2.29) and the results of SPICE simulations

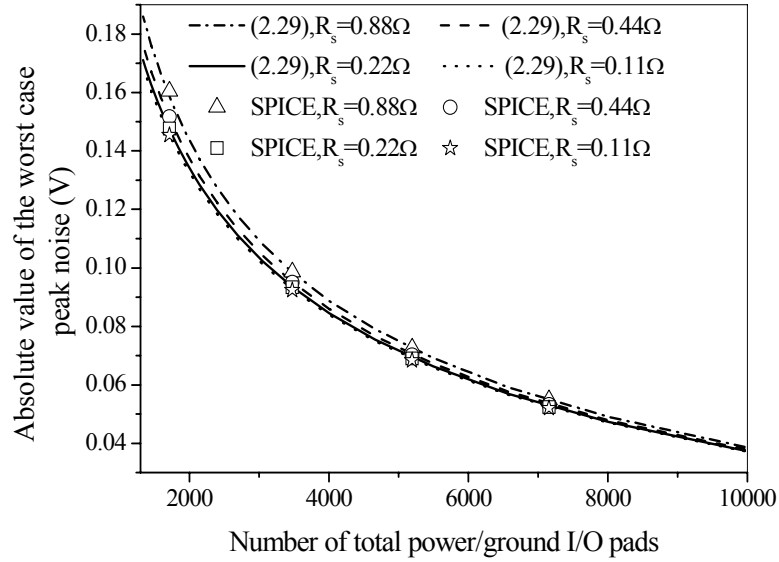


Figure 2.16 The worst case peak noise as a function of the number of pads: Comparison between (2.29) and the results of SPICE simulations

Grid fineness does not influence the worst case peak noise value if the pad size keeps constant, as shown in Figure 2.14. Increasing the pad size can also help reduce the worst case peak noise, but is not as efficient as increasing the amount of decaps and decreasing the I/O number, as shown in Figure 2.15. From the above plots, it is clear that these compact physical models can be used to gain physical insight into the tradeoffs between chip and package level resources.

## 2.7 Technology trends of power supply noise

The models can also be used to project the power noise trends for different generations of technology. In this section, the worst case peak noise value is calculated for a high performance microprocessor unit (MPU) for each generation from the 65 nm node (year 2007) to the 18 nm node (year 2018) [47]. In analyzing the technology trends for the power supply noise (the data of each physical parameter at the 65 nm node has

already been shown in previous sections). The values and scaling factors of each parameter for future generations are obtained as follows

- The analysis is performed for a grid made of the top two metal levels.
- The total number of power/ground pads, chip area, supply voltage, power dissipation, on-chip clock frequency, and the equivalent oxide thickness (EOT) are selected based on the ITRS projections [47].
- For Intel microprocessors at the 180 nm, 130 nm, 90 nm, and 65 nm nodes [54, 55, 56, 48], metal thickness and signal wire pitch for the top two wiring levels do not scale with technology. The numbers for the 65 nm node are taken for each technology generation.
- Reducing the package level inductance is associated with high costs. Therefore, we assume a constant  $L_p$  (0.5 nH) as a safe assumption [51, 57].

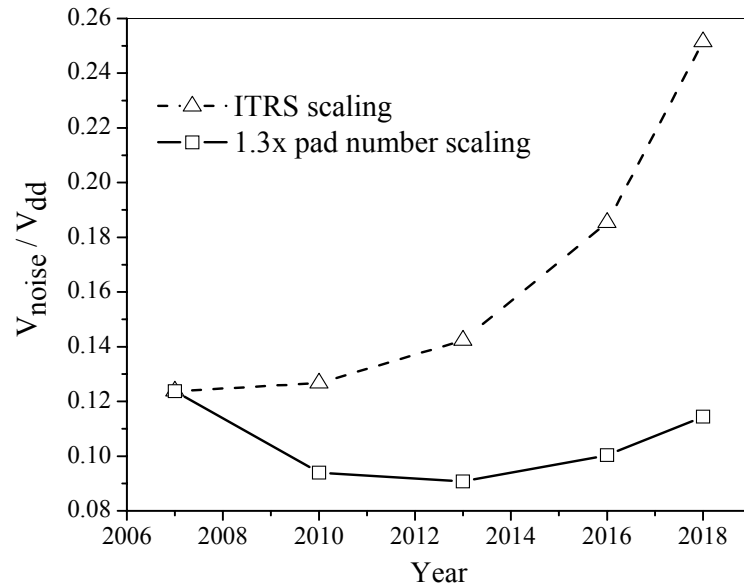


Figure 2.17 Technology trends of the worst case peak noise



Figure 2.17 suggests that supply noise could reach 25%  $V_{dd}$  at the 18 nm node compared to 12%  $V_{dd}$  for current technologies if the ITRS scaling trends are followed. Excessive noise can cause severe difficulties for circuit designers, and new solutions to tackle this supply noise problem are needed in the future.

The importance of scaling package parameters such as the number of I/O pads is also indicated in Figure 2.17. It can be seen that by increasing the pad number by 1.3x each generation, the supply noise can be kept well under control.

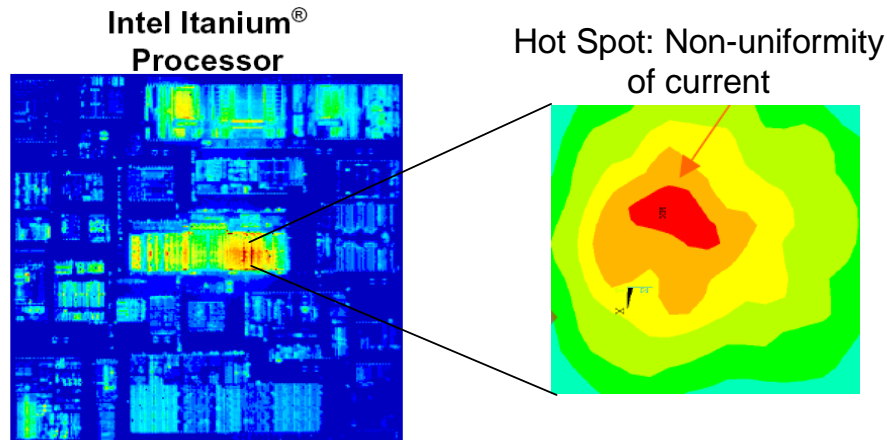
## 2.8 Conclusion

In this chapter, a set of compact physical models is derived to describe the frequency characteristics, time domain transients and the worst case peak noise value for the first droop power supply noise of GSI power distribution networks. The models incorporate the distributed nature of on-chip grids and display high accuracy. The models enable quick full waveform recognition of the power supply noise without doing dedicated simulations. Designers can also perform chip/package co-design of power distribution networks and tradeoff multiple design considerations such as wiring resource allocation, decap insertion and pad allocation. The models allow designers to explore a large design space, supporting the power grid analysis not only for the state-of-the-art design but also for the scaling trends of future technology nodes. There is less than 4% difference between the worst case peak noise model and the results of SPICE simulations.

## CHAPTER 3: COMPACT PHYSICAL MODELS FOR POWER SUPPLY NOISE WITH THE CONSIDERATION OF HOT SPOTS

### 3.1 Introduction

The increasing functional complexities of microprocessors result in a non-uniform power (current) density distribution across the die. Local power densities with greater than  $300 \text{ W/cm}^2$ , also referred as hot spots, are not rare for today's high-performance chips [34]. These hot spots not only require advanced thermal solutions, but also challenge the design of power distribution systems. The non-uniformity of the current density distribution also leads to high noise levels at current crowding locations.



Source: [www.intel.com](http://www.intel.com)

Figure 3.1 Current map of Intel Itanium® Processor [2]

In Chapter 2, blockwise models are proposed to address the power integrity problem of power hungry blocks. The models are based on an assumption that the switching current and decap distributions are uniform throughout the functional block. Figure 3.1 shows the current map of Intel Itanium® [2]. It can be observed that for most

functional blocks, the current distribution is uniform, and the blockwise models can be well applied. However, the current induced by some critical blocks at the center of the chip displays high non-uniformity, and the blockwise models can not be used for these blocks. Thus, in this chapter, the blockwise models are further extended to more general cases by removing the assumption of uniform switching current conditions [58]. The new generalized analytical physical models enable quick recognition of the first droop noise for arbitrary functional block sizes and non-uniform current switching conditions.

### 3.2 Mathematical Modeling for Hot Spots

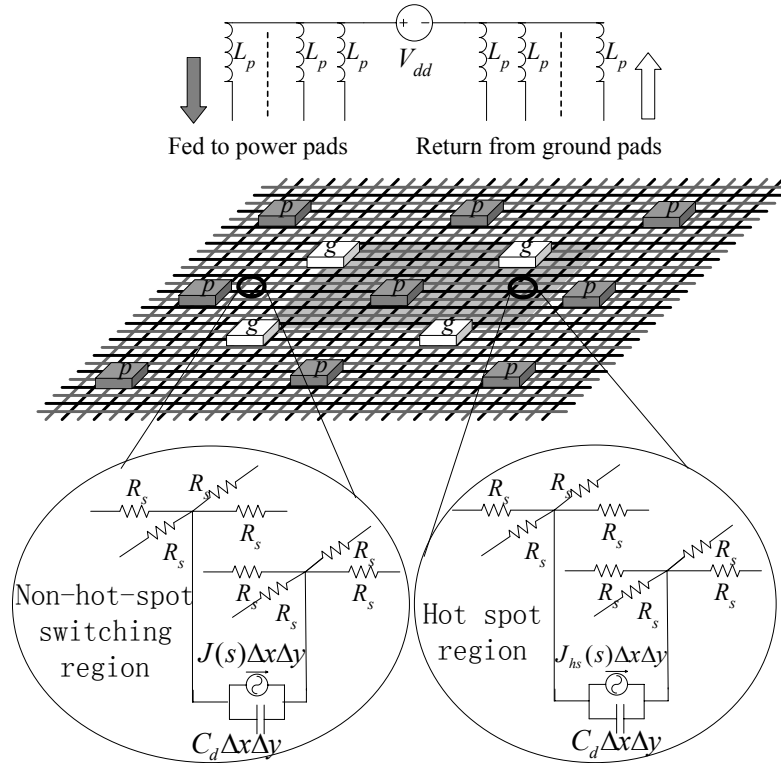


Figure 3.2 Simplified circuit model for GSI power distribution system with a hot spot

A simplified circuit model that accounts for the hot spot is shown in Figure 3.2 and is an extension of Figure 2.4. Symbols  $R_s$ ,  $C_d$ ,  $\Delta x$ ,  $\Delta y$ , and  $L_p$  are the same as those

used in Section 2.3. The current density for an active block, as shown in the non-shaded region, is represented by  $J(s)$  in the Laplace domain. In a high-performance chip, high local power dissipation can result in hot spots, as shown in the shaded region in Figure 3.2, where  $J_{hs}(s)$  denotes the current density inside hot spot.

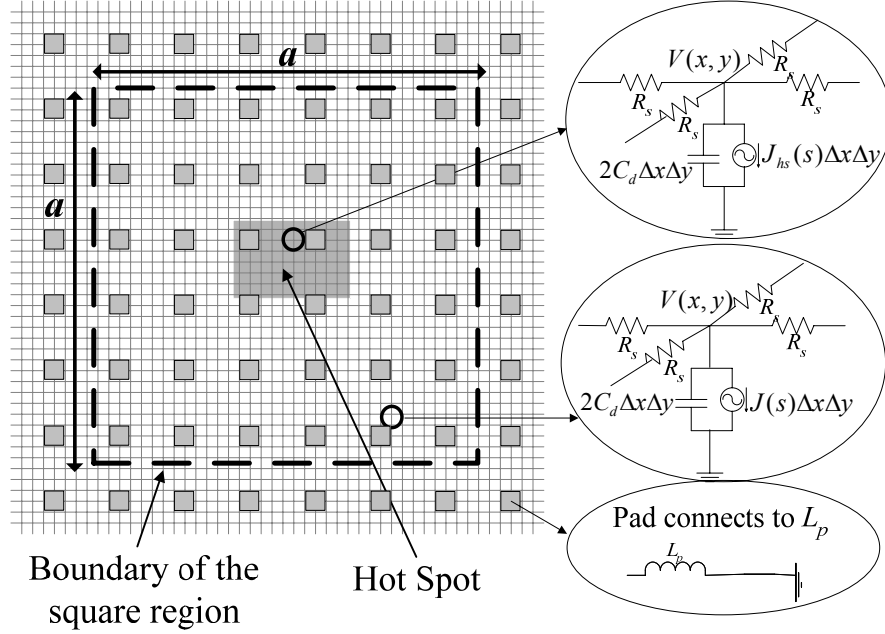


Figure 3.3 Circuit model for single grid structure and the square region allocated for the analysis

The on-chip power distribution system consists of power and ground grids, and this double grid structure can be decoupled into two individual grids, as shown in Figure 3.3. The main objective of this work is to accurately model the power supply noise caused by hot spots. We consider one hot spot in this section, but the results can be extended to more hot spots by superposition. Hot spots are typically small compared to the chip area, and we can always consider a large square region that contains all the power/ground pads with considerable contributions to the supply current of a hot spot.

Partial differential equation (3.1) can describe the frequency characteristics of the power noise  $V(x,y,s)$  at each node in this square region.

$$\nabla^2 V(x,y,s) = R_s J(s) + 2V(x,y,s) \cdot sR_s C_d + \Phi(x,y,s), \quad (3.1)$$

where  $\Phi(x,y,s)$  is the source function of this equation and can be written as

$$\begin{aligned} \Phi(x,y,s) = & -\sum_{i=1}^M R_s [J_{hs}(s) - J(s)] \cdot \Delta x \cdot \Delta y \cdot \delta(x - x_{spi}) \delta(y - y_{spi}) \\ & - \frac{R_s}{sL_p} \sum_{j=1}^N V_{padj}(s) \cdot \delta(x - x_{padj}) \delta(y - y_{padj}) \end{aligned} \quad (3.2)$$

The first term of  $\Phi(x,y,s)$  represents the current sources associated with switching nodes in the hot spot region.  $M$  is the total number of nodes within the hot spot, and  $(x_{spi}, y_{spi})$  represents the location of each node inside the hot spot. The second term represents the voltage drop on each package inductance ( $L_p$ ) associated with I/O pads.  $N$  is the total number of pads.  $V_{padj}(s)$  and  $(x_{padj}, y_{padj})$  denote the voltage and location of each pad, respectively.

By choosing a region large enough for the analysis, there would be virtually no current flowing through the boundaries, thereby producing boundary conditions of the second kind for (3.1) [46],

$$\frac{\partial V}{\partial y} \Big|_{x=0} = 0, \quad \frac{\partial V}{\partial y} \Big|_{x=a} = 0, \quad \frac{\partial V}{\partial x} \Big|_{y=0} = 0, \quad \frac{\partial V}{\partial x} \Big|_{y=a} = 0, \quad (3.3)$$

where  $a$  denotes the size of the square region chosen for analysis.

Equation (3.3) is the combination of a Poisson's equation and a Helmholtz equation. If we let

$$V(x,y,s) = u(x,y,s) - \frac{J(s)}{2sC_d}, \quad (3.4)$$

then (3.1) is modified into a pure Helmholtz equation

$$\nabla^2 u(x, y, s) = 2u(x, y, s) \cdot sR_s C_d + \Phi(x, y, s), \quad (3.5)$$

which also satisfies boundary condition of the second kind.

$$\frac{\partial u}{\partial y} \Big|_{x=0} = 0, \quad \frac{\partial u}{\partial y} \Big|_{x=a} = 0, \quad \frac{\partial u}{\partial x} \Big|_{y=0} = 0, \quad \frac{\partial u}{\partial x} \Big|_{y=a} = 0. \quad (3.6)$$

The source function can be rewritten into

$$\begin{aligned} \Phi_u(x, y, s) = & - \sum_{i=1}^M R_s [J_{hs}(s) - J(s)] \cdot \Delta x \cdot \Delta y \cdot \delta(x - x_{spi}) \delta(y - y_{spi}) \\ & - \frac{R_s}{sL_p} \sum_{j=1}^N \left[ u_{padj}(s) - \frac{J(s)}{2sC_d} \right] \cdot \delta(x - x_{padj}) \delta(y - y_{padj}). \end{aligned} \quad (3.7)$$

The solution of (3.5) can be obtained by using Green's function  $G(x, y, \xi, \eta, s)$ . The solution is

$$\begin{aligned} u(x, y, s) = & \sum_{i=1}^M R_s [J_{hs}(s) - J(s)] \cdot \Delta x \cdot \Delta y \cdot G(x, y, x_{spi}, y_{spi}, s) \\ & - \frac{R_s}{sL_p} \sum_{j=1}^N \left[ u_{padj}(s) - \frac{J(s)}{2sC_d} \right] \cdot G(x, y, x_{padj}, y_{padj}, s). \end{aligned} \quad (3.8)$$

However, in (3.8)  $u_{padk}(s)$  ( $k=1..N$ ) is still an unknown for each pad, but it is known that  $u_{padk}(s)$  should also satisfy (3.1) and therefore (3.8). If we substitute  $u_{padk}(s)$  ( $k=1..N$ ) back into (3.8), we have (3.9).

Equation set (3.9) includes  $N$  equations and  $N$  unknowns. The voltage  $u_{padk}(s)$  ( $k=1..N$ ) associated with each pad can be solved from (3.9), and  $u(x, y, s)$  can also be calculated accordingly. Consequently, the frequency characteristics of the power supply noise for a single power/ground grid  $V(x, y, s)$  with a hot spot can be solved analytically from (3.4).

$$\begin{aligned}
u_{pad1}(s) &= \sum_{i=1}^M R_s [J_{hs}(s) - J(s)] \cdot \Delta x \cdot \Delta y \cdot G(x_{pad1}, y_{pad1}, x_{spi}, y_{spi}, s) \\
&\quad - \frac{R_s}{sL_p} \sum_{j=1}^N \left[ u_{padj}(s) - \frac{J(s)}{2sC_d} \right] \cdot G(x_{pad1}, y_{pad1}, x_{padj}, y_{padj}, s) \\
\hdashline
u_{padk}(s) &= \sum_{i=1}^M R_s [J_{hs}(s) - J(s)] \cdot \Delta x \cdot \Delta y \cdot G(x_{padk}, y_{padk}, x_{spi}, y_{spi}, s) \\
&\quad - \frac{R_s}{sL_p} \sum_{j=1}^N \left[ u_{padj}(s) - \frac{J(s)}{2sC_d} \right] \cdot G(x_{padk}, y_{padk}, x_{padj}, y_{padj}, s) \\
\hdashline
u_{padN}(s) &= \sum_{i=1}^M R_s [J_{hs}(s) - J(s)] \cdot \Delta x \cdot \Delta y \cdot G(x_{padN}, y_{padN}, x_{spi}, y_{spi}, s) \\
&\quad - \frac{R_s}{sL_p} \sum_{j=1}^N \left[ u_{padj}(s) - \frac{J(s)}{2sC_d} \right] \cdot G(x_{padN}, y_{padN}, x_{padj}, y_{padj}, s)
\end{aligned} \tag{3.9}$$

The time domain transient noise can be obtained by performing an inverse Laplace transform on  $V(x,y,s)$ . The peak noise for each node can also be identified by adding up the transient noises of power and ground grids.

### 3.3 Case study

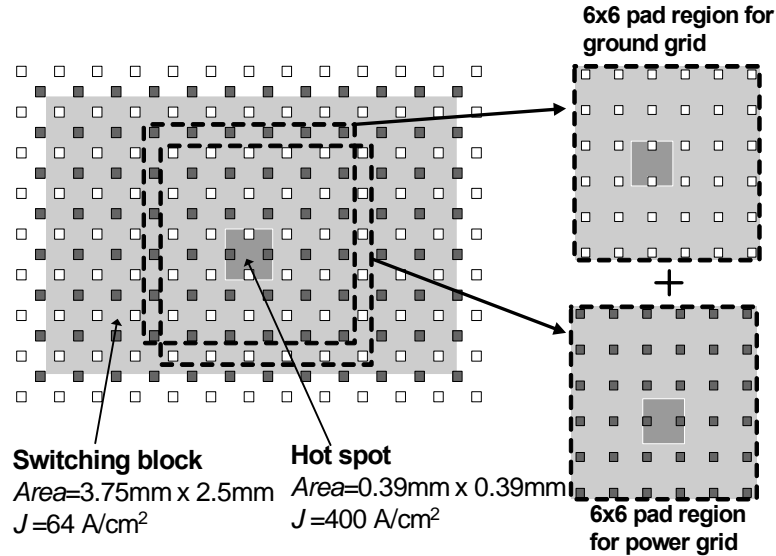


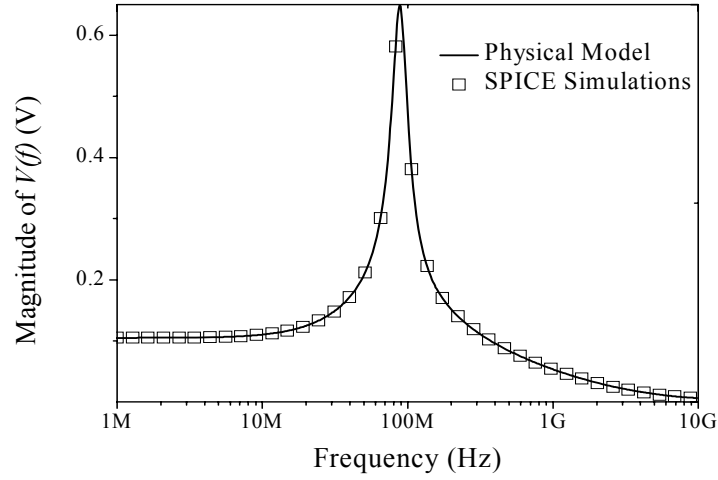
Figure 3.4 Illustration of the switching block, the hot spot and the 6x6 pad regions allocated for the analysis

A case study is performed for a functional block with grids on the top two metal levels of a chip designed at the 45 nm node. The functional block contains a large number of pads (over 100 power and ground pads) and has a uniform current density distribution except for a hot spot region with an extremely high current density (such as the interface circuitry for data cache unit in the Power5 [42]), as shown in Figure 3.4. In this analysis, the switching functional block occupies 3.75 mm x 2.5 mm chip area and has an on-current density of 64 A/cm<sup>2</sup>, which is the average current density given by the ITRS [47] for the 45 nm node. The hot-spot region is assumed to have an on-current density of 400 A/cm<sup>2</sup>. This hot spot occupies a 0.39 mm x 0.39 mm region located at the center of the switching block. According to the ITRS projections for chip area and number of I/Os at the 45 nm node, the distance between two power (ground) pads can be calculated as 375  $\mu$ m. As the top level interconnects do not often scale with technology, the wire dimensions for the Intel's 65 nm microprocessors are assumed to be the same as those of the 45 nm chips [48]. The segment resistance  $R_s$  is calculated as 0.22  $\Omega$ , and the number of power (ground) wires between two power (ground) pads is calculated as 43. In this calculation, the empirical value of 0.5 nH is also used as the package inductance associated with each power/ground I/O [51], and it is assumed that 20% of the chip area is allocated for decaps. The value of EOT for decaps is taken to be 0.65nm according to the ITRS projections for the 45 nm node [47].

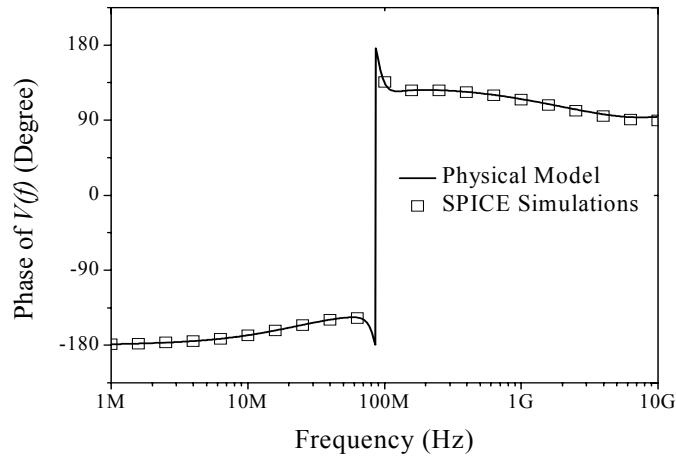
As shown in Figure 3.4, the double grid structure needs to be divided into two single grids as previously noted. To apply the new model, a 6x6 pad region around the hot spot is selected for each grid. It is found that less than 1% of the total supply current consumed by the hot spot region flows through the pads outside the region, and thus a



6x6 pad region is sufficient for the analysis. Figure 3.5 illustrates the frequency domain noise responses (magnitude and phase responses) at the center point of the hot spot. The results are also compared against SPICE simulation results, and the new model shows less than 1% error.



(a)



(b)

Figure 3.5 Frequency domain noise response for the center point of the hot spot. (a) Magnitude response, (b) Phase response

The total transient noise voltage at the center point of the hot spot can also be obtained using inverse Laplace transform and is represented by the solid line shown in Figure 3.6. Compared with the results of SPICE simulation (square symbols), the peak noise value has less than 1% error.

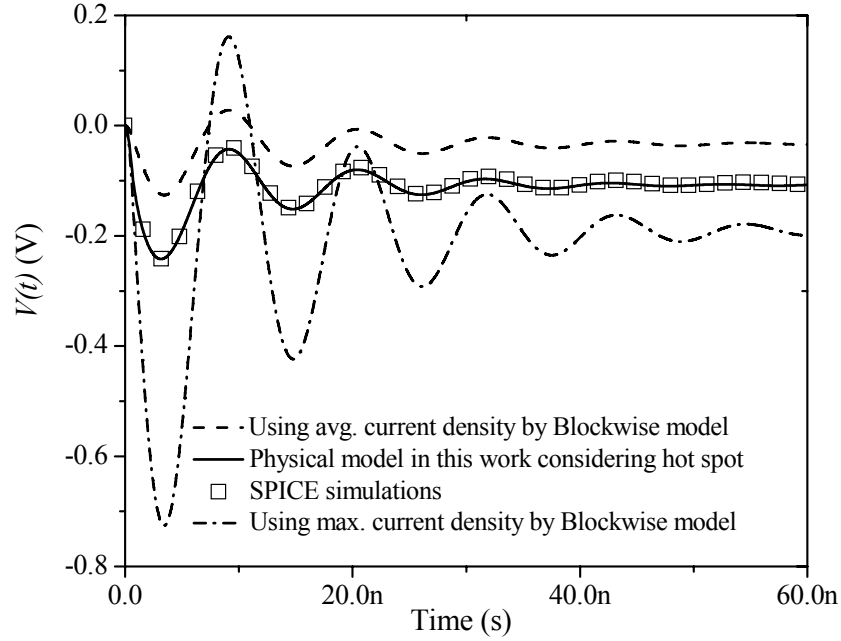


Figure 3.6 Transient noise waveforms using SPICE simulations and different models

To further understand the significance of this modified model in this case, it is necessary to look at the error from the blockwise model that ignores the non-uniform switching current caused by the hot spot. The average current density for the functional block is approximately  $70 \text{ A/cm}^2$ . By using this average current, the transient noise response based on the blockwise models proposed in Chapter 2 is shown by the dashed line in Figure 3.6. The dash-dotted line is the noise response when the maximum current density within the functional block ( $400 \text{ A/cm}^2$ ) is applied in the blockwise model. It is

noted that if the non-uniformity of the current is neglected and the average current density is used instead, the peak noise value is underestimated by 50%. If the maximum current density for the entire block is used to estimate noise, the peak-noise voltage is overestimated by three-times.

### **3.4 Chip/Package Co-Design and Solutions to Suppress Noise**

To suppress the power supply noise to a safe level, either an on-chip solution (adding more decaps) or a package level solution (adding more power/ground I/O pads) can be adopted. Decaps are effective when the capacitance value is large enough and when the capacitors are close to the hot spot. Adding decaps is costly for hot spots since the logic is already dense and the layout is already crowded. Decaps also consume substantial gate leakage power. In this situation, package-level high density chip I/O techniques, such as Sea of Leads [53], can be an alternative option. The new physical models can help designers to identify the noise levels of hot spots, calculate how many more pads are needed, and fulfill chip/package co-design.

Adding more P/G pads locally can be quite effective in lowering the power supply noise of the case studied in the previous section. To investigate this point, three cases are compared: (i) no extra pads; (ii) 4 extra pads; (iii) 12 extra pads are utilized in the hot spot region as illustrated in Figure 3.7 (a). As the peak noise changes almost linearly with the increase of current density within the hot spot, as shown in Figure 3.7 (a), adding more pads can always provide more I/O paths for the switching current and therefore reduce the peak noise. For example, for the hot spot current density of  $400 \text{ A/cm}^2$ , the peak noise is approximately 240 mV (Figure 3.7 (b)). The peak noise can be reduced to

165 mV (by 30%) by adding 4 pads and to 130 mV (by 45%) by adding 12 pads into the hot spot region.

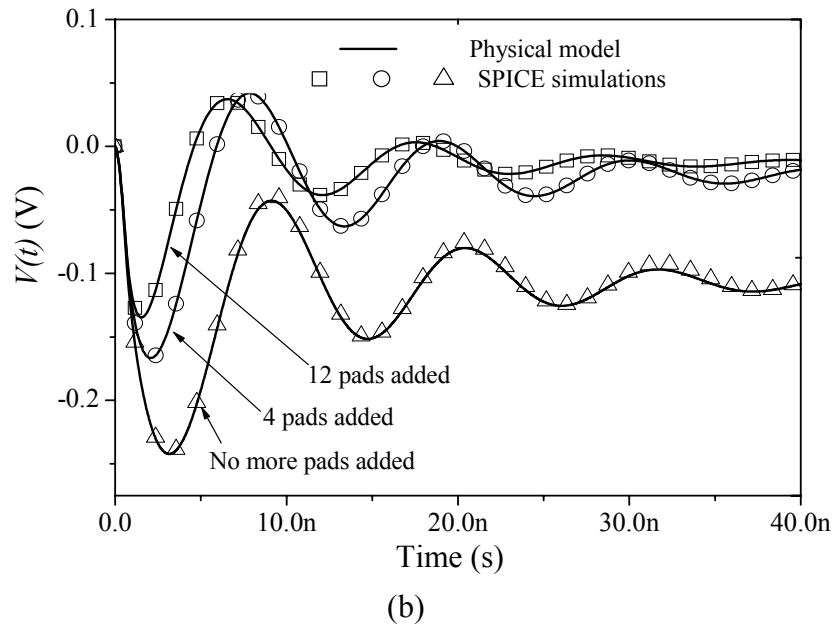
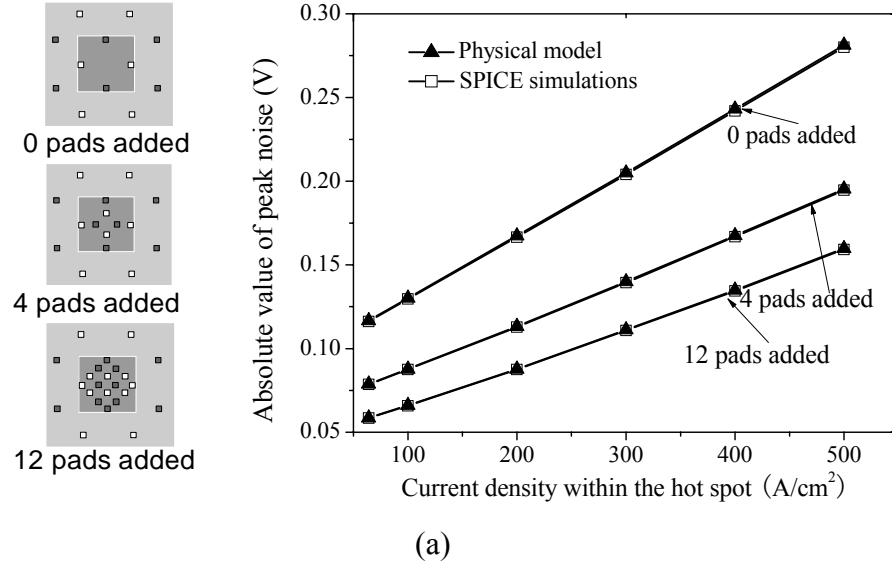


Figure 3.7 (a) Configurations of added pads within the hot spot and peak noise for different pad allocation schemes. (b) Noise waveforms for different pad allocation schemes with a hot spot current density of  $400 A/cm^2$

### **3.5 Conclusion**

Generalized analytical physical models are derived to predict the first droop power supply noise for non-uniform current switching conditions and arbitrary functional block sizes. The models are capable of capturing the impact of package parameters and the distributed nature of power grids, and are also able to deal with the non-uniformity of the current density distribution brought by power-hungry blocks or hot spots. The models can help designers balance various chip and package parameters, provide noise suppressing solutions, and fulfill chip/package co-designs. Comparisons between the models and the results of SPICE simulations are shown in a case study, and there is less than 1% error for both the frequency domain noise model and the projected peak noise value.

# **CHAPTER 4: COMPACT PHYSICAL MODELING AND DESIGN**

## **IMPLICATIONS OF POWER DISTRIBUTION NETWORKS FOR**

### **THREE-DIMENSIONAL (3-D) CHIP STACKS**

#### **4.1 Introduction**

Three-dimensional (3-D) nanosystems can provide enormous advantages in achieving multi-functional integration, improving system speed and reducing the power consumption for future generations of ICs [35]. 3-D chip stacks have been used in commercial products, and today's applications are mainly focused on low power portable devices, such as flash memories and wireless chips. At the high performance end, industry has already started to pave the way for microprocessor stacking and microprocessor-memory stacking, which will extend Moore's Law beyond its expected limits and help break the bottleneck of the memory bandwidth problem for multi-core microprocessors [36, 37]. Through-silicon-vias (TSVs) and micro-bumps are the key technologies to fulfill 3-D chip stacks for high performance applications, and they eliminate the need for long-metal wires that connect today's 2-D chips together, instead, relying on short vertical connections etched through the silicon wafer [36]. These TSVs and micro-bumps enable multiple chips to be stacked together, allowing greater amounts of information to be passed between chips.

However, stacking multiple high-performance dice may result in severe power integrity problems. As shown in Figure 4.1, if multiple high power microprocessors are stacked together and flip-chip technology for 3-D chip stacking is used, several hundred amperes of current (or even more) needs to be delivered to a limited footprint area. Also

the supply current flows through the micro-bumps and narrow TSVs that may exhibit large parasitic inductance. These may potentially lead to a large  $\Delta I$  noise if stacked chips switch simultaneously. Thus, power distribution networks in 3-D systems need to be accurately modeled and carefully designed. In this chapter, analytical models are derived from a set of partial differential equations that describe the frequency-dependent characteristics of the power supply noise in each stack of the chips to obtain physical insight into the rather complex power delivery networks in 3-D systems [59].

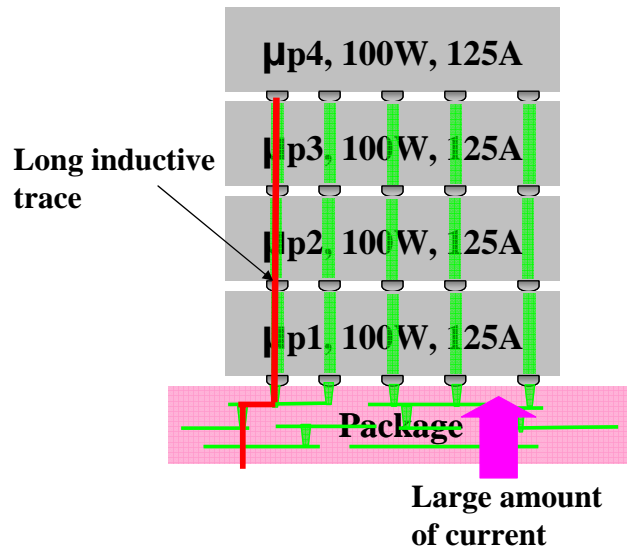


Figure 4.1 Power integrity problem of 3-D chip stack

## 4.2 Mathematical Modeling for 3-D Chip Stacks

In 3-D stacked systems, power is fed from the package through power I/O bumps distributed over the bottom-most chip and then to the upper chips using TSVs and micro-bumps. Each chip is composed of various functional blocks whose footprint can cover a large number of power and ground I/O pads. Power supply noise is modeled assuming the switching current, decap distributions, and TSV allocation within a functional block are uniform. The footprint can be divided into cells, which are identical square regions

between a pair of adjacent quarter power and ground pads, as shown in Figure 4.2. It can be assumed that no current passes in the normal direction relative to the cell borders in each die. Under these assumptions, one cell is enough for the power supply noise analysis.

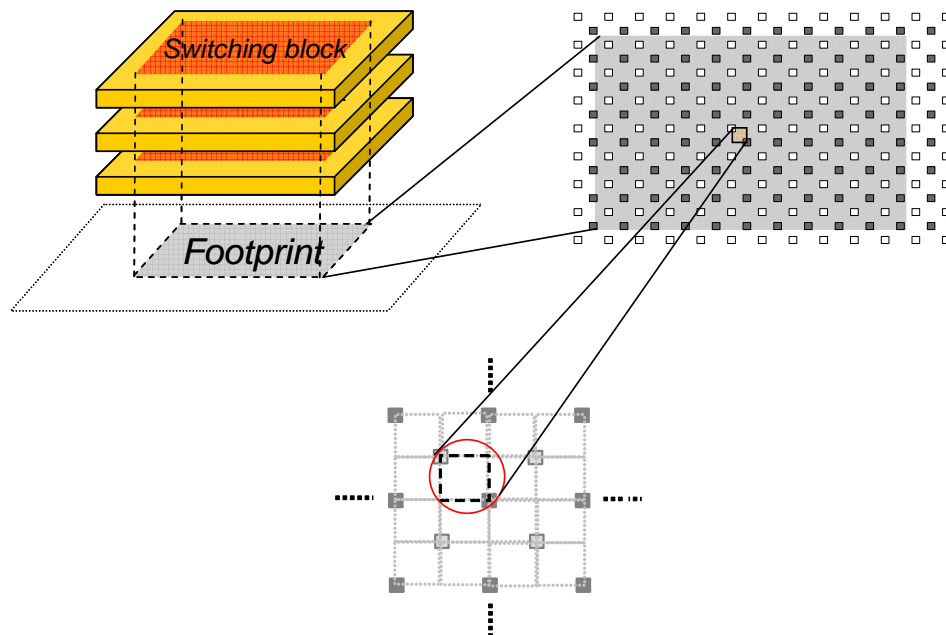


Figure 4.2 Division of the 3-D stack footprint

A simplified circuit model to analyze the power distribution network of 3-D systems is shown in Figure 4.3. A wire between two nodes on the  $i$ -th die is simply modeled as a lumped resistance  $R_{si}$ . The decoupling capacitance per unit area of the  $i$ -th die is represented by  $C_{di}$ . The current density for an active block of Die  $i$  is represented by  $J_i(s)$  in the Laplace domain. Inductance  $L_p$  is the per pad loop inductance associated with the package, connected to the bottom-most die (Die 1). Each silicon TSV is modeled as connected inductor  $L_{via}$  and resistor  $R_{via}$  in series (this includes the parasitics of the micro-bumps when they are used between dice). Symbols  $\Delta x$  and  $\Delta y$  represent the



distances between two adjacent power (or ground) nodes in the same wiring level for  $x$  and  $y$  directions, respectively.

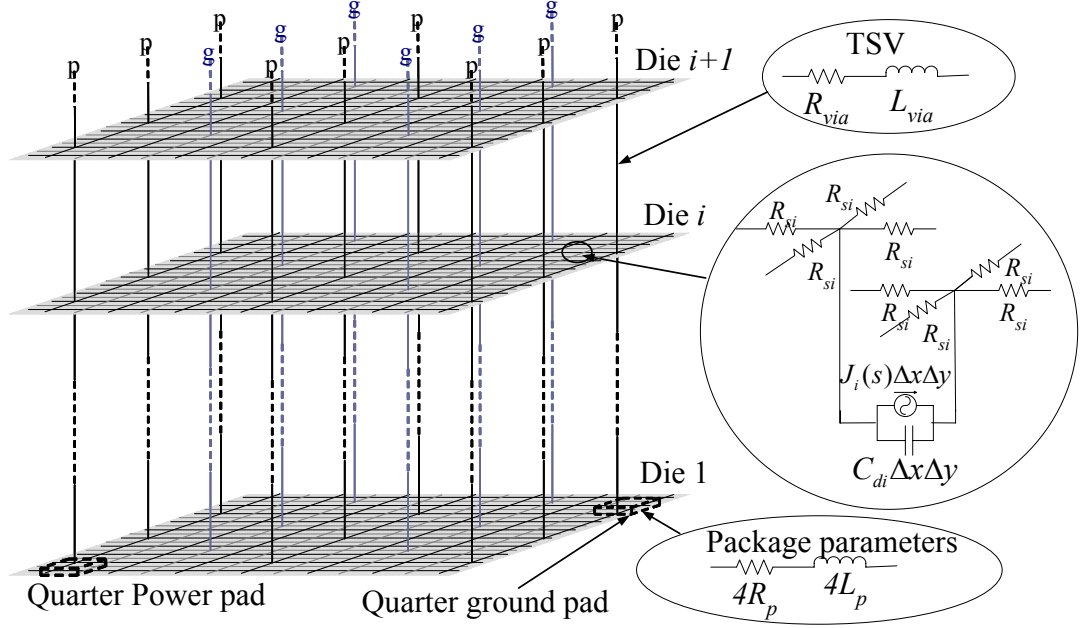


Figure 4.3 Simplified circuit model for 3-D stacked system

The whole structure consists of power and ground grids that can be decoupled into two single grids, as shown in Figure 4.4. Similar to the model derived in Section 2.3, the following partial differential equation describes the frequency characteristics of the power supply noise  $V_i(x, y, s)$  for each node for the  $i$ -th stacked die.

$$\nabla^2 V_i(x, y, s) = R_{si} J_i(s) + 2V_i(x, y, s) \cdot sR_{si} C_{di} + \Phi_i(x, y, s) \quad (4.1)$$

where  $\Phi_i(x, y, s)$  is the source function of PDE of Die  $i$  (except for Die 1) and can be written as

$$\Phi_i(x, y, s) = R_{si} \cdot \sum_{k=1}^{N_{via}} \left( \frac{V_{(i-1)viak} - V_{iviak}}{sL_{viak} + R_{viak}} - \frac{V_{iviak} - V_{(i+1)viak}}{sL_{viak} + R_{viak}} \right) \cdot \delta(x - x_{viak}) \delta(y - y_{viak}) \quad (4.2)$$

Equation (4.2) is derived to account for the discontinuity caused by the TSVs in a die,  $N_{via}$  denotes the total number of vias in each die, and  $V_{iviak}$  is the voltage of Via  $k$

connected to Die  $i$ . Moreover, the source functions are used to make mathematical connections between Die  $(i-1)$ , Die  $i$ , and Die  $(i+1)$ . The source function for Die 1 is written as

$$\Phi_1(x, y, s) = -\frac{R_{s1}}{4sL_p} V_{pad}(s) \cdot \delta(x)\delta(y) + R_{s1} \cdot \sum_{k=1}^{N_{via}} \left( -\frac{V_{1via k} - V_{2via k}}{sL_{via k} + R_{via k}} \right) \cdot \delta(x - x_{via k})\delta(y - y_{via k}). \quad (4.3)$$

In (4.3), the first term accounts for the contribution of the package inductance, where  $V_{pad}$  is defined as the voltage of the power/ground pad in Die 1.

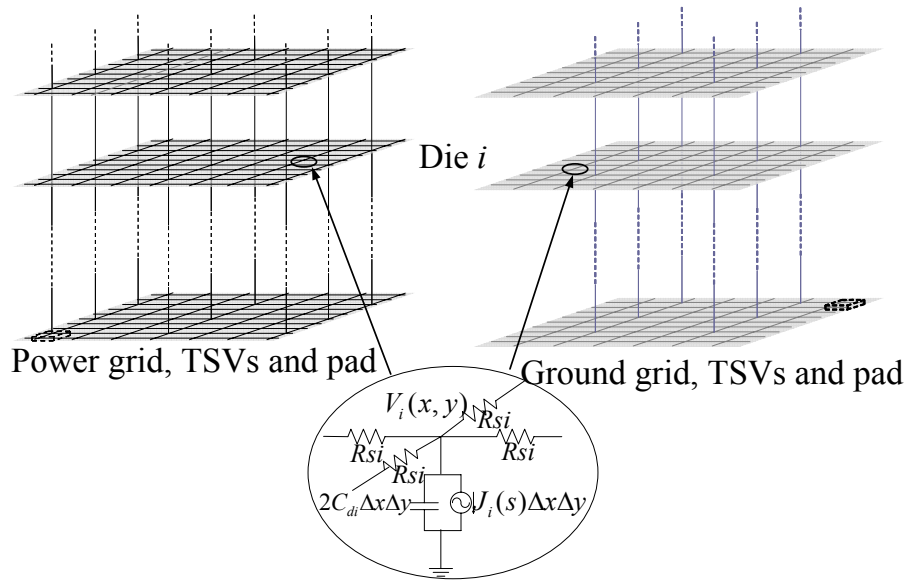


Figure 4.4 Circuit model for single grid structure

Since no current flows normally through the cell boundaries, the partial differential equation should satisfy the following boundary conditions:

$$\frac{\partial V_i}{\partial y} \Big|_{x=0} = 0, \quad \frac{\partial V_i}{\partial y} \Big|_{x=a} = 0, \quad \frac{\partial V_i}{\partial x} \Big|_{y=0} = 0, \quad \frac{\partial V_i}{\partial x} \Big|_{y=a} = 0, \quad (4.4)$$

where  $a$  is the cell size.

Similar to the derivation process in (3.4)-(3.8), (4.1) can be transformed into a pure Hemholtz equation and be solved analytically by applying the boundary condition of the second kind described by (4.4). The supply noise at Die  $i$  and Die 1 are

$$V_i(x, y, s) = R_{si} \cdot \sum_{k=1}^{N_{via}} \left( \frac{V_{(i-1)via_k} - V_{ivia_k}}{sL_{via_k} + R_{via_k}} - \frac{V_{ivia_k} - V_{(i+1)via_k}}{sL_{via_k} + R_{via_k}} \right) \cdot G(x, y, x_{via_k}, y_{via_k}, s) - \frac{J_i(s)}{2sC_{di}}, \quad (4.5)$$

and

$$V_1(x, y, s) = -\frac{R_{s1}}{4sL_p} V_{pad}(s) \cdot G(x, y, 0, 0, s) + R_{s1} \cdot \sum_{k=1}^{N_{via}} \left( -\frac{V_{1via_k} - V_{2via_k}}{sL_{via_k} + R_{via_k}} \right) \cdot G(x, y, x_{via_k}, y_{via_k}, s) - \frac{J_1(s)}{2sC_{d1}}. \quad (4.6)$$

where  $G(x, y, \xi, \eta, s)$  is the Green's function of a Helmholtz equation with boundary condition of the second kind. Since,  $V_{pad}$  and  $V_{ivia_k}$  are unknowns in (4.5) and (4.6), but they should be the solutions of (4.5) and (4.6). We can put them on the left-hand sides of (4.5) and (4.6) and solve for them. If there are  $K$  vias in each die, an equation set with  $[K(N-1)+1]$  equations and  $[K(N-1)+1]$  unknowns ( $V_{pad}$  and each  $V_{ivia_k}$ ) needs to be solved. Consequently, the frequency characteristics of the power supply noise for Die  $i$   $V_i(x, y, s)$  can be solved analytically from (4.5) and (4.6).

The time domain transient noise can be obtained by performing an inverse Laplace transform on  $V_i(x, y, s)$ . The peak noise for each node can also be identified by adding up the transient noises of power and ground grids.

### 4.3 Model Validation

A comparison between the physical model and the results of SPICE simulations is performed for a stack of five high performance MPUs at the 45 nm node considering the top two metal levels of each chip, as shown in Figure 4.5. The setup of each chip is the same as the simulations performed in Section 3.3. It is assumed that five critical functional

blocks in each of the dice share the same footprint and switch simultaneously with an identical on-current density of  $100 \text{ A/cm}^2$ . The five chips are stacked together through micro-bumps of  $100 \text{ }\mu\text{m}$  diameter and TSVs of  $50 \text{ }\mu\text{m}$  diameter and height of  $200 \text{ }\mu\text{m}$ . The effective inductance (half of the loop inductance for two adjacent traces) can be calculated as  $0.06 \text{ nH}$  by RAPHAEL [14]. For the figures to show the chip stacking structures in Sections 4.3-4.5, the die with gray shade denotes the die which is switching, and the arrow points to the die for which the power supply noise is to be examined. The worst case noise, which is the main concern in digital systems, normally occurs at the corners of the grid cell (furthest from power/ground pads). This is similar to the previous findings in the case of a single chip in Chapter 2. In Figure 4.5, all dice are shaded and this represents that all dice are switching.

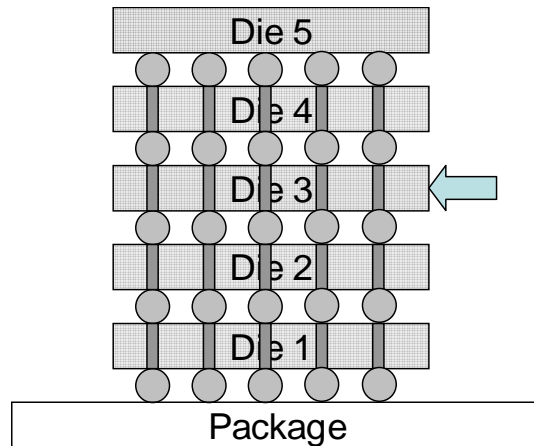


Figure 4.5 Five chip stacking for model validation: All dice are switching, and Die 3 is examined

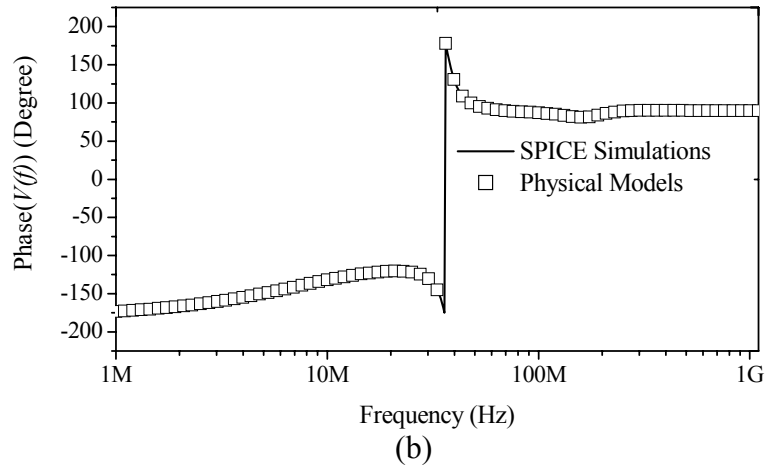
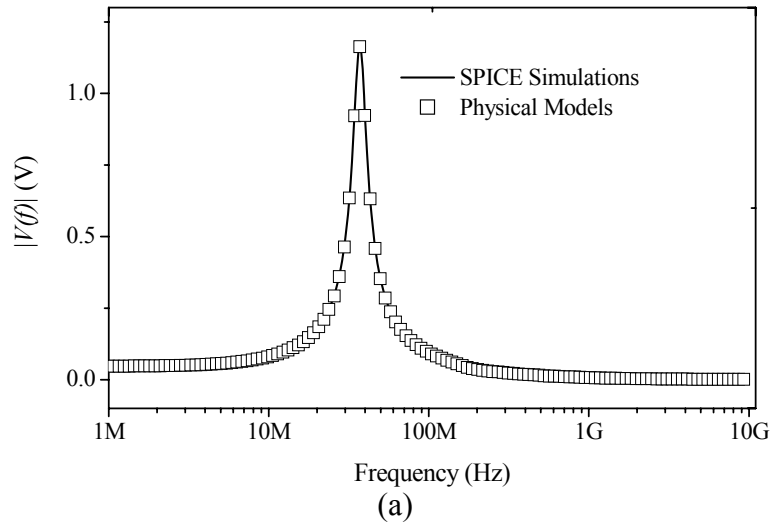


Figure 4.6 Frequency response of the worst case noise for the third die: (a) Magnitude; (b) Phase.

Figure 4.6 illustrates the frequency domain response (magnitude and phase) for the worst case noise of the third die. The results are also compared against the results of SPICE simulations, and the new models show less than 4% error. The transient supply noise of the worst case scenario is also obtained and is represented by the solid line in Figure 4.7. Comparing with the results of SPICE simulations (square dots), the peak noise value has less than 4% error.

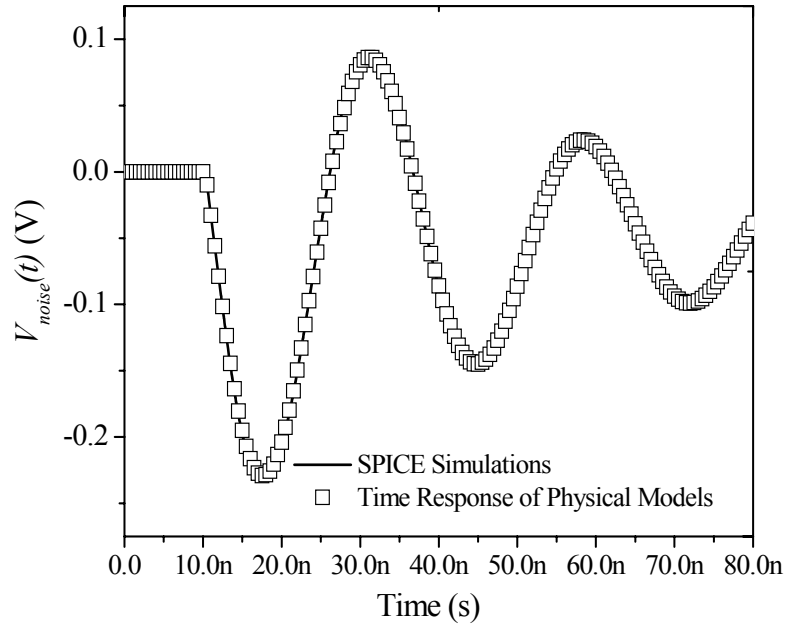


Figure 4.7 Time domain response of the worst case noise for the third die

#### 4.4 Design Implication for 3-D Integration

In the following context, the models will be used to address the power integrity problem of 3-D integration systems. To make a worst case scenario analysis, the worst case peak noise value will be considered.

The case when a single die is switching is considered first. This can be the case when one die dissipates considerably larger power relative to the other dice in the stack. An example of such a system is a processor die with several memory dice. A stack of 10 dice is modeled in Figure 4.8 (a). As the switching die becomes further away from the package, an increase of worst case peak noise can be seen from Figure 4.8 (b). This is mainly because of the longer current trace and therefore a larger parasitic inductance and resistance as the switching block is further away.

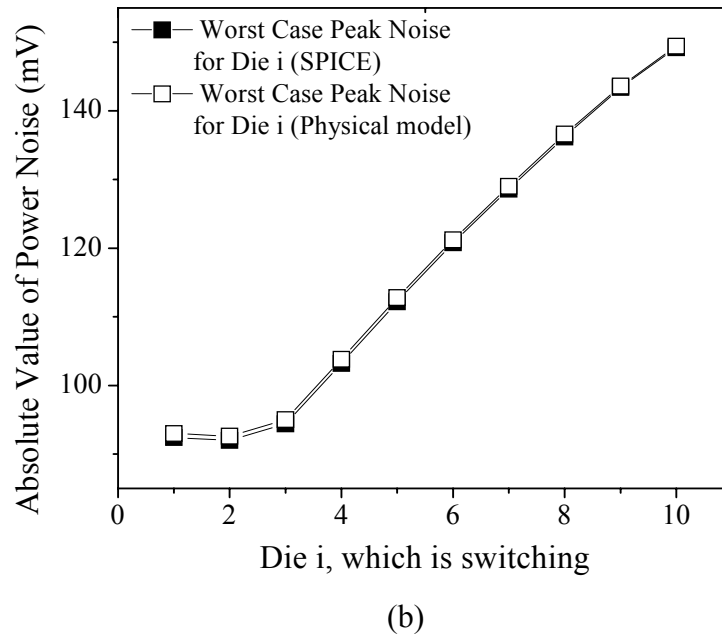
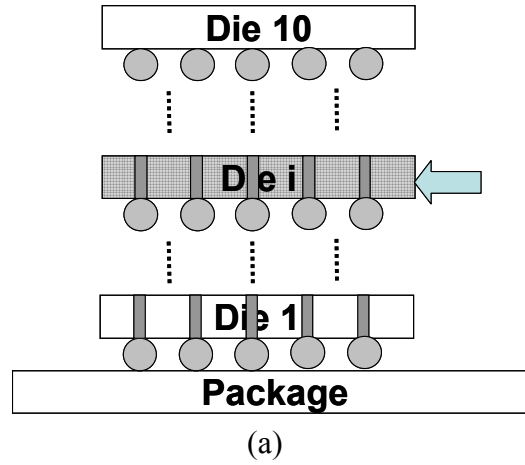


Figure 4.8 Single die switching, changing the switching die

The increasing functional complexities of digital systems results in a higher level of integration, and more and more dice will be stacked together to achieve this goal. It can be seen in Figure 4.9 that if the total number of the stacked dice is increasing, the noise level for the topmost die decreases when the number of dice is less than 6. This is because non-switching dice behave as decaps for the switching die. However, when the number of dice increases beyond 6, the increase in decaps can not compensate the impact

of the longer inductive TSV traces and micro-bumps associated with those added dice, which result in the increase of the noise level.

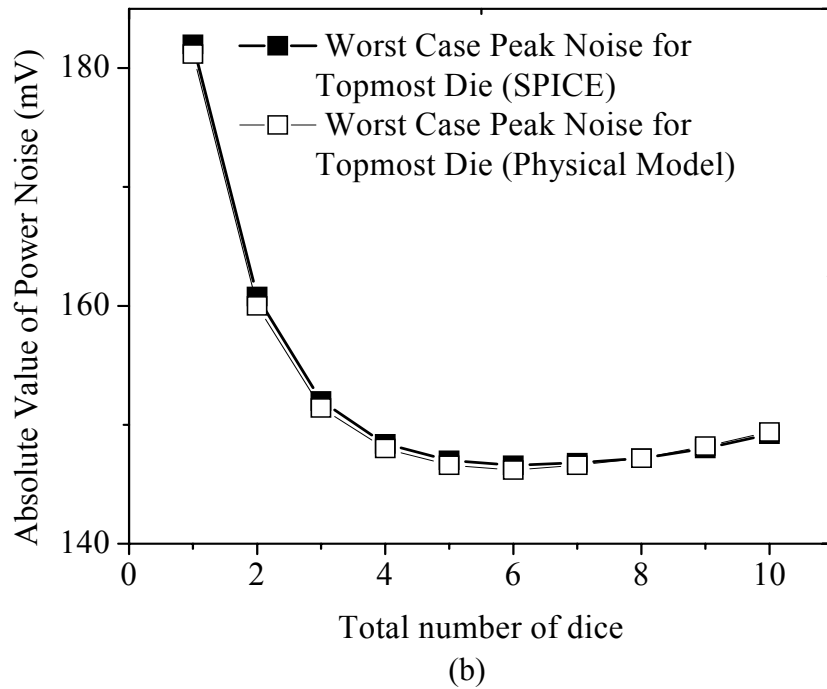
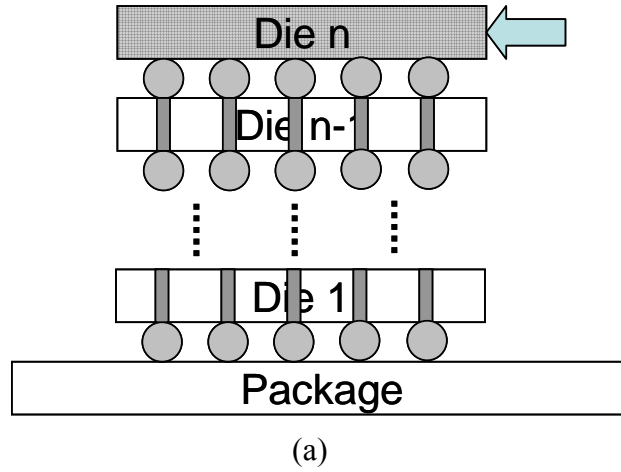


Figure 4.9 Single die switching, increasing total number of dice

If only one die is switching, the noise is smaller than the single chip case (2-D case), because the switching dice can use the decaps of those non-switching dice in the 3-



D stacks. However, normally the activities of the two blocks with the same footprints are highly correlated because an important purpose of 3-D integration is to put the blocks that communicate most as close to each other as possible, as shown in Figure 4.10.

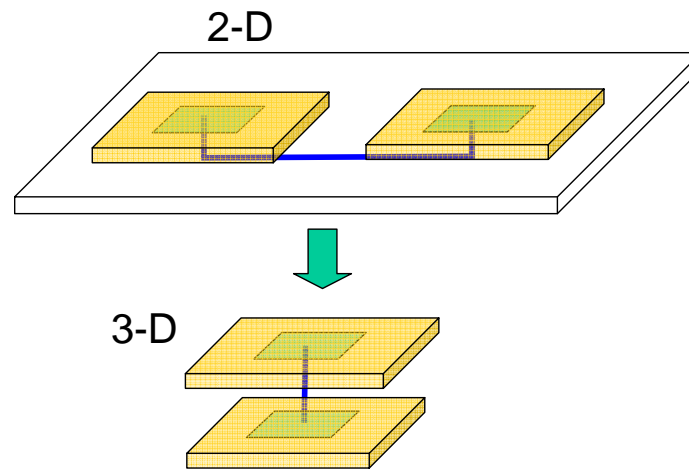


Figure 4.10 Making shorter interconnects between communicated blocks by using 3-D integration

Therefore, we must consider the worst case scenario when all the functional blocks sharing the same footprint switch simultaneously, as shown in Figure 4.11. If the total number of dice is increased and the noise levels of the topmost and bottommost levels are examined, it can be seen that when all dice are switching the noise produced in a 3-D integrated system is unacceptable when compared to a single chip case. This is especially true for the topmost die where the noise level changes dramatically (180 mV for the single die case as opposed to 790 mV for the 10 dice case). Even for the bottommost die, methods of suppressing the noise need to be identified.

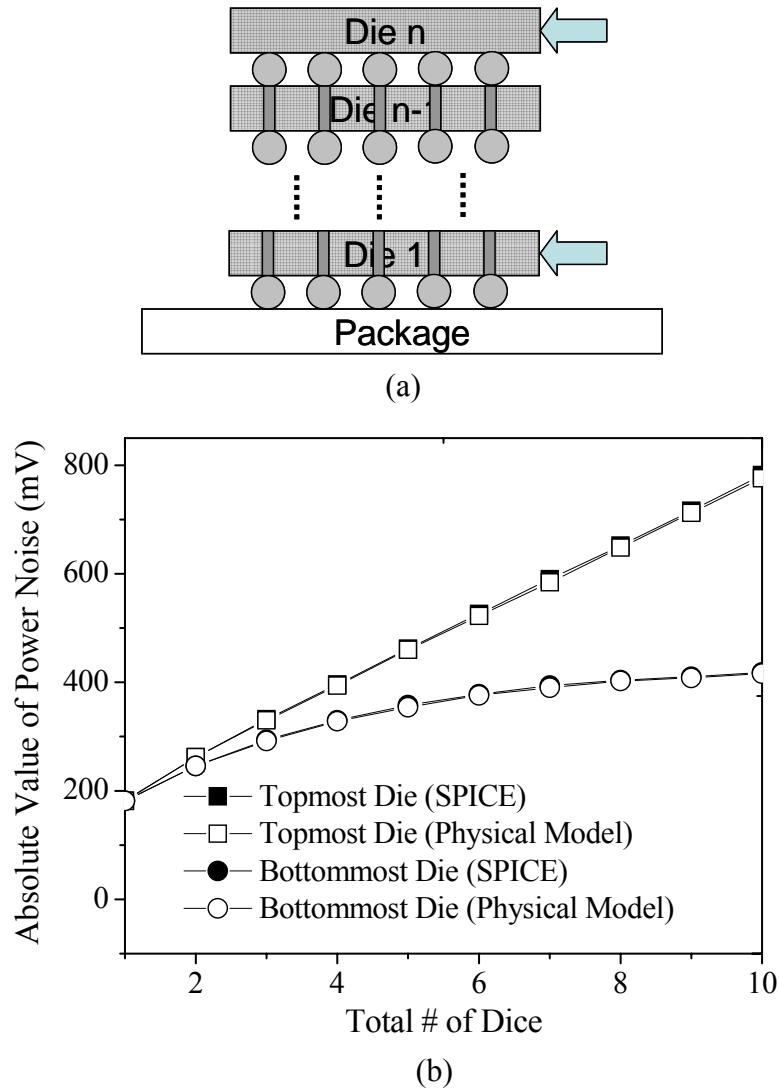


Figure 4.11 All dice switching, increasing total number of dice

#### 4.5 Solutions for Suppressing Noise Level in 3-D systems

Traditionally, as presented in Chapter 3, to suppress the noise to a safe level, designers can either add more decaps in a logic chip or add more power/ground I/O pads between the chip and package. In 3-D systems, power integrity problems arise from the third dimension, and the solutions can be pushed into the third dimension as well. In this section, new method are presented in a “3-D” way to tackle the 3-D problem.

##### “Decaps” Die

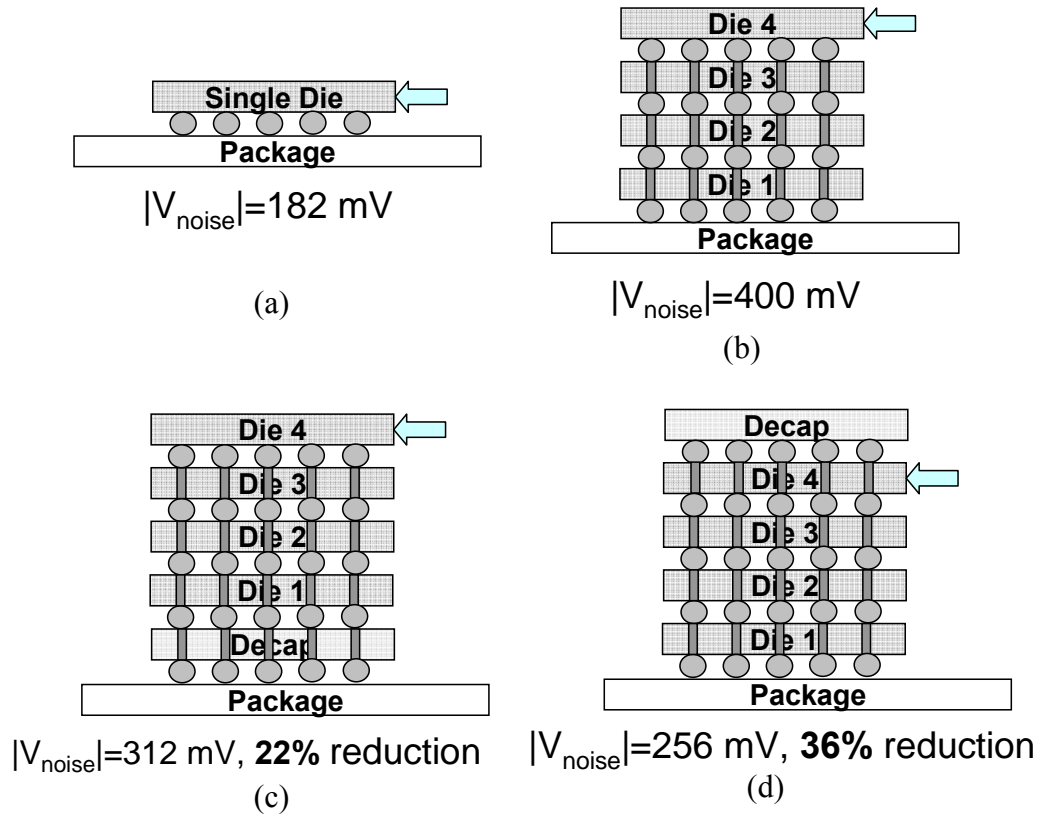


Figure 4.12 Effect of adding one "decap" die. (a) Single die switching; (b) Four switching dice without "decap" die; (c) Four switching dice with "decap" die at the bottom; (d) Four switching dice with "decap" die on the top

If a whole die is used as decaps (100% area is occupied by decaps), and the "decap" die is stacked with other dice, the noise can be suppressed to some extent. For example, as shown in Figure 4.12, if the same setup as previous sections is adopted and 4 dice with one "decap" die are stacked together, putting the "decap" die on the top results in 36% reduction in the worst case peak noise of Die 4 (256 mV in Figure 4.12 (d) compared to 400 mV Figure 4.12 (b)). Putting the "decap" die at the bottom of the stack can lead to 22% reduction for Die 4 (312 mV in Figure 4.12 (c) compared to 400 mV in Figure 4.12(b)). Although there are improvements resulting from the "decap" die, still more "decap" dice need to be added to achieve the noise level of a single die (2-D case, 182 mV in Figure 4.12 (a)).

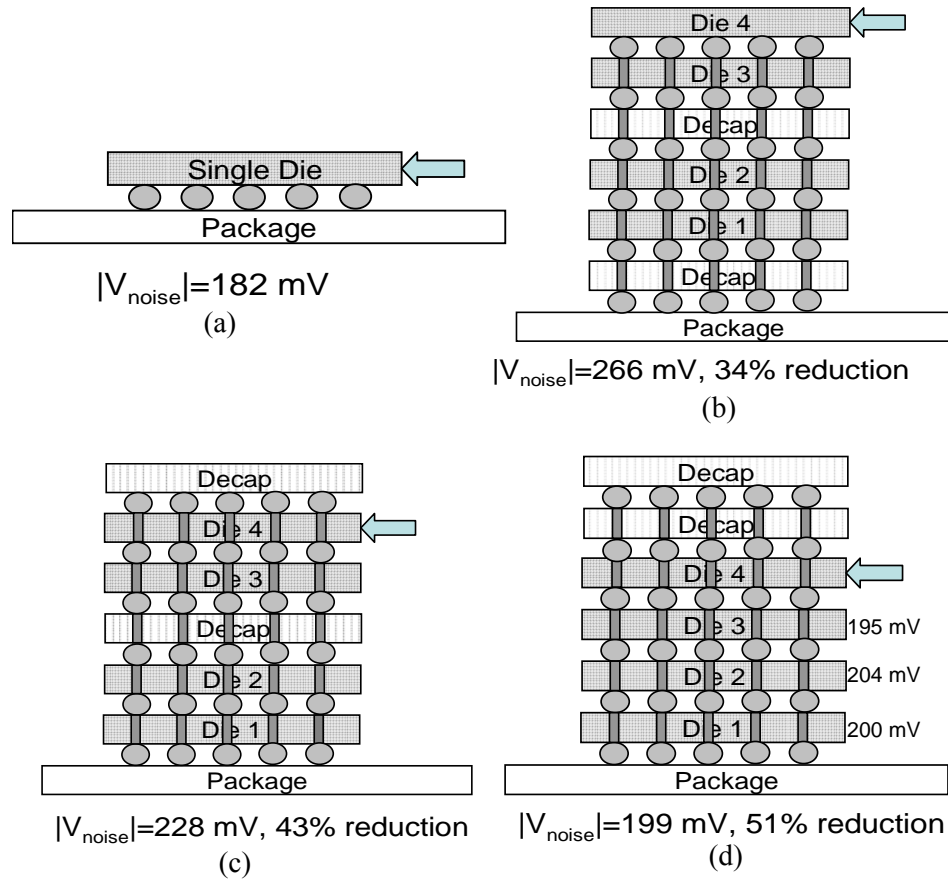


Figure 4.13 Effect of adding two “decap” dice. (a) Single die switching; (b) One “decap” die at the bottom and the other in the middle; (c) One “decap” die in the middle and the other on the top; (d) Both “decap” dice on the top

Figure 4.13 (b), (c) and (d) illustrate the case of different schemes of using two “decap” dice. By putting the two “decap” dice on the top, the noise of Die 4 can be suppressed to 199 mV. The noise levels of other dice (Die 1-3) which are away from the decap dice are also kept around 200 mV, which is close to the level of single chip (2-D). It can be seen that putting the “decap” dice on the top is the best scheme to suppress the noise of the fourth die, as shown in Figure 4.13 (d).

Instead of adding a “decap” die, it is more efficient if high-k material is used between the power and ground planes (on-chip) [59]. A drawback of this technique is the cooling problem since “decap” dice block the cooling path for other dice. It should be

emphasized that cooling also presents major challenges to 3-D integration and the newly developed microfluid cooling technique can potentially alleviate this cooling problem [60].

### Adding More TSVs

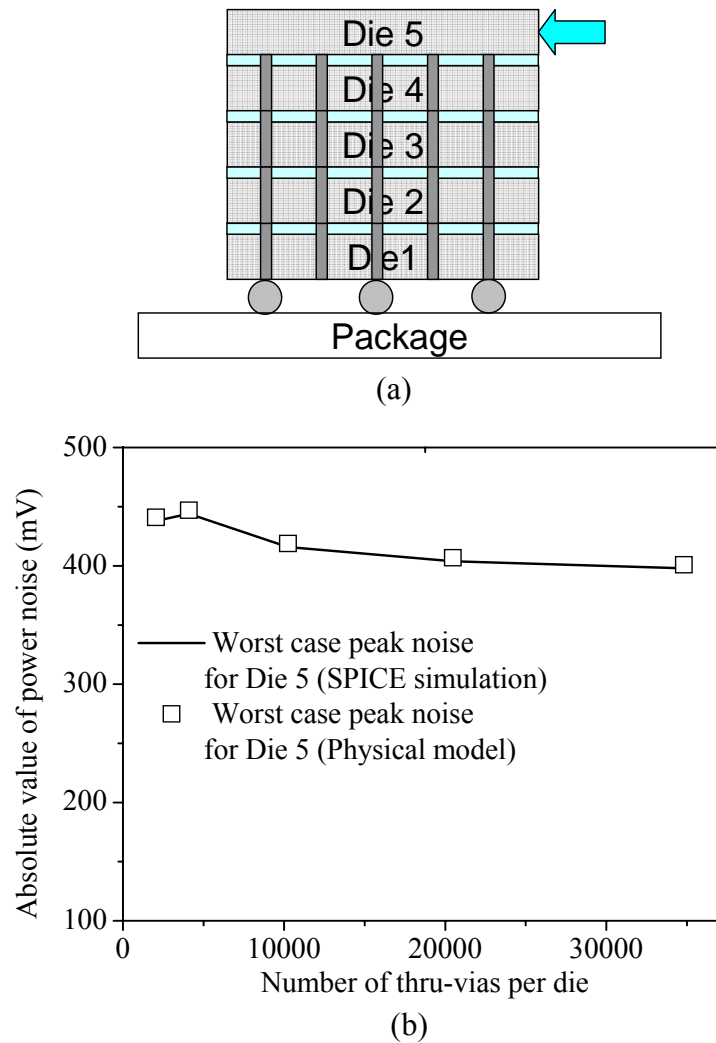


Figure 4.14 Effect of adding more TSVs: fixing the number of power/ground I/Os

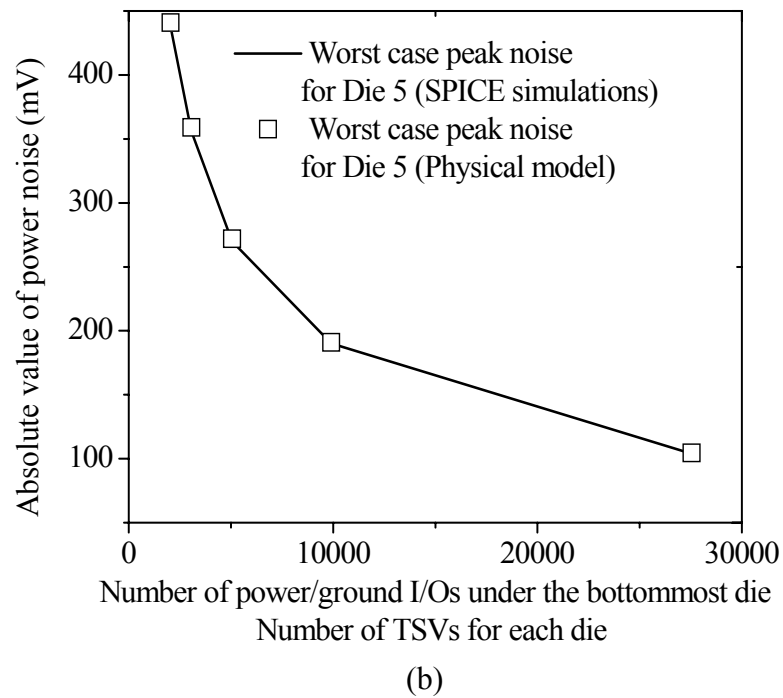
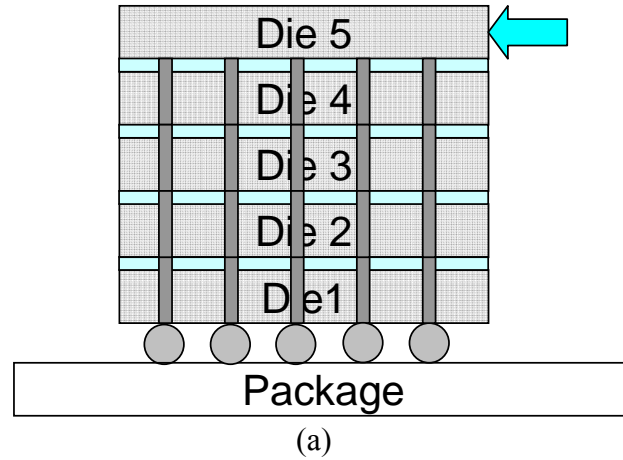


Figure 4.15 Effect of adding more TSVs and power/ground I/Os

Another possible solution is to use more TSVs. To examine the efficiency of increasing the number of TSVs, in the first case, a five die stacking structure is used, and the total number of power/ground I/Os is fixed as 2048. As shown in Figure 4.14, it is noted that one can not benefit by solely increasing the number of TSVs. Because the parasitics of TSVs are much smaller than those of the package, and only small changes

for noise level can be obtained by increasing the number of TSVs. Adding more TSVs might even make designers lose benefits because TSVs consume die area that would be potentially used for decaps or additional circuits for noise suppressing purposes. In the second case, the numbers of both P/G pads and TSVs in each dice are increased. This causes the power supply noise to greatly reduce and even reach the level of a single chip, as shown in Figure 4.15. These two cases show that the bottleneck is still power/ground I/Os as they have a critical role in determining the power supply noise. The inductance of the package is the dominant part throughout the whole power delivery path for the first droop noise. Therefore, the power integrity problem needs an I/O solution that can provide high-density interconnection without sacrificing the mechanical attributes needed for reliability.

## **4.6 Conclusion**

In this chapter, analytical physical models are derived to incorporate the impact of 3-D integration on the first droop power supply noise. The models enable power design engineers to identify the challenges in power delivery network design due to the larger supply current and longer power delivery paths in 3-D integrated systems. The physical models not only consider the distributed nature of the power grids and decaps, but also capture the impacts of package parameters and 3-D design parameters. The models have less than 4% error compared to the results of SPICE simulations. Based on the models, design guidelines are also proposed to address the power integrity problem for 3-D integration. The relationships between the power supply noise, decap insertion, power/ground I/O allocation, and TSVs allocation are discussed quantitatively. Schemes

for reducing the power supply noise in 3-D integration systems are also proposed and their impacts on future 3-D system designs are also emphasized in this chapter.



## **CHAPTER 5: COMPACT PHYSICAL MODELS FOR CHIP AND PACKAGE POWER AND GROUND DISTRIBUTION NETWORKS**

### **5.1 Introduction**

Flip-chip technology introduces significant chip and package design complexities [29], and allows progressively faster devices packed at greater densities. However, they demand careful co-design of chip- and package-level power distribution networks to avoid excessive power supply noise with the minimum use of resources. The first droop noise is a ringing effect tightly associated with the chip and package resources (on-chip decaps and package inductance), and thus optimum designs require integrated chip/package co-design efforts to quickly evaluate the tradeoffs available in silicon chips and packages. This chip/package combination must be viewed throughout various design phases especially for power integrity aware designs.

However, in reality, two sets of tools (PEEC method and SPICE simulations for chips and electromagnetic solvers for packages) are used for chip and package modeling. Those tools are based on different methodologies and bridging them becomes challenging. It is of great importance to build up a tight linkage between chip and package designs, and this requires a unified platform that can fulfill real co-designs. In this chapter, a compact physical modeling scheme is used to make it happen.

Compact physical models for the first droop power supply noise have been proposed in Chapter 2-4 to address the power delivery issues caused by power hungry blocks, hot spots and 3-D integration. These models embody the distributed nature of on-chip power grids, decaps and current sources. A simple representation of the package was

used with terminating chip I/O interconnects with lumped inductors. In this chapter, these models are further extended. Distributed representations of power and ground planes are added to accurately consider the noise propagation in the package and help complete fast and accurate chip/package co-designs [61]. In the package, multiple power/ground planes can be simplified into a plane pair [10]. The entire chip surface and package areas are chosen for the modeling. The frequency characteristics of power supply noise voltages on the chip and in the package are described mathematically by two partial differential equations (PDEs), respectively. These models are validated using the commercialized power integrity validation tool SPEED2000 [22], and the analysis is performed for a ceramic package module designed by IBM.

## **5.2 Mathematical Modeling for Chip and Package Power Distribution Networks**

An overview of the power delivery system of GSI systems is illustrated in Figure 5.1. The supply current comes from the DC-DC converter at the board level, and is fed into the package through a ball grid array (BGA) [29]. The current then flows through power planes and vias in the package, enters the chip through a solder bump I/O array (IBM uses the term C4, Controlled Collapse Chip Connection, for the chip to package I/Os), and finally is distributed to on-chip circuitry by on-chip power/ground grids. The current returns through an opposite path. The power delivery traces in the package and on the board are associated with a certain amount of inductance, which results in voltage fluctuation in the power distribution network. To decouple the power supply noise, decaps are allocated at the chip and package levels to bypass the high frequency components in the noise transients. On-chip grids and decaps together with power ground planes and vias play important roles for the first droop power supply noise. The first

droop noise is the focus of this chapter, and a perfect board is assumed since board parameters only influence the second and third droops. It is assumed that the power and ground BGAs are AC shorted at board level.

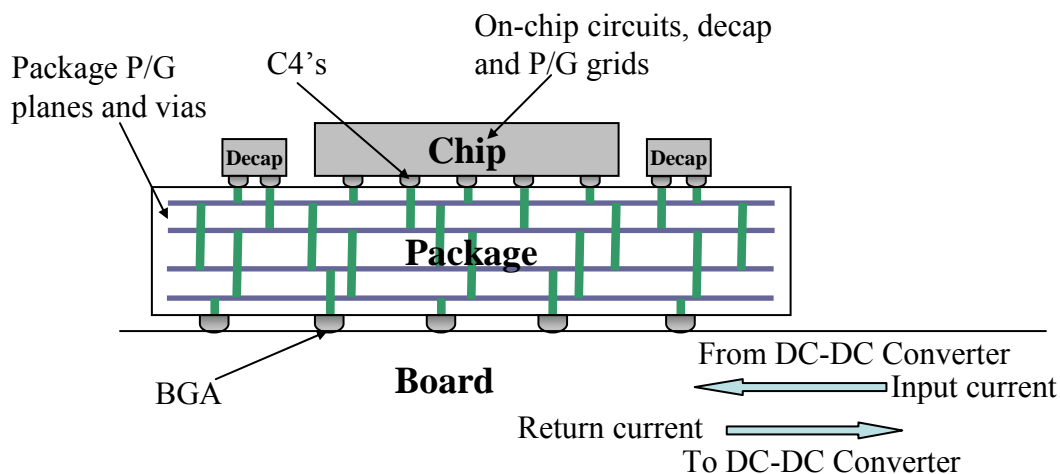


Figure 5.1 Power delivery system of GSI systems

Compact physical models for the first droop power supply noise have been presented in Chapters 2-4, and simplified circuit modes are used to analyze on-chip power distribution networks. In this chapter, the same circuit model is built up for the chip-level power distribution networks. As shown in Figure 5.2, a wire between the two nodes is simply modeled as a lumped resistance  $R_c$ . The amount of decoupling capacitance per unit area is represented by  $C_c$ . The current density for an active block is represented by  $J_c(s)$  in the Laplace domain. Symbols  $\Delta x$  and  $\Delta y$  represent the distances between two adjacent power (or ground) nodes in the same wiring level for  $x$  and  $y$  directions, respectively.

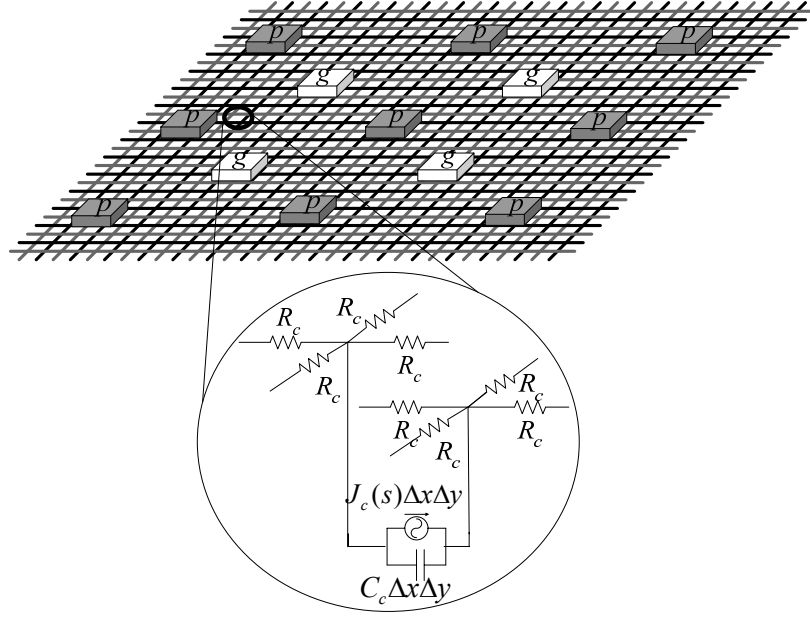


Figure 5.2 Simplified circuit model of on-chip power/ground grids

A high-performance GSI system requires a package with multiple power/ground planes, because a large number of signal lines have to be routed on several interconnect levels in the package. These signal interconnect levels have to be placed between or over power/ground planes to have an impedance-controlled environment and also to prevent coupling of signal lines in different levels. To speed up the analysis process, multiple power/ground planes can be snapped into a plane pair [10]. There were quite a few past works, such as [62], which proposed equivalent circuit models similar to the simplified circuit models for on-chip grids to handle a power/ground plane pair. As shown in Figure 5.3, a power/ground plane pair can be divided into unit cells, and if dielectric loss is neglected, each cell consists of an equivalent circuit with  $R_p$ ,  $L_p$  and  $C_p$ . Symbol  $R_p$  represents the distributed resistance per unit area,  $L_p$  is the distributed inductance per unit area, and  $C_p$  is the distributed parallel plate capacitance per unit area between power and ground planes. The calculation of  $R_p$ ,  $L_p$  and  $C_p$  can be found in [62]. Symbols  $\Delta x_p$  and  $\Delta y_p$  represent the sizes of each unit cell in  $x$  and  $y$  directions, respectively.

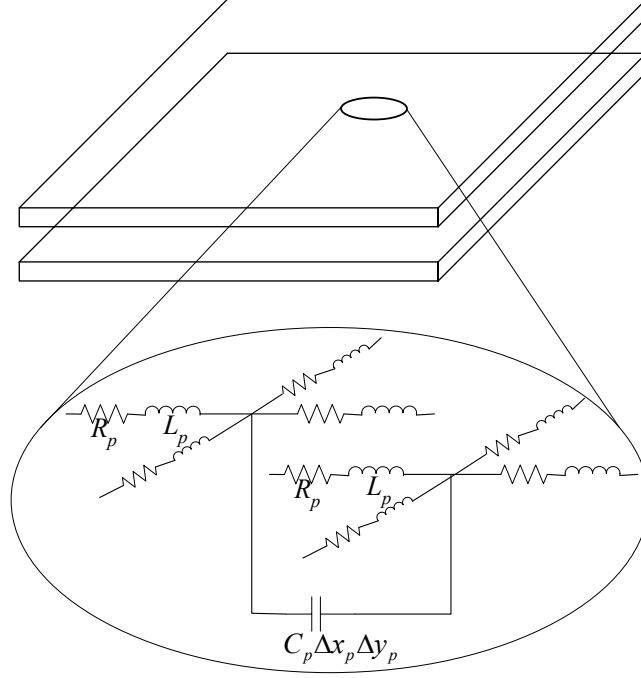


Figure 5.3 Simplified circuit model of package power/ground planes

The chip and package are connected through C4 bumps and I/O pads. The parasitics of a C4 bump and I/O pad pair is represented by a resistor  $R_{c4}$  and an effective inductor  $L_{c4}$ . Similarly, each BGA bump, connecting package and board, can be represented by a resistor  $R_{bga}$  and an effective inductor  $L_{bga}$ .

Therefore, the following partial differential equation (PDE) can describe the frequency characteristics of the power supply noise  $V(x,y,s)$  for each node on a single power/ground grid [61].

$$\nabla^2 V_c(x, y, s) = V_c(x, y, s) \cdot R_c \cdot 2sC_c + \Phi_c(x, y, s). \quad (5.1)$$

Also, in the package, a power/ground plane is naturally a continuous planar surface. Thus, the PDE to describe the frequency characteristics of a power/ground plane can be written as

$$\nabla^2 V_p(x, y, s) = V_p(x, y, s) \cdot (R_p + sL_p) \cdot 2sC_p + \Phi_p(x, y, s). \quad (5.2)$$

In (5.1) and (5.2),  $\Phi_c(x,y,s)$  and  $\Phi_p(x,y,s)$  are the source functions for the chip and package PDEs, respectively. They can be described as

$$\Phi_c(x, y, s) = - \sum_{i=1}^{N_{sp}} R_c J_c(s) \cdot \Delta x \cdot \Delta y \cdot \delta(x - x_{spi}) \delta(y - y_{spi}) - \sum_{j=1}^{N_{C4}} \left[ \frac{R_c}{(sL_{C4j} + R_{C4j})} \cdot (V_{c-C4j}(s) - V_{p-C4j}(s)) \right] \cdot \delta(x - x_{C4j}) \delta(y - y_{C4j}) \quad (5.3)$$

and

$$\Phi_p(x, y, s) = \sum_{i=1}^{N_{C4}} \left[ \frac{(R_p + sL_p)}{(sL_{C4i} + R_{C4i})} \cdot (V_{c-C4i}(s) - V_{p-C4i}(s)) \right] \cdot \delta(x - x_{C4i}) \delta(y - y_{C4i}) - \sum_{j=1}^{N_{BGA}} \left[ \frac{(R_p + sL_p)}{(sL_{BGAj} + R_{BGAj})} V_{p-BGAj}(s) \right] \cdot \delta(x - x_{BGAj}) \delta(y - y_{BGAj}) - \sum_{k=1}^{N_{Decap}} \left[ \frac{(R_p + sL_p)}{\left( sL_{ESLk} + R_{ESRk} + \frac{1}{sC_{Decap}} \right)} V_{p-Decapk}(s) \right] \cdot \delta(x - x_{Decapk}) \delta(y - y_{Decapk}) \quad (5.4)$$

Function  $\Phi_c(x,y,s)$  is used to account for the impact of on-chip switching circuits, power/ground I/O pads and C4 bumps on the chip PDE. The first summation in  $\Phi_c(x,y,s)$  represents the switching current induced from on-chip functional blocks, and  $(x_{spi}, y_{spi})$  is the location of each switching node. The second summation represents the current flowing through the power/ground I/O pads and C4 bumps. The location of each C4 bump is denoted by  $(x_{C4j}, y_{C4j})$ . The term  $[V_{c-C4j}(s) - V_{p-C4j}(s)]$  is the voltage across the  $j$ -th I/O pad and C4 bump pair. It should be noted that  $\Phi_c(x,y,s)$  is also a function of package voltages  $[V_{p-C4j}(s)]$  at the C4 locations.

Function  $\Phi_p(x,y,s)$  is used to account for the impact of power/ground I/O pads, C4 bumps, BGA bumps, and package level decaps on the package PDE. The first summation in  $\Phi_p(x,y,s)$  represents the current flowing through the power/ground I/O pads and C4 bumps. The second summation represents the current flowing through the BGA bumps.

The location of each BGA bump is denoted by  $(x_{BGAj}, y_{BGAj})$ . The third summation represents the current bypassed by package-level decaps, and  $(x_{Decapk}, y_{Decapk})$  is the location of Decap  $C_{Decapk}$ . Also,  $L_{ESLk}$  and  $R_{ESLk}$  denote the equivalent series inductance (ESL) and equivalent series resistance (ESR) of each decap, respectively. Also, we can observe that  $\Phi_p(x,y,s)$  is a function of chip voltages  $[V_c - C_{4j}(s)]$  at the C4 locations.

$\Phi_c(x,y,s)$  and  $\Phi_p(x,y,s)$  also make links between two PDEs using the voltages across I/O pads and C4 bumps.

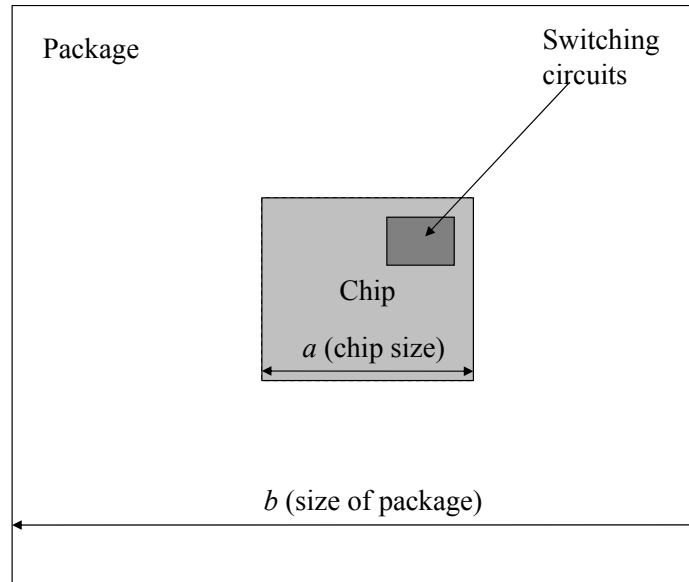


Figure 5.4 The setup of the boundary conditions for PDEs

As shown in Figure 5.4, if the whole chip surface and whole package areas are chosen for the modeling, it is found that there is no current flowing normally through the chip and package boundaries. Thus, both the two PDEs have boundary conditions of the second kind. If the shapes of the chip and package are square, the boundary conditions for the PDEs can be written as

$$\frac{\partial V_c}{\partial y} \Big|_{y=0} = 0, \quad \frac{\partial V_c}{\partial y} \Big|_{y=a} = 0, \quad \frac{\partial V_c}{\partial x} \Big|_{x=0} = 0, \quad \frac{\partial V_c}{\partial x} \Big|_{x=a} = 0, \quad (5.5)$$

and

$$\frac{\partial V_p}{\partial y} \Big|_{y=0} = 0, \quad \frac{\partial V_p}{\partial y} \Big|_{y=b} = 0, \quad \frac{\partial V_p}{\partial x} \Big|_{x=0} = 0, \quad \frac{\partial V_p}{\partial x} \Big|_{x=b} = 0, \quad (5.6)$$

where  $a$  and  $b$  are the sizes of the chip and package, respectively.

PDEs (5.1) and (5.2) can be solved analytically by putting in the boundary conditions of the second kind described by (5.5) and (5.6) [46]. The noise voltages for the chip and package are

$$V_c(x, y, s) = - \sum_{i=1}^{N_{sp}} R_c J_c(s) \cdot \Delta x \cdot \Delta y \cdot G(x, y, x_{spi}, y_{spi}, s) - \sum_{j=1}^{N_{C4}} \left[ \frac{R_c}{(sL_{C4j} + R_{C4j})} \cdot (V_{c-C4j}(s) - V_{p-C4j}(s)) \right] \cdot G(x, y, x_{C4j}, y_{C4j}, s), \quad (5.7)$$

and

$$V_p(x, y, s) = \sum_{i=1}^{N_{C4}} \left[ \frac{(R_p + sL_p)}{(sL_{C4i} + R_{C4i})} \cdot (V_{c-C4i}(s) - V_{p-C4i}(s)) \right] \cdot G(x, y, x_{C4i}, y_{C4i}, s) - \sum_{j=1}^{N_{BGA}} \left[ \frac{(R_p + sL_p)}{(sL_{BGAj} + R_{BGAj})} V_{p-BGAj}(s) \right] \cdot G(x, y, x_{BGAj}, y_{BGAj}, s) - \sum_{k=1}^{N_{Decap}} \left[ \frac{(R_p + sL_p)}{\left( sL_{ESLk} + R_{ESRk} + \frac{1}{sC_{Decap}} \right)} V_{p-Decapk}(s) \right] \cdot G(x, y, x_{Decapk}, y_{Decapk}, s), \quad (5.8)$$

where  $G(x, y, \xi, \eta, s)$  is the Green's function of a Helmholtz equation with boundary condition of the second kind. However, in (5.7) and (5.8), the voltages associated with I/O pads and C4 bumps ( $V_{c-C4j}$  &  $V_{p-C4j}$ ), BGA ( $V_{BGAj}$ ) and decaps ( $V_{Decapk}$ ) are unknowns. However, those voltages should also be the solutions of (5.7) and (5.8), and those voltages associated with the two PDEs can be put on the right-hand sides of (5.7) and (5.8). If there are  $N_{C4}$  C4 bumps,  $N_{BGA}$  BGA bumps, and  $N_{Decap}$  decaps, an equation set with  $(2N_{C4} + N_{BGA} + N_{Decap})$  equations and  $(2N_{C4} + N_{BGA} + N_{Decap})$  unknowns needs to be



solved. Consequently, the frequency characteristics of the power supply noise for the chip and package can be solved analytically from (5.7) and (5.8).

The transient noise in the time domain can be obtained by performing an inverse Laplace transform on  $V_c(x,y,s)$  and  $V_p(x,y,s)$ . The peak noise for each node can also be identified by adding up the transient noises of power and ground grids or power and ground planes.

### 5.3 Model Validation by SPICE

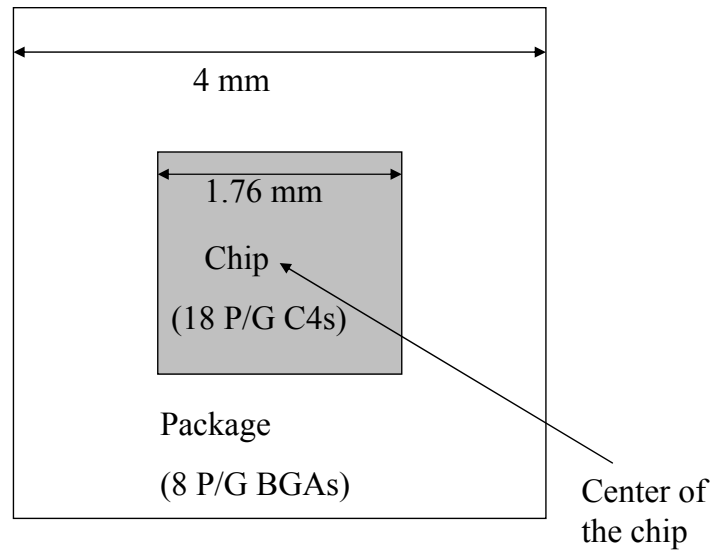


Figure 5.5 Setup for the comparison between SPICE simulations and compact physical models

A comparison between the compact physical models and the results of SPICE simulations is first performed. As SPICE can not handle large networks, a smaller size problem is chosen to validate the accuracy of the compact physical models, as shown in Figure 5.5. The study is performed for a 1.76 mm x 1.76mm chip with 18 C4 bumps (3x3 array for power and 3x3 array for ground) and a 4 mm x 4 mm package with 8 BGA bumps (2x2 array for power and 2x2 array for ground). This setup can be 100 times

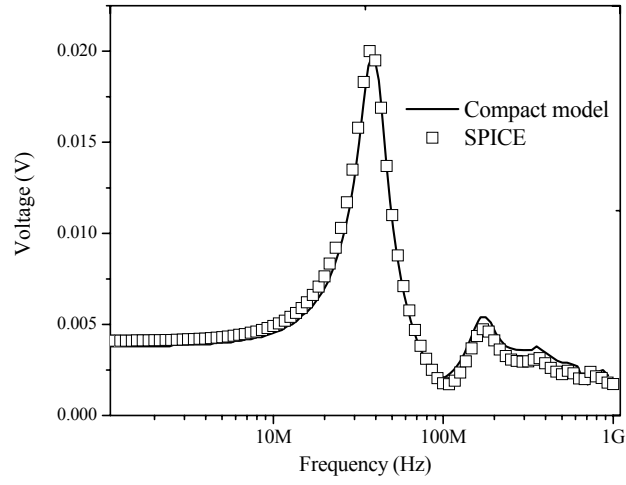
smaller than regular sized chip and package. Empirical values are used for key parameters, as shown in Table 5.1 [9, 61].

Table 5.1 Parameters for the comparison between SPICE and compact physical models

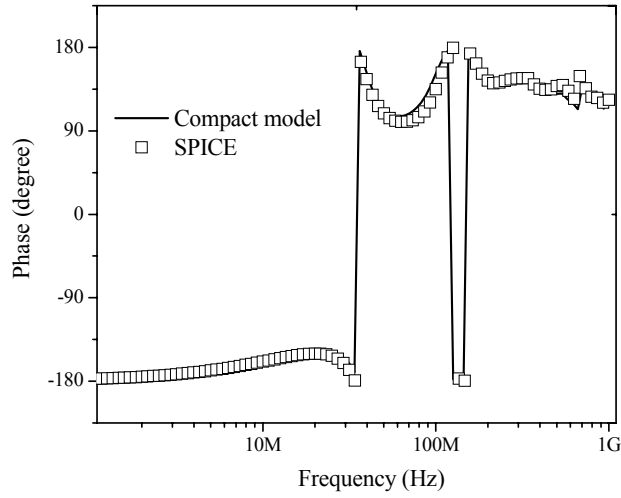
<i>Symbol</i>	<i>Value</i>	<i>unit</i>
$R_c$	0.3	$\Omega$
$R_p$	0.05	$\Omega$ per unit area
$L_p$	0.2	nH per unit area
$C_p$	2	nF/cm <sup>2</sup>
$C_c$	530	nF/cm <sup>2</sup>
$L_{c4}$	0.01	nH
$R_{c4}$	0.002	$\Omega$
$L_{BGA}$	0.4	nH
$R_{BGA}$	0.01	$\Omega$
$J_c$	10	A/cm <sup>2</sup>

It is assumed that there are 50 power/ground wires between two power/ground C4 bumps, and each segment is modeled by lumped RC circuit in SPICE simulations. The package is also discretized based on the same fineness as on-chip grids, and each package segment is modeled by lumped RLC circuit in SPICE simulations. Also it is assumed that the whole chip is switching with a current density of 10 A/cm<sup>2</sup>.

The comparison between the frequency domain results of new models and SPICE simulations for the noise at the center of the chip is shown in Figure 5.6. It can be observed from Figure 5.6 that the results from compact models fit SPICE simulation results very well.



(a)



(b)

Figure 5.6 Frequency domain response at the center of the chip: (a) Magnitude; (b) Phase

#### 5.4 Model Validation by SPEED2000

It is of great importance to evaluate the efficiency of the models to be used in real industry designs. These models are also validated using the commercialized power integrity validation tool SPEED2000 [22]. This analysis is performed for a ceramic package module with 15 plane layers designed by IBM. As shown in Figure 5.7, a 13.2 mm x 13.2 mm chip is put on the top of a 33 mm x 33 mm package, and the chip is

surrounded by sixteen 250 nF package-level decaps. The schematic view of the package in SPEED2000 is shown in Figure 5.8.

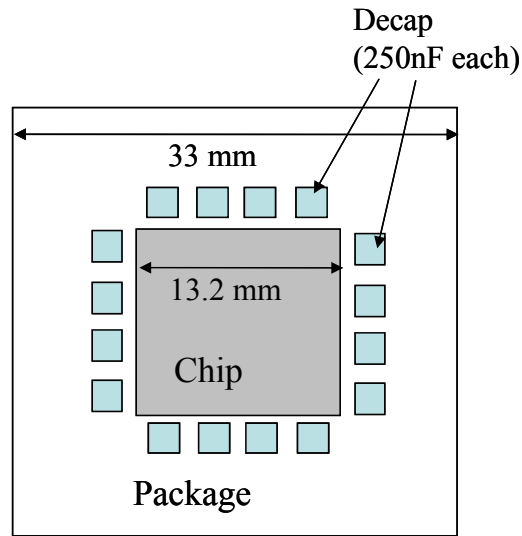


Figure 5.7 Configuration of the IBM ceramic package module

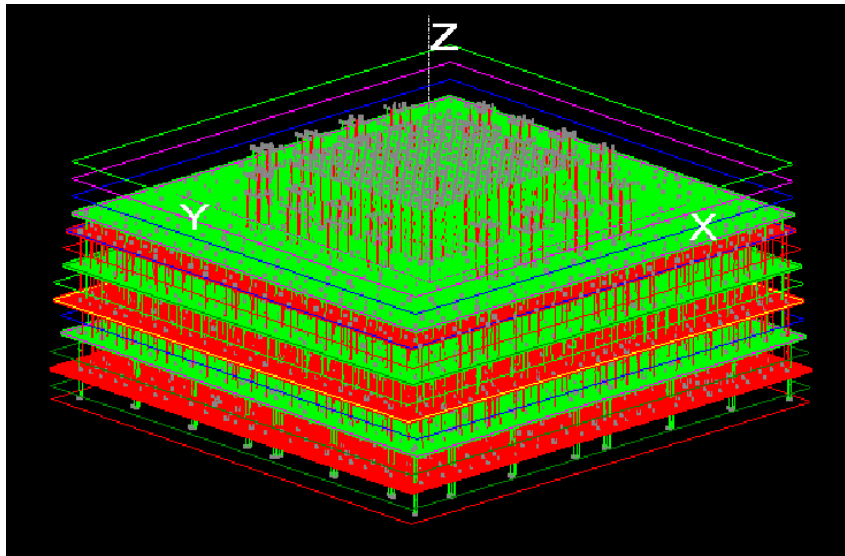


Figure 5.8 Schematic view of the package module in SPEED2000

To speed up the simulation process of the compact physical model yet not lose much accuracy, the package is divided into 10x10 cells and the chip is divided into 12x12

parts, as shown in Figure 5.9. In each cell of the chip and package, all the C4 and BGA bumps are snapped together as one element. In this case, it is also assumed that the whole chip is switching with current density of  $10 \text{ A/cm}^2$ .

Figure 5.10 shows the noise waveforms at the center of the chip and one decap location in the package as indicated in Figure 5.9. The models can well predict the resonance of the first droop power supply noise and have less than 10% error in predicting the first peak of the noise waveform compared to SPEED2000. It is meaningful to have an accurate prediction for the first peak. Because the first peak is a negative peak that lowers the supply voltage, slows down the circuits and therefore limits the maximum frequency the circuits can achieve. The physical models also give a good projection for the ringing effect (10% error for the resonance frequency at the center of the chip) by considering the distributed nature of the on-chip decoupling capacitance and package-level plane inductance. In Figure 5.10, it can also be observed that the models have larger error (50%) in predicting the second peak (a positive peak). This is because an assumption, that the package planes are uniform, is made to derive the models. However, in a real design, there are over tens of thousands of vias in the package, which play important roles in deciding the resistance of the planes, and the non-uniformity of the vias leads to the difficulties to accurately calculate the damping factor for the noise, and influence the accuracy of the later peaks. The physical models derived in this chapter need around 30 minutes to finish the simulation while SPEED2000 takes 9-10 hours. The compact physical models can at least have 10x speed-up.

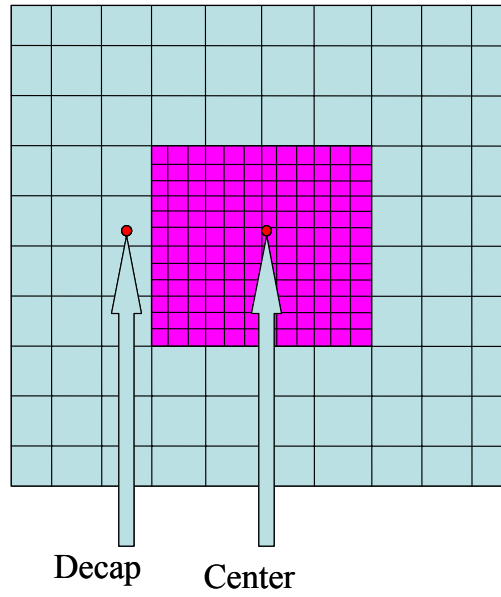


Figure 5.9 Divisions of the chip and package

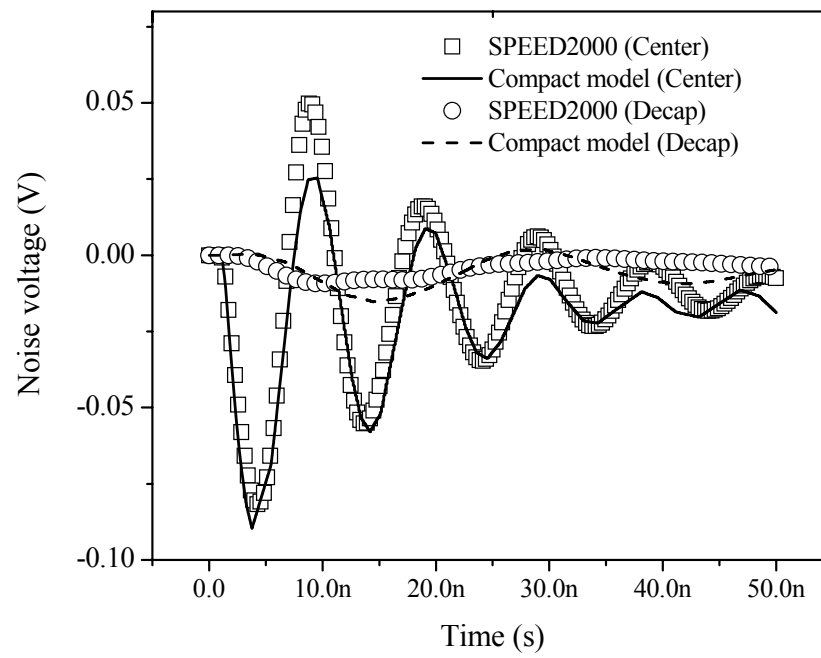


Figure 5.10 Waveforms for the two locations as shown in Figure 5.9, the center of the chip and a decap location in the package

## 5.5 Conclusion

For the first time, compact physical models are derived in this chapter, which allow designers to make a fast recognition of the power supply noise at chip and package levels. The models are able to incorporate the distributed nature of on-chip power grids and package power/ground planes. Designers can perform chip/package co-design for power distribution networks and tradeoff multiple design considerations such as metal resource allocation (on-chip power/ground grid and package-level power/ground plane), decap insertion (amount of decaps on the chip and in the package) and I/O allocation (sizes and number of C4 and BGA bumps) during early stages of design. The models are validated by commercial tools SPICE and SPEED2000. A ceramic package designed by IBM is used, and it is validated that there is less than 10% difference between the model predictions and the results of commercial tool SPEED2000 when predicting the first negative peak noise value and the time of noise occurrence. The new models also have at least 10x speed-up compared to SPEED2000 in early stage designs.

## **CHAPTER 6: COMPACT PHYSICAL MODELING TO MINIMIZE ENERGY-PER-BIT FOR ON-BOARD LC TRANSMISSION LINES UNDER NOISE CONDITIONS**

### **6.1 Introduction**

As stated in Chapter 1, chips become more and more power-limited in the GSI era [5, 47]. At the same time, the demand for I/O bandwidth increases due to the increases in the speed and integration level of chips, and hence the power dissipation in I/O drivers becomes a major issue. Low-swing signaling is suggested as a viable technique for lowering the power dissipation [40, 41]. It is desirable to send a certain amount of information dissipating the minimum possible energy. Thus, energy-per-bit, the energy consumed to transfer a single bit data, can be considered as a metric for comparing the energy efficiency of interconnects.

In a noise-free system, the energy-per-bit of a noise free channel decreases monotonically as the signal voltage swing decreases. Due to the receiver sensitivity and some noise sources independent to the signal swing, however, lowering signal swing requires larger bit duration such that a signal can reach a level large enough to fight against the noise. This increase in bit duration, however, increases the energy-per-bit. Thus, there is an optimal signal swing at which the energy-per-bit is minimized. In [41], the author has presented the work about balancing the signal swing and circuit overheads to reduce the total power consumption for on-chip interconnect systems. This chapter, for the first time, illustrates the trade-off between the signal swing and bit duration, and for a



given noise condition, calculates the theoretical value for the minimum energy-per-bit and optimal voltage swing for on-board LC transmission lines.

Tuning down the signal voltage swing can be used to save energy for on-board interconnects. The maximum energy saving and area overhead for this technique are also derived. By using a proper power supply scheme, the energy-per-bit can be decreased by 2.5x while consuming about 1.5x more on-board wiring area.

## 6.2 Noise Aware Bit-rate Limit

The step response of a transmission line [63] is

$$f(t) = \begin{cases} \operatorname{erfc} \sqrt{\frac{\beta}{t - l\sqrt{LC}}} & \text{for } t \geq l\sqrt{LC} \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

where  $\operatorname{erfc}()$  is the complementary error function,  $l$  is the wire length,  $L$  and  $C$  are external inductance and capacitance, respectively, and  $\beta$  (with a dimension of second) is a normalized unit related to the dimensional and material properties of the wire. And hence the bit-rate limit of an LC transmission line  $b$  (with a dimension of 1/s) has the form [64]

$$b \leq \frac{1}{k\beta} = \frac{1}{k} \left( \frac{16\sigma Z_0^2}{\mu_0 \mu_r} \right) \left( \frac{p}{l} \right)^2, \quad (6.2)$$

where  $\mu_0$  and  $\mu_r$  are the permeability of vacuum and the relative permeability of the conductor, respectively,  $\sigma$  is the conductivity of metal, and  $Z_0$  denotes the characteristic impedance. This result indicates that the bit-rate limit is proportional to  $(p/l)^2$ , the squared ratio of the perimeter to the length of the conductor. The choice of  $k$  determines the eye-opening dimension, and the calculation of  $k$  is discussed later in this section.

In [38] the worst case noise analysis method is presented, and if unbounded thermal, flicker and shot noises are ignored, the bounded noise sources are classified into

two categories: the proportional component, which is composed of the noise sources that are proportional to the signal swing such as crosstalk and signal induced power supply noise, and the independent component, which represents the noise sources that are independent of the signal swing such as signal unrelated power supply noise. Receiver sensitivity is an input offset voltage needed to generate a full swing output. Therefore, it is reasonable to attribute receiver sensitivity into the category of the independent noise. Consequently, the overall noise budget  $V_N$  is obtained by simply summing them up,

$$V_N = K_N V_S + V_{IN}, \quad (6.3)$$

where  $V_S$  is the signal swing, and  $K_N$  is the dependent noise coefficient and  $V_{IN}$  is the independent noise voltage. In [38], typical values of  $K_N$  and  $V_{IN}$  are listed for board-level interconnections as  $K_N = 0.25$ , and  $V_{IN} = 0.05V_{DD}$ , where  $V_{DD}$  denotes supply voltage, namely, the full swing voltage. If taking the noise budget into account, the eye-opening dimension will be determined by the noise budget when reduced signal swings are utilized. Because the step input response is a complementary error function ( $erfc()$ ) [63], the noise aware bit-rate can be obtained as (6.5) by selecting  $k$  described by (6.4)

$$k = \left( \frac{1}{erfc^{-1}\left(0.5 + \frac{V_N}{V_S}\right)} \right)^2. \quad (6.4)$$

Substituting equations (6.4) and (6.3) in (6.2) yields

$$b \leq \left[ erfc^{-1}\left(0.5 + K_N + \frac{V_{IN}}{V_S}\right) \right]^2 \left( \frac{16\sigma Z_0^2}{\mu_0 \mu_r} \right) \left( \frac{p}{l} \right)^2. \quad (6.5)$$

The inverse error function in (6.5) is not an easy term to manipulate in mathematics, and a polynomial series approximation is therefore introduced. It is found

that a first order polynomial approximation [65] has maximum error less than 5% and is accurate enough when  $\left(0.5 + K_N + \frac{V_{IN}}{V_S}\right)$  is between 0.5 and 1, which is the range of the term in real circuit parameter implementation based on the empirical values of  $K_N$  and  $V_{IN}$  given in [38]. Thus, the revised noise aware bit-rate limit can be obtained,

$$b \leq \left[ \frac{\sqrt{\pi}}{2} \left( 1 - \left( 0.5 + K_N + \frac{V_{IN}}{V_S} \right) \right) \right]^2 \left( \frac{16\sigma Z_0^2}{\mu_0 \mu_r} \right) \left( \frac{p}{l} \right)^2. \quad (6.6)$$

### 6.3 Minimized Energy-Per-Bit

To minimize the energy-per-bit, an interconnect needs to work at its bit-rate limit so that the bit duration is minimized. Energy-per-bit can therefore be written as

$$E_{bit} = \frac{P}{b}. \quad (6.7)$$

where power,  $P$ , decreases monotonically with scaling down the voltage swing. Bit rate limit, however, decreases as the voltage swing scales because of independent noise sources. Generally, power consumption of drivers is dominant and that of receivers can be ignored [66].

To realize reduced signal swing techniques, the interconnection system is constructed as shown in Figure 6.1 [41]. The wire is ended with matched impedance with the value of wire's characteristic impedance  $Z_0$  (often set to be 50  $\Omega$ ). The driver uses reduced supply voltage,  $V_{SSD} = V_{DD}/2 - V_S/2$  and  $V_{DDD} = V_{DD}/2 + V_S/2$ , and the DC voltage of the wire is set to  $V_{DD}/2$  such that the receiver and the driver can work at a proper point.

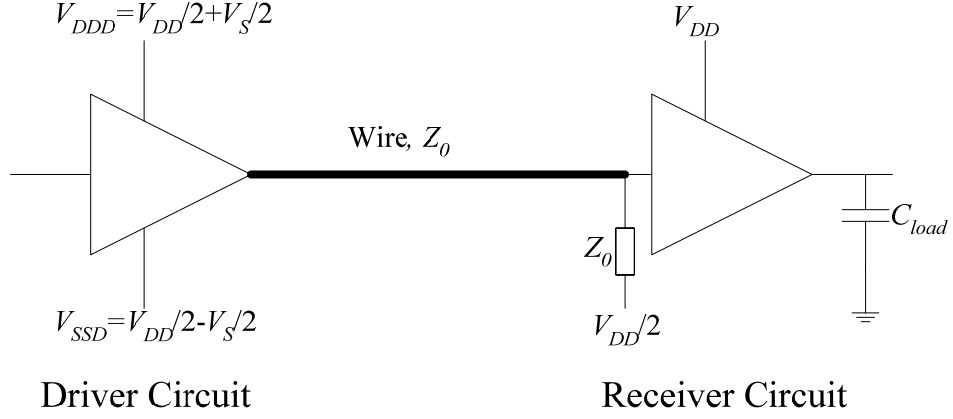


Figure 6.1 Power supply scheme for on-board interconnection

A power supply scheme [41] can be adopted based on the circuit structure in Figure 6.1. Two supplies,  $V_S/2$  and  $-V_S/2$ , each connected to the common stable node  $V_{DD}/2$ . Either of the two supplies delivers the current with amount of  $V_S/2Z_0$  or  $-V_S/2Z_0$  to the receiver. The power consumption is

$$P = \frac{V_S^2}{4Z_0}. \quad (6.8)$$

Equation (6.8) needs to be doubled, if differential signaling is used. Using (6.6), (6.7), and (6.8) obtains

$$E_{bit} \geq \frac{V_S^2}{\left[ \sqrt{\pi} \left( 1 - (0.5 + K_N + \frac{V_{IN}}{V_S}) \right) \right]^2 \left( \frac{16\sigma Z_0^3}{\mu_0 \mu_r} \right) \left( \frac{p}{l} \right)^2}. \quad (6.9)$$

letting the derivative of (6.9) equal to zero, the optimal signal swing that minimizes the energy-per-bit is calculated as

$$V_{S-opt} = \frac{4V_{IN}}{1 - 2K_N}, \quad (6.10)$$

which is independent of material and dimensional properties of the wire.

Energy-per-bit normalized to the full swing energy-per-bit is plotted in Figure 6.2. It demonstrates that a more severe noise condition results in a larger optimal signal swing and a larger minimum energy-per-bit. Below the optimal voltage value, the energy-per-bit increases sharply with the decrease of the signal swing. In the case of  $K_N=0.25$  and  $V_{IN}=0.05V_{DD}$ , the minimum normalized energy-per-bit is 0.410 with a  $0.4V_{DD}$  signal swing.

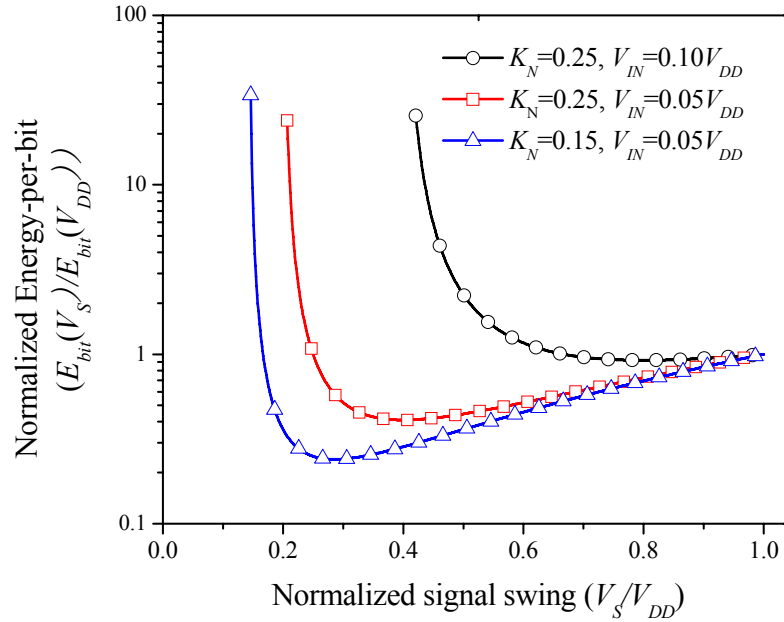


Figure 6.2 Normalized energy-per-bit vs. signal swing in different noise conditions

#### 6.4 Wiring Area Overhead

The energy saving that low-swing signaling can offer comes at the price of smaller bandwidth and hence for a given aggregate bandwidth, there is a wiring area overhead associated with low-swing signaling. To quantify this overhead, a similar approach presented in [64] is followed here.

Data flux density  $\Phi_D$  is the interconnect bandwidth per unit wire width. It is desired to have a large data flux density to transfer as many bits per second as possible using a constant wiring area in high performance systems. As shown in Figure 6.3, if data buses fully utilize a chip edge, aggregate bandwidth for each chip edge could be obtained by multiplying the maximum data flux density with chip dimension.

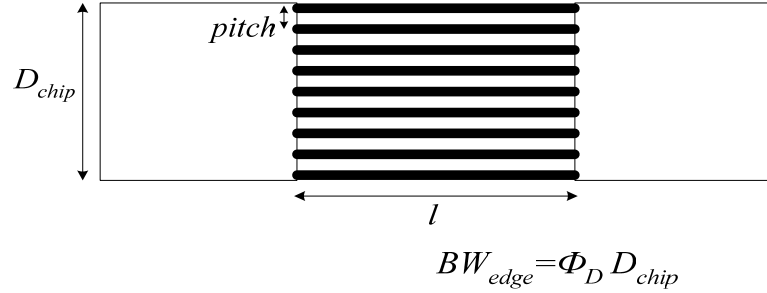


Figure 6.3 Illustration of data flux density,  $\Phi_D$  (bit/s · cm), where chip edge length  $D_{chip}$  and wire width  $w$  both have dimensions of cm

The result in [64] is based on two assumptions. First, because of the proximity effect, current flows mostly through the lower and upper regions of wires that are close to ground planes and the effective perimeter of a wire can be taken as twice wire width  $2w$ . Secondly, off-chip or on-board wire width is set to be half of the wire pitch. Therefore, the maximum data flux density is determined by the ratio of bit-rate limit to wire width and can be calculated from (6.6)

$$\Phi_D = \frac{b}{2w} \leq \left[ \frac{\sqrt{\pi}}{2} \left( 1 - \left( 0.5 + K_N + \frac{V_{IN}}{V_S} \right) \right) \right]^2 \left( \frac{64\sigma Z_0^2}{\mu_0 \mu_r} \right) \frac{w}{2l^2}. \quad (6.11)$$

To achieve the maximum data flux density, on-board wires should be made as wide as possible. The maximum frequency that a driver can switch  $f_{max}$ , however, introduces a limit on the maximum interconnect bandwidth.

$$\Phi_D \leq \frac{f_{\max}}{2w} \leq \left[ \frac{\sqrt{\pi}}{2} \left( 1 - (0.5 + K_N + \frac{V_{IN}}{V_S}) \right) \right]^2 \left( \frac{64\sigma Z_0^2}{\mu_0\mu_r} \right) \frac{w}{2l^2}. \quad (6.12)$$

Hence the optimized wire width is the width at which interconnect bandwidth becomes equal to the maximum switching frequency of drivers:

$$w_{opt} = \frac{l}{\left[ \frac{\sqrt{\pi}}{2} \left( 1 - (0.5 + K_N + \frac{V_{IN}}{V_S}) \right) \right]} \sqrt{\frac{f_{\max}}{\left( \frac{64\sigma Z_0^2}{\mu_0\mu_r} \right)}}. \quad (6.13)$$

It is assumed that the maximum off-chip driver speed  $f_{\max}$  is equal to the ITRS projections for the chip-to-board clock frequency [47]. By (6.12) and (6.13), the maximum data flux density on-board wires can present is

$$\Phi_D \leq \frac{\left[ \frac{\sqrt{\pi}}{2} \left( 1 - (0.5 + K_N + \frac{V_{IN}}{V_S}) \right) \right]}{2l} \sqrt{\left( \frac{64\sigma Z_0^2}{\mu_0\mu_r} \right) \cdot f_{\max}}. \quad (6.14)$$

From (6.14), it can be observed that larger  $V_S$  results in a larger maximum data flux density. Thus, to achieve the maximum data flux density, one needs to utilize full swing signal ( $V_S = V_{DD}$ ).

Equation (6.13) shows that as wire length decreases, the optimal wire width decreases. However, the wire width is limited by the minimum resolvable line width  $w_{min}$ , implying that a minimum length exists below which the maximum data flux density keeps constant, whose value is

$$\Phi_D = \frac{f_{\max}}{w_{min}}. \quad (6.15)$$

Figure 6.4 plots the maximum data flux density versus wire length for the 65 nm, 45 nm and 32 nm technology nodes. The value of  $f_{\max}$  and  $w_{min}$  are taken from the ITRS, and  $V_{IN}$  and  $K_N$  are set to be  $0.01V_{DD}$  and 0.25, respectively. It is noted that the critical

length, below which the maximum data flux density keeps constant, is also the point, beyond which the wire will exploit its physical limit to achieve the maximum performance. This length will decrease with the technology scaling down, 50 cm for the 65 nm node, 33 cm for the 45 nm node and 17 cm for the 32 nm node. That is to say, more and more on-board wires will be working at the region where the performance of the wires is governed by the physical limit. Figure 6.5 shows that as the noise condition exacerbates, this critical length scales down even further.

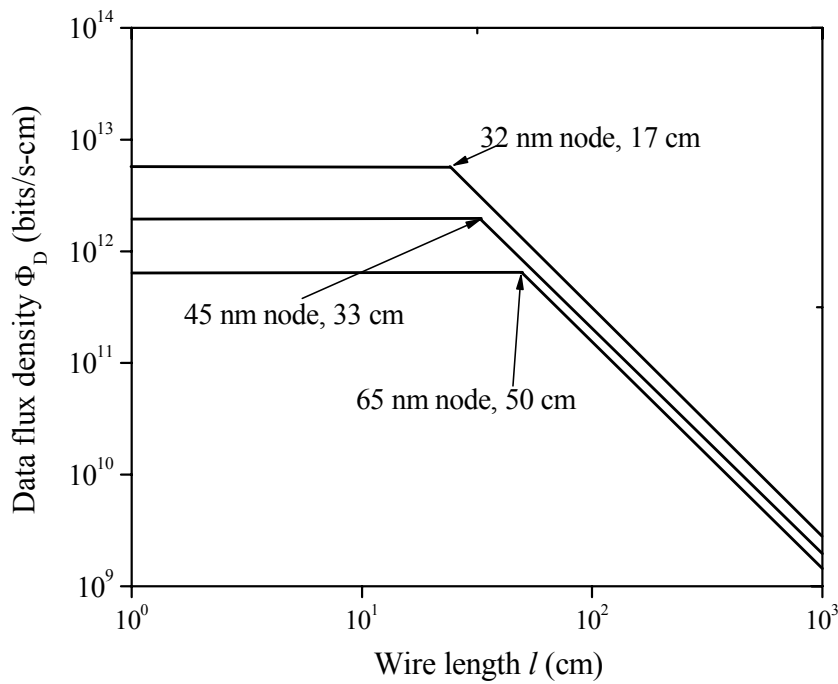


Figure 6.4 Data flux density vs. wire length marked by critical length of each technology generation

To quantify the area overhead of low-swing signaling, only interconnects longer than the critical length are considered as the largest over-head corresponds to those interconnects that are limited by wires but not drivers. To meet the bandwidth



requirement in the full swing case, more wiring area (for given wiring area in each layer, more layers are needed, and for given number of wiring layers, more wiring area in each layer is needed) is needed when the reduced swing technique is used.

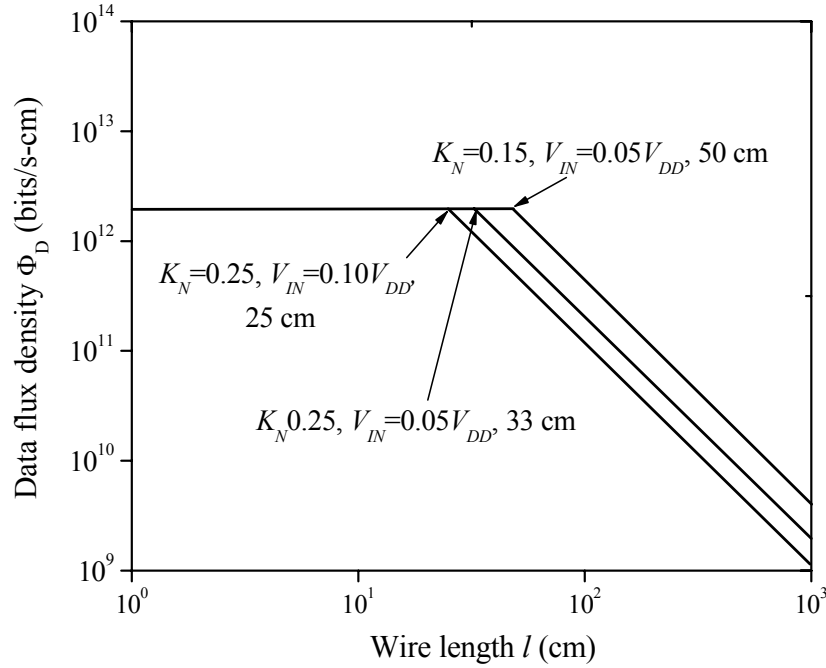


Figure 6.5 Data flux density vs. wire length marked by the critical length in different noise conditions based on the ITRS projections for the 45 nm node

Energy Benefit Factor ( $EBF$ ) is defined as how many times the energy-per-bit is reduced, compared to the full swing case, when the low swing technique is adopted, therefore

$$EBF = \frac{E_{bit}(V_{DD})}{E_{bit}(V_S)}. \quad (6.16)$$

Area Overhead Factor ( $AOF$ ) is defined as how much more wiring area is used compared to the full swing case when the low swing technique is used for a constant aggregate bandwidth

$$AOF = \frac{b(V_{DD}) - b(V_S)}{b(V_S)}. \quad (6.17)$$

The trade-off between energy saving (*EBF*) and wiring area overhead (*AOF*) is shown in Figure 6.6. The maximum improvement in energy efficiency is 2.44× and requires 1.56× larger wiring area.

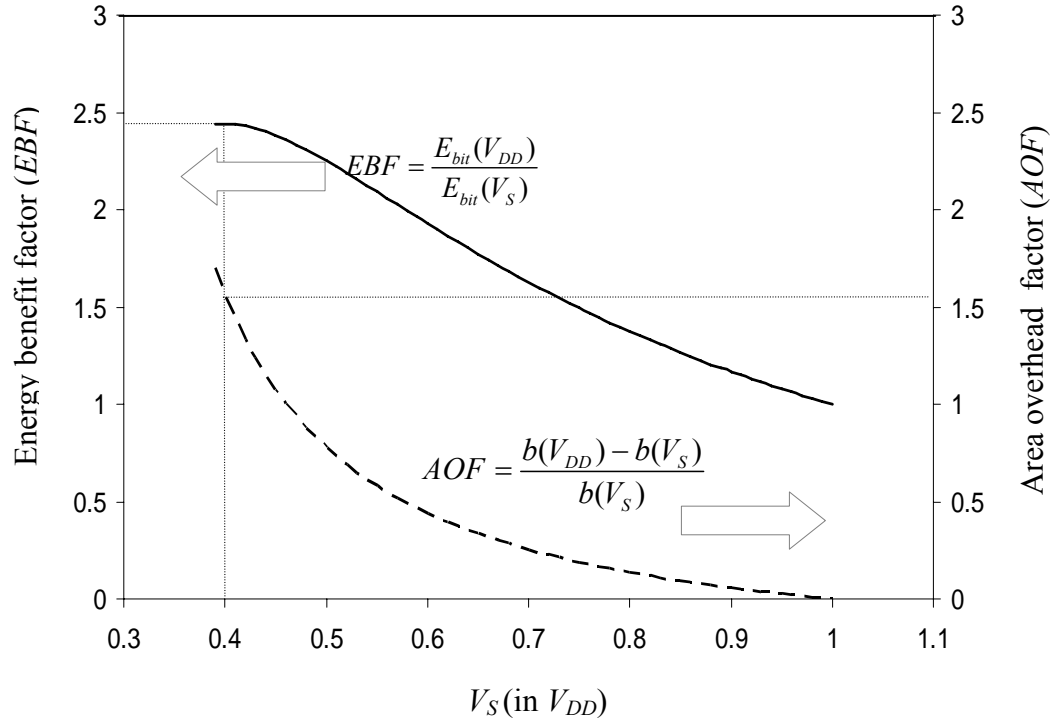


Figure 6.6 Energy benefit factor and Area overhead factor vs. signal swing when  $K_N=0.25$ ,  $V_{IN}=0.05V_{DD}$ .

The power supply scheme used in this analysis relies on using a secondary supply voltage, which needs another voltage regulator. Thus, this technique is well applicable to the systems with less extra overhead on the secondary power supply, such as a dual supply chip, where the secondary supply is already available for the core logic part. (e.g. for a  $0.6V_{DD}$  secondary power supply, 1.87× energy benefit can be obtained).

## 6.5 Conclusion

In this chapter, the minimum energy required to transfer one bit of data through chip-to-chip interconnects for a given noise condition is calculated. It is shown that because of the noise sources independent of the signal swing as well as the sensitivity of receivers, the bit duration has to increase as the signal swing increases. This increase in the bit duration increases the energy-per-bit and hence negates the power saving that lowering the signal amplitude offers. It is shown that there is an optimal voltage swing that is independent of interconnect length and cross-sectional dimensions and is determined by the noise condition only. For a typical noise condition for on-board single-ended interconnects ( $K_N=0.25$ ,  $V_{IN}=0.10V_{DD}$ ), the optimal signal swing is  $0.4V_{DD}$  that leads to  $2.44\times$  energy benefit at the cost of  $1.56\times$  increase in wiring area for a constant aggregate bandwidth compared to the full swing case. To achieve a considerable energy benefit from the low swing techniques, it is critical to increase the noise immunity of the channels. Differential signaling, for instance, is less vulnerable to noise and is hence a good candidate for low-swing interconnection [38]. This chapter contains a list of the key results and contributions of the research presented in this thesis

## **CHAPTER 7: CONCLUSIONS AND FUTURE WORKS**

In this chapter, the key conclusions of this dissertation are summarized and possible extensions of this dissertation are discussed. These extensions include: 1) Power supply noise analysis for multicore microprocessors; 2) Optimizations of power and fluidic I/O and TSV networks in 3-D chip stacks.

### **7.1 Conclusions of Dissertation**

The main objective of this thesis is to derive a set of compact physical models addressing power integrity issues in high performance GSI and 3-D system designs. These compact physical models facilitate quick assessment of the first droop power supply noise and the noise's impact on the high-speed link performance without extended dedicated simulations. The models can also help designers gain valuable physical insights into the complicated power delivery system and tradeoffs among various important chip and package design parameters during the early stages of design.

The main contributions of this dissertation are as follows:

1. Novel blockwise compact physical models are derived to describe the frequency characteristics, time domain transients, and the worst case peak noise value for the first droop of power supply noise for the power hungry blocks in GSI chips. The models support the power grid analysis not only for state-of-the-art design but also for the scaling trends of future technology nodes. The models display high accuracy, and there is less than 4% discrepancy between the models and the results of SPICE simulations.

2. New compact physical models are introduced to predict the power supply noise with the consideration of hot spots. The models specifically address the non-uniformity problem for the power density distribution brought by hot spots. The models give less than 1% error compared against the results of SPICE simulations.
3. Efficient compact physical models are developed to incorporate the impact of 3-D integration on the power supply noise. The models enable designers to identify the challenges in the power delivery network design in 3-D chip stacks because of the larger supply current and longer power delivery path brought by TSVs and Micro-bumps in 3-D integrated systems. The models have less than 4% error compared to the results of SPICE simulations. Based on the models, design guidelines are also proposed to address the power integrity problem for 3-D integration.
4. For the first time compact physical models are built to incorporate the distributed nature of on-chip power/ground grids and package-level power/ground planes on the same mathematical platform. The models are validated by commercial tools SPICE and SPEED2000. A ceramic package designed by IBM is assumed, and it is validated that there is less than 10% difference between the model predictions and the commercial tool SPEED2000. The new models also have at least 10x speed-up.
5. A new model is introduced to investigate the minimum energy required to transfer a single bit of data through on-board interconnects for a given noise condition. It is discovered that the trade-off between signal swing and bit duration leads to an

optimal signal swing at which the energy-per-bit is minimized. The model can be used to calculate the theoretical value for the minimum energy-per-bit and optimal voltage swing. The maximum energy saving and area overhead for the low swing technique are also derived.

## **7.2 Power Supply Noise Analysis for Multicore Microprocessors**

A multicore microprocessor implements several microprocessing cores on a single die. Each core independently implements processor operations such as superscalar execution, pipelining, and multithreading [67]. The processors also share the same interconnect to the rest of the system. Noise mitigation is an increasingly difficult problem as more and more cores are integrated in the same die where circuitry becomes denser and power demand increases. Switching events in the core logic can cause large transient current demand. This transient current will cause a large power supply noise which flows through the shared power distribution network and adversely impacts the performance of other cores.

A good understanding of how the power distribution network topology influences the power supply noise is of great significance. Using models derived in this dissertation enables quick predictions of the noise levels for certain cores. Based on the noise information, different topologies for the power distribution network can be selected. Fully connected on-chip grid and package level power and ground planes can allow power hungry cores to use the decaps associated with other non-switching cores. However, fully connected networks permit noise propagation among different cores, and the noise condition of a switching core will be further exacerbated if all its adjacent cores are switching. Separate power distribution networks can be adopted to isolate two blocks

with very high switching activities, but decap resources will be limited in this case. Therefore, hybrid networks can be a good option and the optimization between the power distribution network topology, core switching patterns, and decap allocation can be a revealing extension of this dissertation.

### **7.3 Optimizations of the Power and Fluidic I/O and TSV Networks in 3-D Chip**

#### **Stacks**

3-D microsystems can provide enormous advantages in achieving multi-functional integration, improving system speed and reducing power consumption for future generations of ICs. However, stacking multiple dice or wafers may result in severe thermal, power integrity and connectivity problems [68]. It is very critical to resolve the challenges of realizing a low cost chip-scale integrated I/O interconnect network that is capable of addressing the heat removal, I/O bandwidth, and power delivery requirements for 3-D integration systems. To identify the interactions and tradeoffs between power, heat removal and connectivity, it is of great significance to extend current models to address the optimization of the electrical and fluidic I/O and TSV networks at the same time. Thus, the ultimate goal of the proposed task is to build a simulator and optimizer to optimize among power delivery schemes, cooling technologies and I/Os and TSV density accurately, efficiently and quickly.

## APPENDIX A: DERIVATION FOR PARTIAL DIFFERENTIAL EQUATION (2.1)

This section gives the rigorous derivation for PDE (2.1). The derivation starts from the calculation of the voltage of an *anisotropic* grid with different resistances in the  $x$  and  $y$  directions, as shown in Figure A.1.

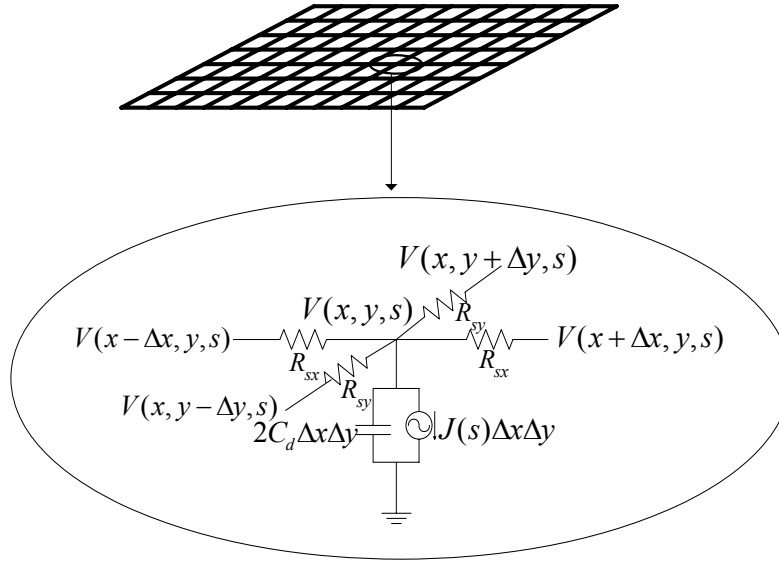


Figure A.1 Circuit model for a node for a single grid

Each node of the grid is connected to four neighboring nodes. The wires between two nodes are modeled by lumped resistances  $R_{sx}$  and  $R_{sy}$  for the  $x$  and  $y$  directions, respectively. Symbol  $J(s)$  is the switching current density in Laplace domain, and  $C_d$  denotes the amount of decoupling capacitance per unit area. Symbols  $\Delta x$  and  $\Delta y$  represent the distances between two nodes at the same wiring level for  $x$  and  $y$  directions, respectively. The voltage at node located at  $(x, y)$  can be calculated from the voltages of the four neighboring nodes located at  $(x+\Delta x, y)$ ,  $(x-\Delta x, y)$ ,  $(x, y+\Delta y)$ , and  $(x, y-\Delta y)$ . Based



on Kirchoff's current law [28], the current flowing into the node is equal to the current flowing out of the node, and

$$\frac{V(x, y, s) - V(x + \Delta x, y, s)}{R_{sx}} + \frac{V(x, y, s) - V(x, y + \Delta y, s)}{R_{sy}} + \frac{V(x, y, s) - V(x - \Delta x, y, s)}{R_{sx}} + \frac{V(x, y, s) - V(x, y - \Delta y, s)}{R_{sy}} = -J(s)\Delta x\Delta y - V(x, y, s) \cdot 2sC_d \cdot \Delta x\Delta y \quad (\text{A.1})$$

The left-hand side (LHS) of (A.1) represents the current flowing towards the four neighboring nodes, and the right-hand side (RHS) is the current associated with the current source and decap.

To facilitate the derivation process, the sheet resistances of a grid are introduced. The sheet resistance is a measure of resistance of thin films that have a uniform thickness. It is assumed that  $R_x$  and  $R_y$  are the sheet resistances (the units are  $\Omega$  or  $\Omega/\square$  [69]) for the grid in  $x$  and  $y$  directions, respectively. The sheet resistances can be calculated from the segment resistances as [28]:

$$R_x = R_{sx} \times \frac{l_{segy}}{l_{segx}} = \frac{\rho}{T_x} \frac{l_{segy}}{W_x} \times \frac{l_{segx}}{l_{segx}} = \frac{\rho}{T_x} \frac{l_{segy}}{W_x}, \quad (\text{A.2})$$

and

$$R_y = R_{sy} \times \frac{l_{segx}}{l_{segy}} = \frac{\rho}{T_y} \frac{l_{segy}}{W_y} \times \frac{l_{segx}}{l_{segy}} = \frac{\rho}{T_y} \frac{l_{segx}}{W_y}, \quad (\text{A.3})$$

where  $l_{segx}$  and  $l_{segy}$  are the length of the wire segments in  $x$  and  $y$  directions respectively, and they are equal to  $\Delta x$  and  $\Delta y$ , respectively. Symbols  $W_x$  and  $W_y$  are the widths of the segments in  $x$  and  $y$  directions,  $T_x$  and  $T_y$  are the thicknesses of the segments in  $x$  and  $y$  directions, and  $\rho$  is the resistivity of the grid metal. Thus, (A.1) can be rewritten to

$$\begin{aligned} & \frac{V(x, y, s) - V(x + \Delta x, y, s)}{R_x \frac{\Delta x}{\Delta y}} + \frac{V(x, y, s) - V(x, y + \Delta y, s)}{R_y \frac{\Delta y}{\Delta x}} + \frac{V(x, y, s) - V(x - \Delta x, y, s)}{R_x \frac{\Delta x}{\Delta y}} \\ & + \frac{V(x, y, s) - V(x, y - \Delta y, s)}{R_y \frac{\Delta y}{\Delta x}} = -J(s)\Delta x\Delta y - V(x, y, s) \cdot 2sC_d \cdot \Delta x\Delta y \end{aligned} \quad (\text{A.4})$$

If both the RHS and LHS of (A.4) are multiplied with a factor of  $1/(\Delta x\Delta y)$ , (A.4) is transformed into

$$\begin{aligned} & \frac{1}{R_x} \frac{V(x, y, s) - V(x + \Delta x, y, s)}{\Delta x^2} + \frac{1}{R_y} \frac{V(x, y, s) - V(x, y + \Delta y, s)}{\Delta y^2} + \frac{1}{R_x} \frac{V(x, y, s) - V(x - \Delta x, y, s)}{\Delta x^2} \\ & + \frac{1}{R_y} \frac{V(x, y, s) - V(x, y - \Delta y, s)}{\Delta y^2} = -J(s) - V(x, y, s) \cdot 2sC_d \end{aligned} \quad (\text{A.5})$$

The number of segments of the grid is usually large; therefore, the grid can be modeled as a continuous planar surface. The segment lengths  $\Delta x$  and  $\Delta y$  are very small, and using the finite element method (FEM), the partial derivative in  $x$  direction of the voltages at the locations  $(x + \Delta x, y)$  and  $(x - \Delta x, y)$  can be approximated by [70]

$$\left. \frac{\partial V}{\partial x} \right|_{(x + \Delta x, y)} \approx \frac{V(x + \Delta x, y, s) - V(x, y, s)}{\Delta x} \quad (\text{A.6})$$

and

$$\left. \frac{\partial V}{\partial x} \right|_{(x - \Delta x, y)} \approx \frac{V(x - \Delta x, y, s) - V(x, y, s)}{\Delta x}, \quad (\text{A.7})$$

respectively.

Using (A.6) and (A.7), the second partial derivative in  $x$  direction of the voltage at the location  $(x, y)$  can be approximated by [70]

$$\begin{aligned}
\frac{\partial^2 V}{\partial x^2} \big|_{(x,y)} &\approx \frac{\frac{\partial V}{\partial x} \big|_{(x+\Delta x,y)} - \frac{\partial V}{\partial x} \big|_{(x-\Delta x,y)}}{\Delta x} \\
&\approx \frac{\frac{V(x+\Delta x,y,s) - V(x,y,s)}{\Delta x} - \frac{V(x+\Delta x,y,s) - V(x,y,s)}{\Delta x}}{\Delta x} \quad (A.8) \\
&= -\frac{V(x,y,s) - V(x+\Delta x,y,s)}{\Delta x^2} - \frac{V(x,y,s) - V(x-\Delta x,y,s)}{\Delta x^2}
\end{aligned}$$

Similiarly, the second partial derivative in  $y$  direction of the voltage at the location  $(x,y)$  can be approximated by

$$\begin{aligned}
\frac{\partial^2 V}{\partial y^2} \big|_{(x,y)} &\approx \frac{\frac{\partial V}{\partial y} \big|_{(x,y+\Delta y)} - \frac{\partial V}{\partial y} \big|_{(x,y-\Delta y)}}{\Delta y} \\
&\approx \frac{\frac{V(x,y+\Delta y,s) - V(x,y,s)}{\Delta y} - \frac{V(x,y-\Delta y,s) - V(x,y,s)}{\Delta y}}{\Delta y} \quad (A.9) \\
&= -\frac{V(x,y,s) - V(x,y+\Delta y,s)}{\Delta y^2} - \frac{V(x,y,s) - V(x,y-\Delta y,s)}{\Delta y^2}
\end{aligned}$$

It is noted that each of above terms is exactly shown in (A.5), using the relationships described by (A.8) and (A.9), for any node  $(x,y)$  on the grid, the difference equation (A.5) can be approximated by a partial differential equation (PDE) [28, 70]

$$\frac{1}{R_x} \frac{\partial^2 V(x,y,s)}{\partial x^2} + \frac{1}{R_y} \frac{\partial^2 V(x,y,s)}{\partial y^2} = J(s) + V(x,y,s) \cdot 2sC_d \quad (A.10)$$

Equation (A.10) can be used to calculate the voltage of an *anisotropic* grid. For *isotropic* grids as discussed in this dissertation,  $R_x=R_y$  and  $l_{segx}=l_{segy}$  ( $\Delta x = \Delta y$ ). From (A.2) and (A.3), it is known that the sheet resistances are equal to the segment resistances in value,

$$R_x = R_y = R_{sx} = R_{sy} = R_s, \quad (A.11)$$

where  $R_s$  is the segment resistance for an isotropic grid in both  $x$  and  $y$  directions. Thus, using (A.11), the original form of (2.1) can be derived from (A.10) as

$$\frac{\partial^2 V(x, y, s)}{\partial x^2} + \frac{\partial^2 V(x, y, s)}{\partial y^2} = R_s J(s) + V(x, y, s) \cdot 2sR_s C_d. \quad (\text{A.12})$$

The derivations for the source function and boundary condition of (2.1) have already been discussed in Section 2.2.

## APPENDIX B: DERIVATION TO OBTAIN THE SOLUTION FOR

(2.5)

This section shows the derivation steps from (2.5) to (2.6).

Equation (2.5) can be transformed into a pure Helmholtz equation [46]. If a new function  $u(x,y,s)$  is introduced and defined as

$$u(x, y, s) = V(x, y, s) + \frac{J(s)}{2sC_d}, \quad (\text{B.1})$$

then

$$V(x, y, s) = u(x, y, s) - \frac{J(s)}{2sC_d}. \quad (\text{B.2})$$

Using  $\left(u(x, y, s) - \frac{J(s)}{2sC_d}\right)$  to represent  $V(x,y,s)$ , the left-hand side (LHS) of (2.5) can be written as

$$\nabla^2 V(x, y, s) = \nabla^2 \left( u(x, y, s) - \frac{J(s)}{2sC_d} \right) = \nabla^2 u(x, y, s); \quad (\text{B.3})$$

The right-hand side (RHS) of (2.5) can be written as

$$\begin{aligned} & R_s J(s) + 2V(x, y, s) \cdot sR_s C_d - R_s \frac{V(\alpha D_{pad}, 0, s)}{4sL_p} \delta(x)\delta(y) \\ &= R_s J(s) + 2 \cdot \left( u(x, y, s) - \frac{J(s)}{2sC_d} \right) \cdot sR_s C_d - R_s \frac{\left( u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right)}{4sL_p} \delta(x)\delta(y) \quad (\text{B.4}) \\ &= R_s J(s) + 2 \cdot u(x, y, s) \cdot sR_s C_d - R_s J(s) - \frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \delta(x)\delta(y) \\ &= u(x, y, s) \cdot 2sR_s C_d - \frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \delta(x)\delta(y) \end{aligned}$$

Equating (B.3) and (B.4) gives a partial differential equation (PDE) of  $u(x,y,s)$ ,

$$\nabla^2 u(x, y, s) = u(x, y, s) \cdot 2sR_s C_d - \frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \delta(x) \delta(y). \quad (\text{B.5})$$

Solving for (B.5) is equivalent to solving for (2.5).

Since the partial derivatives of  $V(x, y, s)$  are equal to the partial derivatives of  $u(x, y, s)$ , as shown in (B.6),

$$\begin{aligned} \frac{\partial V(x, y, s)}{\partial x} &= \frac{\partial \left( u(x, y, s) - \frac{J(s)}{2sC_d} \right)}{\partial x} = \frac{\partial u(x, y, s)}{\partial x}, \\ \frac{\partial V(x, y, s)}{\partial y} &= \frac{\partial \left( u(x, y, s) - \frac{J(s)}{2sC_d} \right)}{\partial y} = \frac{\partial u(x, y, s)}{\partial y}, \end{aligned} \quad (\text{B.6})$$

the boundary conditions of (2.2) can be expressed as

$$\begin{aligned} \frac{\partial u(x, y, s)}{\partial x} \Big|_{y=0} &= 0, \quad \frac{\partial u(x, y, s)}{\partial x} \Big|_{y=a} = 0 \\ \frac{\partial u(x, y, s)}{\partial y} \Big|_{x=0} &= 0, \quad \frac{\partial u(x, y, s)}{\partial y} \Big|_{x=a} = 0. \end{aligned} \quad (\text{B.7})$$

Equation (B.5) is a pure Helmholtz equation with a boundary condition of the second kind.

From [46], it is known that a Helmholtz equation has a general form

$$\nabla^2 u(x, y) + \lambda u(x, y) = \Phi(x, y), \quad (\text{B.8})$$

where  $\Phi(x, y)$  is the source function and  $\lambda$  is independent of location. If this equation satisfies the boundary condition of the second kind in a square region with area of  $axa$ ,

$$\begin{aligned} \frac{\partial u(x, y)}{\partial x} \Big|_{y=0} &= f_1(y), \quad \frac{\partial u(x, y)}{\partial x} \Big|_{y=a} = f_2(y) \\ \frac{\partial u(x, y)}{\partial y} \Big|_{x=0} &= f_3(x), \quad \frac{\partial u(x, y)}{\partial y} \Big|_{x=a} = f_4(x), \end{aligned} \quad (\text{B.9})$$

then the solution of (B.8) based on this boundary condition is [46]

$$\begin{aligned}
u(x, y) = & \int_0^a \int_0^a \Phi(\xi, \eta) G(x, y, \xi, \eta) d\eta d\xi \\
& - \int_0^a f_1(\eta) G(x, y, 0, \eta) d\eta + \int_0^a f_2(\eta) G(x, y, a, \eta) d\eta \cdot \\
& - \int_0^a f_3(\xi) G(x, y, \xi, 0) d\xi + \int_0^a f_4(\xi) G(x, y, \xi, a) d\xi
\end{aligned} \tag{B.10}$$

$G(x, y, \xi, \eta)$  is the Green's function, and can be written either as a single series

$$\begin{aligned}
G(x, y, \xi, \eta) = & \frac{1}{a} \sum_{n=0}^{\infty} \frac{\varepsilon_n \cos(p_n x) \cos(p_n \xi)}{\beta_n \sinh(\beta_n a)} H_n(y, \eta) \\
= & \frac{1}{a} \sum_{m=0}^{\infty} \frac{\varepsilon_m \cos(q_m y) \cos(q_m \eta)}{\mu_m \sinh(\mu_m a)} Q_m(x, \xi) \cdot,
\end{aligned} \tag{B.11}$$

where

$$\begin{aligned}
p_n = \frac{\pi n}{a}, \quad H_n(y, \eta) = & \cosh(\beta_n \eta) \cosh[\beta_n(a - y)] \text{ for } y > \eta \\
= & \cosh(\beta_n y) \cosh[\beta_n(a - \eta)] \text{ for } \eta > y \\
q_m = \frac{\pi m}{a}, \quad Q_m(x, \xi) = & \cosh(\mu_m \xi) \cosh[\mu_m(a - x)] \text{ for } x > \xi, \\
= & \cosh(\mu_m x) \cosh[\mu_m(a - \xi)] \text{ for } \xi > x \\
\beta_n = \sqrt{p_n^2 - \lambda}, \quad \mu_m = \sqrt{q_m^2 - \lambda}, \quad \varepsilon_n = & 1 \text{ for } n = 0 \\
= & 2 \text{ for } n \neq 0
\end{aligned} \tag{B.12}$$

or as a double series

$$\begin{aligned}
G(x, y, \xi, \eta) = & \frac{1}{a^2} \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{\varepsilon_n \varepsilon_m \cos(p_n x) \cos(q_m y) \cos(p_n \xi) \cos(q_m \eta)}{p_n^2 + q_m^2 - \lambda} \\
p_n = \frac{\pi n}{a}, \quad q_m = \frac{\pi m}{a}
\end{aligned} \tag{B.13}$$

In (B.5)

$$\begin{aligned}
\lambda = -2sR_s C_d, \quad f_1(y) = 0, \quad f_2(y) = 0, \quad f_3(x) = 0, \quad f_4(x) = 0, \text{ and} \\
\Phi(x, y) = -\frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \delta(x) \delta(y) \cdot
\end{aligned} \tag{B.14}$$

It is noted that  $\lambda$  ( $G(x, y, \xi, \eta)$  also becomes a function of  $s$ , because it is associated with  $\lambda$ ) and  $\Phi(x, y)$  become functions of  $s$ . Thus, the source function and Green's function can be rewritten to  $\Phi(x, y, s)$  and  $G(x, y, \xi, \eta, s)$ , respectively.

Now, the Helmholtz equation (B.5) can be solved by using (B.10). Substituting the boundary conditions and source function into (B.10), the terms related to the boundary condition are equal to zero and only the source function term remains in the equation,

$$\begin{aligned}
u(x, y, s) &= \int_0^a \int_0^a \Phi(\xi, \eta, s) G(x, y, \xi, \eta, s) d\eta d\xi \\
&= \int_0^a \int_0^a -\frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \delta(x) \delta(y) G(x, y, \xi, \eta, s) d\eta d\xi. \quad (B.15) \\
&= -\frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \cdot G(x, y, 0, 0, s)
\end{aligned}$$

In (B.15),  $u(\alpha D_{pad}, 0, s)$  is still unknown. Equation (B.11) itself can be used to solve for  $u(\alpha D_{pad}, 0, s)$ . When  $x = \alpha D_{pad}$ ,  $y = 0$ , (B.15) changes to

$$u(\alpha D_{pad}, 0, s) = -\frac{R_s}{4sL_p} \left[ u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \right] \cdot G(\alpha D_{pad}, 0, 0, 0, s). \quad (B.16)$$

Voltage  $u(\alpha D_{pad}, 0, s)$  can be solved as

$$u(\alpha D_{pad}, 0, s) = \frac{\frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}. \quad (B.17)$$

Substituting (B.17) back into (B.15),  $u(x, y, s)$  can be solved as

$$\begin{aligned}
u(x, y, s) &= -\frac{R_s}{4sL_p} \left[ \frac{\frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} - \frac{J(s)}{2sC_d} \right] \cdot G(x, y, 0, 0, s) \\
&= -\frac{R_s}{4sL_p} \left[ \frac{\frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s) - \frac{J(s)}{2sC_d} - \frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} \right] \cdot G(x, y, 0, 0, s) \\
&= \frac{R_s}{4sL_p} \left[ \frac{\frac{J(s)}{2sC_d}}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} \right] \cdot G(x, y, 0, 0, s)
\end{aligned} \quad (B.18)$$



Because  $V(x, y, s) = u(x, y, s) - \frac{J(s)}{2sC_d}$ ,  $V(\alpha D_{pad}, 0, s)$  can be solved using (B.17),

$$\begin{aligned}
V(\alpha D_{pad}, 0, s) &= u(\alpha D_{pad}, 0, s) - \frac{J(s)}{2sC_d} \\
&= \frac{\frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} - \frac{J(s)}{2sC_d} \\
&= \frac{\frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s) - 1 - \frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} \\
&= -\frac{\frac{J(s)}{2sC_d}}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}
\end{aligned} \tag{B.19}$$

and  $V(x, y, s)$  can be solved using (B.18),

$$\begin{aligned}
V(x, y, s) &= u(x, y, s) - \frac{J(s)}{2sC_d} \\
&= \frac{R_s}{4sL_p} \left[ \frac{\frac{J(s)}{2sC_d}}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} \right] \cdot G(x, y, 0, 0, s) - \frac{J(s)}{2sC_d} \\
&= \frac{\frac{J(s)}{2sC_d} \cdot \frac{R_s}{4sL_p} \cdot G(x, y, 0, 0, s) - \frac{J(s)}{2sC_d} - \frac{J(s)}{2sC_d} \cdot \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} \\
&= \frac{\frac{R_s}{4sL_p} \cdot \frac{J(s)}{2sC_d} [G(x, y, 0, 0, s) - G(\alpha D_{pad}, 0, 0, 0, s)] - \frac{J(s)}{2sC_d}}{1 + \frac{R_s}{4sL_p} \cdot G(\alpha D_{pad}, 0, 0, 0, s)} \\
&= -\frac{s \cdot \frac{J(s)}{2C_d} - \frac{J(s) \cdot R_s}{2C_d \cdot 4L_p} [G(x, y, 0, 0, s) - G(\alpha D_{pad}, 0, 0, 0, s)]}{s^2 + s \cdot \frac{R_s}{4L_p} G(\alpha D_{pad}, 0, 0, 0, s)}
\end{aligned} \tag{B.20}$$

The final form of (B.20) is exactly the same as (2.6), and the solution to (2.5) is therefore obtained.

## REFERENCES

- [1] G. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, No. 8, pp. 114-117, April 1965.
- [2] [www.intel.com](http://www.intel.com), Nov. 23, 2005.
- [3] J. D. Meindl, "Gigascale Integration: Is the Sky the Limit?" *IEEE Circuits & Devices*, vol. 12, pp. 19-24, 32, Nov, 1996.
- [4] J. Davis et al., "Interconnect Limits on Gigascale Integration (GSI) in the 21st Century," *Proceedings of IEEE*, vol. 89, no. 3, Mar, 2001.
- [5] J. D. Meindl, "Low Power Microelectronics: Retrospect and Prospect," *Proceedings of IEEE*, vol. 83, pp. 619-635, Apr. 1995.
- [6] M. Swaminathan, E. Engin, *Power Integrity: Modeling and Design for Semiconductor and Systems*, Prentice Hall PTR, 1st Edition, 2007.
- [7] H. Zheng, B. Krauter and L.T. Pileggi, "Electrical Modeling of Integrated-Package Power/Ground Distributions," *IEEE Design and Test of Computer*, vol. 20, no. 3, pp. 23-31, May-June, 2003.
- [8] K. L. Wong, T. Rahal-Arabi, M. Ma, and G. Taylor, "Enhancing Microprocessor Immunity to Power Supply Noise with Clock-Data Compensation," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, April, 2006.
- [9] W. D. Becker, J. Eckhardt, R. W. Frech, G. A. Katopis, E. Klink, M. F. McAllister, T. G. MacNamara, P. Muench, S. R. Richter, and H. H. Smith, "Modeling, Simulation, and Measurement of Mid-Frequency Simultaneous Switching Noise in Computer Systems," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, part B*, vol. 21, pp. 157-163, May 1998.
- [10] B. Garben, M. F. McAllister, W. D. Becker, and R. Frech, "Mid-Frequency Delta-I Noise Analysis of Complex Computer System Boards with Multiprocessor Modules and Verification by Measurements," *IEEE Transactions on Advanced Packaging*, vol. 24, pp. 294-303, Aug. 2001.

- [11] H. Chen, and J. Neely, "Interconnect and Circuit Modeling Techniques for Full-Chip Power Supply Noise Analysis," *IEEE Transaction on Components, Packaging, and Manufacturing Technology, part B*, vol. 21, pp. 209-215, Aug. 1998.
- [12] A. E. Ruehli, "Equivalent Circuits Models for Three Dimensional Multiconductor Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 22, no. 3. pp. 216--220, 1974.
- [13] M. Kamon, J. Tsuk, and J. White, "Fast Henry: A Multipole Accelerated 3-D Inductance Extraction Program," *Proceedings of the 30th Design Automation Conference*, Dallas, June 1993.
- [14] Raphael: Interconnection Analysis Program, TMA Inc, 1996
- [15] P. A. Brennan, N. Raver, and A. Ruehli, "Three Dimensional Inductance Computations with Partial Element Equivalent Circuits," *IBM Journal of Research and Development*, vol. 23, pp. 661-668, Nov. 1979
- [16] L.W. Nagel, and D. Pederson, "SPICE: Simulation Program with Integrated Circuits Emphasis," University of California, Berkeley, CA, ERL-M382, 1973
- [17] J. He, M. Celik, and L. Pileggi," SPIE: Sparse PEEC Inductance Extraction," *Proceedings of the 34th Design Automation Conference*, Anaheim, 1997.
- [18] K. Shakeri and J. D. Meindl, "Relative Inductance Extraction Method," *Proceedings of the IEEE 2004 Custom Integrated Circuits Conference*, Orlando, October, 2004.
- [19] A. Odabasioglu, M. Celik and L. T. Pileggi, "PRIMA: Passive Reduced-Order Interconnect Macromodeling Algorithm," *IEEE Transactions on Computer-Aided Design*, vol. 17, no. 8, pp. 645-654, August 1998.
- [20] I. M. Elfadel and D. D. Ling, "A Block Rational Arnoldi Algorithm for Multipoint Passive Model-Order Reduction of Multiport RLC Networks," *IEEE/ACM Proceeding of International Conference on Computer-Aided Design*, San Jose, Nov. 1997, pp. 66-71.

- [21] J. Fang and J. Ren, "Locally Conformed Finite-Difference Time-Domain Algorithm of Modeling Arbitrary Shape Planar Metal Strips," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, pp. 830-838, May 1993.
- [22] Speed2000 product brochure,  
<http://www.sigritty.com/products/speed2000/Speed20001.pdf>, Aug. 27, 2007
- [23] J. H. Kim and M. Swaminathan, "Modeling of Irregular Shaped Power Distribution Planes Using the Transmission Matrix Method," *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, vol. 24, pp. 334–346, Aug. 2001.
- [24] M. Choi and A. C. Cangellaris, "A Quasi Three-Dimensional Distributed Electromagnetic Mode for Complex Power Distribution Networks," *Proceedings of the 51st Electronic Component and Technology Conference*, pp. 83–86, Orlando, May 2001.
- [25] O.P. Mandhana, "Design Oriented Analysis of Package Power Distribution System Considering Target Impedance for High Performance Microprocessors", *Electrical Performance of Electronic Packaging Conference*, Boston, Oct. 2001.
- [26] T. Rahal-Arabi, G. Taylor, M. Ma, and C. Webb, "Design and Validation of the Pentium® III and Pentium® 4 Processors Power Delivery," *2002 Symposium on VLSI Circuits Digest of Technical Papers*, pp. 220-223, June, 2002.
- [27] G. Huang, D. Sekar, A. Naeemi, K. Shakeri, and J. D. Meindl, "Compact Physical Models for Power Supply Noise and Chip/Package Co-Design of Gigascale Integration", *Electronic Component and Technology Conference*, Reno, June, 2007.
- [28] K. Shakeri and J. D. Meindl, "Compact Physical IR-Drop Models for Chip/Package Co-Design of Gigascale Integration (GSI)," *IEEE Transactions on Electron Devices*, vol. 52, no. 6, pp. 1087-1096, June, 2005.
- [29] R. Tummala, *Fundamentals of Microsystems Packaging*, McGraw Hill, 2001.
- [30] N. Na, T. Budell, E. Tremble, and I. Wemple "The Effects of On-chip and Package Decoupling Capacitors and an Efficient ASIC Decoupling Methodology", *Electronic Component and Technology Conference*, pp. 556-567, June, 2004.

- [31] M. Powell et al., "Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep-Submicron Cache Memories," *Proceedings of ACM International Symposium on Low-Power Electronics and Design*, pp. 90-95, 2000,.
- [32] Q. Wu, M. Pedram, and X. Wu, "Clock-Gating and Its Application to Low Power Design of Sequential Circuits," *IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 3, Mar. 2000, pp. 415-420.
- [33] Arindam Mukherjee, Malgorzata Marek-Sadowska, "Clock and Power Gating with Timing Closure," *IEEE Design and Test of Computers*, vol. 20, no. 3, pp. 32-39, May/Jun, 2003
- [34] M. Prakash, "Cooling Challenges for Silicon Integrated Circuits," *Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, Las, Vegas, May, 2004.
- [35] K. Banerjee, S. J. Souri, P. Kapur, and K. C. Saraswat, "3D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration", *Proceedings of the IEEE*, vol. 89, no. 5, May, 2001, pp. 602-633
- [36] John U. Knickerbocker, Paul S. Andry, Bing Dang, Raymond R. Horton, Chirag S. Patel, Robert Polastre, Katsuyuki Sakuma, Edmund Sprogis, Cornelia K. Tsang, Steven L. Wright : "3D Chip Stacks & Silicon Packaging Technology using Through-Silicon-Vias (TSV) for Systems Integration" , *3D System Integration Conference (3D-SIC)* , 2007
- [37] Held, J., Bautista, J. and Koehl, S., "From a Few Cores to Many: A Tera-scale Computing Research Overview," Research at Intel White Paper.
- [38] W. Dally and J. Poulton, *Digital Systems Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [39] Semiconductor Industry Association, "International Technology Roadmap for Semiconductor (ITRS)," 2004 version.
- [40] H. Zhang, V. George, J. Rabeay, "Low-swing on-chip signaling techniques: effectiveness and robustness," *IEEE Trans. VLSI Syst*, vol. 8, no. 3, pp. 264-272, 2000.

- [41] C. Svensson, "Optimum Voltage Swing on On-Chip and Off-Chip Interconnect", *IEEE J. of Solid-State Circuits*, vol. 36, no. 7, pp. 1108-1112, July, 2001.
- [42] B. Sinharoy, et al., "POWER5 system microarchitecture," *IBM J. Research & Development*, Vol. 49, No. 4/5, 2005
- [43] A. Dharchoudhury, R. Panda, D. Blaauw, R. Vaidyanathan, "Design and Analysis of Power Distribution Networks in PowerPC Microprocessors," *Design Automation Conference*, 15-19 June 1998 pp: 738 – 743.
- [44] M. K. Gowan, L. L. Biro, D. B. Jackson, "Power Considerations in the Design of the Alpha 21264 Microprocessor," *Design Automation Conference*, June, 1998.
- [45] S. Pant and E. Chiprout, "Power Grid Physics and Implications for CAD," *Proceedings of the 43rd Design Automation Conference*, Anaheim, pp. 119–204, 2006.
- [46] A. D. Polyanin, *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, 2002.
- [47] Semiconductor Industry Association, "International Technology Roadmap for Semiconductor (ITRS)," 2005 version.
- [48] Bai, et al., "A 65nm Logic Technology Featuring 35nm Gate Lengths,. Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-k ILD and 0.57 $\mu$ m SRAM Cell," *International Electron Device Meeting Technical Digest*, pp. 657-660, 2004.
- [49] S. Morton. "Inductance: Implications and Solutions for High-Speed Digital Circuits On-Chip Signaling," *Proceedings of International Solid-State Circuit Conference* 2002.
- [50] Z. Qi, H. Li, S. X.-D. Tan, L. Wu, Y. Cai, and X. Hong, "Fast Decap Allocation Algorithm For Robust On-Chip Power Delivery," *Proceedings of. International Symposium on Quality Electronic Design*, March, 2005.
- [51] A. Muramatsu, M. Hashimoto and H. Onodera, "Effects of On-Chip Inductance on Power Distribution Grid," *International Symposium on Physical Design*, April, 2005.

- [52] W. H. Lee, S. Pant and D. Blaauw, "Analysis and Reduction of. On-chip Inductance Effects in Power Supply Grids," *Proceedings of. International Symposium on Quality Electronic Design*, March, 2004.
- [53] M. S. Bakir, H. A. Reed, H. D. Thacker, P. A. Kohl, K. P. Martin, and J. D. Meindl, "Sea of Leads (SoL) ultrahigh density wafer level chip input/output interconnections," *IEEE Trans. Electron Devices*, vol. 50, no. 10, pp. 2039-2048, Oct. 2003
- [54] S. Yang et al, "A High Performance 180nm Generation Logic Technology," *Proceedings of International Electron Device Meeting*, San Francisco, Dec. 1998, pp. 197--200.
- [55] S. Tyagi, et al, "A 130nm Generation Logic Technology Featuring 70nm Transistors, Dual Vt Transistors and 6 Layers of Cu Interconnects", *International Electron Device Meeting Technical Digest*, San Francisco, Dec. 2000.
- [56] C. H. Jan et al, "90nm Generation, 300mm Wafer Low k ILD/Cu Interconnect Technology," *Proceedings of the IEEE 2003 International Interconnect Technology Conference*, San Francisco, pp. 15-17, June, 2003.
- [57] S. R. Nassif and O. Fakhouri, "Technology Trends in Power-Grid-Induced Noise," *Proceedings of the 2002 International Workshop on System-level Interconnect Prediction*, pp. 55-59, April, 2002.
- [58] G. Huang, D. Sekar, A. Naeemi, K. Shakeri, and J. D. Meindl, "Physical Model for Power Supply Noise and Chip/Package Co-Design in Gigascale Systems with the Consideration of Hot Spots", *IEEE Custom Integrated Circuits Conference*, San Jose, September, 2007.
- [59] G. Huang, M. Bakir, A. Naeemi, H. Chen, and J. D. Meindl, "Power delivery for 3D chip stacks: physical modeling and design implication," *Proc. Electrical Performance of Electronic Packaging*, pp. 205-208, Atlanta, Oct, 2007,
- [60] B. Dang, M. S. Bakir, and J. D. Meindl, "Integrated Thermal-Fluidic I/O Interconnects for An On-Chip Microchannel Heat Sink," *IEEE Electron Device Letter*, vol. 27, pp. 117-119, 2006.
- [61] Gang Huang, Azad Naeemi, Tingdong Zhou, Daniel O'Connor, Bhupindra Singh, Andrew Muszynski, Wiren D. Becker, James Venuto and James D. Meindl,

- "Compact Physical Models of Chip and Package Power and Ground Distribution Networks for Gigascale Integration," *Proceedings of IEEE Electronic Components and Technology Conference (ECTC)*, pp. 646-652, Orlando, May, 2008
- [62] J. H. Kim and M. Swaminathan, "Modeling of Irregular Shaped Power Distribution Planes using the Transmission Matrix Method", *IEEE Transactions on Advanced Packaging*, Vol. 24, No. 3, pp. 334-346, Aug. 2001.
  - [63] R. Wigington, and N. Nahman, "Transient Analysis of Coaxial Cables Considering Skin Effect," *Proc. IRE*, vol. 45, pp.166-174, 1957.
  - [64] A. Naeemi, Xu Jianping, A. V. Mule', T. K. Gaylord, J. D. Meindl, "Optical and electrical interconnect partition length based on chip-to-chip bandwidth maximization," *Photonics Technology Letters, IEEE*, vol. 16 , issue 4 , pp. 1221 – 1223, April 2004.
  - [65] Daniel Zwillinger, *Standard Mathematical Tables and Formulae*, CRC Press,1996
  - [66] Dake Liu et al., "Power consumption estimation in CMOS VLSI chips," *IEEE J. Solid-State Circuits*, vol,29, pp.663-670, June, 1994.
  - [67] Sharkey, A. Buyuktosunoglu, P. Bose. "Evaluating Design Tradeoffs in On-Chip Power Management for CMPs", *International Symposium on Low Power Electronics and Design (ISLPED)*, August 2007
  - [68] Muhannad Bakir, Calvin King, Deepak Sekar, Hiren Thacker, Bing Dang, Gang Huang, Azad Naeemi, and James D Meindl, "Ultra-Compact and High-Performance 3D Integrated Heterogeneous Systems Using Microliquid Cooling", *IEEE Custom Integrated Circuits Conference (CICC)*, San Jose, September 2008.
  - [69] Van Zant, Peter (2000). *Microchip Fabrication*. New York: McGraw-Hill.
  - [70] P. Zarkesh-Ha, "Global Interconnect Modeling for a Gigascale System-on-a-Chip (GSoC)," Ph. D. Dissertation, 2001, Georgia Institute of Technology.



## **VITA**

### **GANG HUANG**

Gang Huang was born in Huadian, China, in April, 1977. He received his B.S. and M.S. degrees in Electronic Engineerin from Tsinghua University in 1999 and 2002, respectively. He obtained another M.S. degree in Electrical and Computer Engineering from Georgia Institute of Technology in 2005. He is currently a Ph.D. candidate under the guidance of Dr. James D. Meindl in the Gigascale Integration (GSI) Group at the Microelectronics Research Center (MiRC) of Georgia Institute of Technology.

From 1998 to 2002, he was a research assistant at EDA (Currently called NICS) lab in Tsinghua University performing research in the area of interconnect modeling and high-level synthesis of digital systems, guided by Dr. Huazhong Yang. From 2003 to present, He has been a graduate research assistant in the GSI group, and working in the areas of modeling and design for power delivery system in Gigascale integration and 3D integration, as well as modeling and optimization for GSI on-chip and off-chip interconnect.