Core Switching Noise for On-Chip
3D Power Distribution Networks

WAQAR AHMAD

TRITA-ICT/ECS AVH 12:06
ISSN 1653-6363
ISRN KTH/ICT/ECS/AVH-12/06-SE
ISBN 978-91-7501-519-4

© Waqar Ahmad, October 2012

KTH School of Information and Communication Technology, SE- 164 40 Kista, Sweden.

Thesis submitted to the School of Information and Communication Technologies, Department of Electronic Systems (ES), KTH Royal Institute of Technology, to be publicly defended on 2012-11-07 at 10:00 hrs in Stockholm, Kista.
Abstract

Reducing the interconnect size with each technology node and increasing speed with each generation increases IR-drop and Ldi/dt noise. In addition to this, the drive for more integration increases the average current requirement for modern ULSI design. Simultaneous switching of core logic blocks and I/O drivers produces large current transients due to power distribution network parasitics at high clock frequency. The current transients are injected into the power distribution planes thereby inducing noise in the supply voltage. The part of the noise that is caused by switching of the internal logic load is core switching noise. The core logic switches at much higher speed than driver speed whereas the package inductance is less than the on-chip inductance in modern BGA packages. The core switching noise is currently gaining more attention for three-dimensional integrated circuits where on-chip inductance is much higher than the board and package inductance due to smaller board, and package. The switching noise of the driver is smaller than the core switching noise due to small driver size and reduced capacitance associated with short on-board wires for three-dimensional integrated circuits. The load increases with the addition of each die. The power distribution TSV pairs to supply each extra die also introduce additional parasitic. The core switching noise may propagate through substrate and consequently through interconnecting TSVs to different dies in heterogeneous integrated system. Core switching noise may lead to decreased device drive capability, increased gate delays, logic errors, and reduced noise margins. The actual behavior of the on-chip load is not well known in the beginning of the design cycle whereas altering the design during later stages is not cost effective. The size of a three-dimensional power distribution network may reach billions of nodes with the addition of dies in a vertical stack. The traditional tools may run out of time and memory during simulation of a three-dimensional power distribution network whereas, the CAD tools for the analysis of 3D power distribution network are in the process of evolution. Compact mathematical models for the estimation of core switching noise are necessary in order to overcome the power integrity challenges associated with the 3D power distribution network design. This thesis presents three different mathematical models to estimate core switching noise for 3D stacked power distribution networks. A time-domain-based mathematical model for the estimation of design parameters of a power distribution TSV pair is also proposed. Design guidelines for the estimation of optimum decoupling capacitance based on flat output impedance are also proposed for each stage of the vertical chain of power distribution TSV pairs. A mathematical model for tradeoff between TSV resistance and amount of decoupling capacitance on each DRAM die is proposed for a 3D-DRAM-Over-Logic system. The models are developed by following a three step approach: 1) design physical model, 2) convert it to equivalent electrical model, and 3) formulate the mathematical model based on the electrical model. The accuracy, speed and memory requirement of the proposed mathematical model is compared with equivalent Ansoft Nexxim models.
Acknowledgments

This thesis is published after four years of hard work which have contributed enormously towards my professional and technical growth. I express my deepest gratitude to the Department of Electronic, Computer, and Software Systems (ECS) of the School of Information and Communication Technologies (ICT) of KTH Royal Institute of Technology for the facilitation of this research work.

First of all, I would like to thank my advisor Prof. Hannu Tenhunen for his ease of cooperation: I would simply describe him as a person with the ability to quickly understand a problem and propose simple solutions. The best thing I liked about him was that he never imposed himself on me rather he helped my natural flow and encouraged every move in my studies and research. However, the initial journey towards PhD studies was complicated as I left my family back home and consequently had to return to settle some practical issues which required my attention. I am indebted to Hannu for being so positive and cooperative during the time of my return when I picked up the strings again. I most appreciate the guidance and support of Li-Rong Zheng, and Qiang Chen during my PhD studies. My thanks go to Dr. Roshan Weerasekera for his help during the beginning of my PhD studies when it was rather difficult to find the right way. I got a lot of inspiration and boosting from him as a senior member of the ELITE project. I am thankful to Awet Yemane Weldezion, Roshan Weerasekera, and all the others involved in the ELITE project for their useful company during the ELITE project meetings and workshops. I enjoyed the lunch time discussions with Prof. Hemani and other friends. Thanks to Prof. Elena Dubrova for her useful tips for my PhD studies.

I deeply acknowledge the sacrifices of my family and children during the course of my PhD studies. They always supported me through thick and thin. My situation dramatically improved after the arrival of my wife and children in Stockholm. I really picked up pace and moved in the right direction at that time. It was all possible because of love and care from my wife Khalida Waqar, my son Yahyea Waqar, my daughter Mariam Waqar, and the sweet addition of little baby Uzair Waqar.

I remember the cricket matches along with my fellow Pakistani PhD students during summer. It was great fun and entertainment which kept us
all together. I am thankful to all my friends and families too numerous to mention here for giving me their dear company and support on the road towards my PhD.

I became seriously ill while writing the thesis when the finishing line was very close. I am grateful to all the doctors and nurses of Karolinska Hospital for their outstanding care and support. Thanks to my wife, for being there by the hospital bed to help me when I needed support to stand up. Special thanks go to Prof. Hannu Tenhunen, Prof. Axel Jantsch, and all others for managing the funding during that difficult time. I am also grateful to Alina Munteanu for her constant cooperation regarding the administrative matters.

Let me take this opportunity to admire the cooperation of the Swedish Institute (SI) and the Higher Education Commission (HEC) of Pakistan. I must pay tribute to HEC for the funding of my PhD studies and to SI for the support in my smooth stay here. I must also thank the travel grants support of the FP7-ICT-215030 (ELITE) project of the 7th framework program.

Last but not the least I would also like to pay tribute to my employer organization for paying me a full salary, which made it financially possible to support my family in Stockholm.

Waqar Ahmad

September, 2012

Stockholm, Sweden.
List of Publications included in the thesis


Summary of Publications included in the thesis

Paper 1. A comprehensive mathematical model is developed in this paper in order to investigate the behavior of power/ground noise as a function of the rise time for an inductive power distribution, TSV pair with decoupling capacitance. A TSV pair forming power and ground is modeled as a series RL circuit. The decoupling capacitance is modeled as a capacitor with small effective series resistance. The switching logic load connected to the power distribution TSV pair is modeled as a linear ramp current source. The switching noise is also modeled as a ramp function. The analysis is performed for the optimal design of the power distribution TSV pair to maintain a certain level of switching noise based on the worst case rise time.

Author’s Contributions: The author conceived the idea, performed modeling and simulation, and wrote the manuscript. To the author’s knowledge, so far there is no model available like this in the literature for the design of the power distribution TSV pair. The model is based on time domain analysis of noise. The model has +/- 5% deviation compared to Ansoft Nexxim. This model is simple to implement using Matlab. The proposed model provides guidelines for the physical design of a power distribution TSV pair based on the on-chip load requirements. In addition to this it is useful for optimal adjustment of power distribution TSV pair design and associated decoupling capacitance for specific load requirements.

Paper 2. In this paper an efficient and accurate model is proposed to estimate the peak-to-peak core switching noise caused by simultaneous switching of logic loads along a vertical chain of the power distribution TSV pairs in a 3D stack of ICs.

Author’s Contributions: The author conceived the idea, performed modeling and simulation, and wrote the paper. The model is accurate with 2-3% deviation compared to Ansoft Nexxim. In addition to this, the model is 3-4 times faster than Ansoft Nexxim while requiring half of the memory. The proposed model is very useful for early estimation of the core switching noise along a chain of power distribution TSV pairs from bottom to top in a three-dimensional (3D) power distribution network. To the author’s knowledge, so far there is no model available like this in the literature for the vertical chain of the power distribution TSV pairs. The proposed model is flexible and takes into account the resistance, inductance, and capacitance of each TSV in the loop of switching current path. The model is proposed for n number of power
distribution TSV pairs in a vertical chain. Physical dimensions of each power TSV pair can be varied according to the requirement of the design. Values of resistance, inductance, and capacitance of pads, solder bumps, and interconnecting metallic lines can also be included in the model.

**Paper 3.** This paper is an extension of paper 2. In addition to the model, design guidelines for approximately flat output impedance for each stage of the vertical chain of power distribution TSV pairs are formulated. This is needed to realize the resonance free scenario over a wide operating frequency range.

The proper design of decoupling capacitance is a challenging issue in order to maintain the output impedance equal to or less than the target impedance for each tier of a three-dimensional (3D) power distribution network.

*Author’s Contributions:* The author conceived the idea, performed modeling and simulation, and wrote the manuscript. The proposed design guidelines are useful for the selection of the value of decoupling capacitance along with effective series resistance (ESR), and effective series inductance (ESL) for each stage of the chain of power distribution TSV pairs. First of all the target impedance for each stage of the vertical chain of power distribution TSV pairs is determined according to allowed ripple and load at each stage. An equation is given to estimate the output impedance. The output impedance should be kept less than or equal to the target impedance. There are separate equations for the estimation of ESR and ESL for the decoupling capacitance. Finally, an equation is given for the estimation of the value of decoupling capacitance. The proposed technique is applicable to a large three-dimensional (3D) power distribution network containing multiple tiers, and multiple vertical power distribution chains.

**Paper 4.** A mathematical model for the estimation of core switching noise within a three-dimensional (3D) stack of dies interconnected through power distribution TSV pairs is proposed in this paper. The core switching noise in this model is a function of the value of resistance, inductance, and capacitance of the power distribution TSVs. In addition to this, the switching noise also depends on the value of the rise time, and amount of the logic load. The proposed model has +/- 3% deviation compared to an equivalent SPICE model. The proposed model is flexible and can be applied to uniform or non-
uniform three-dimensional (3D) power distribution network. The proposed model can easily be solved for unknown voltages using Matlab. The known and unknown voltages along with RLC values of TSVs, RLC values of power distribution grid, logic load associated with each node, decoupling capacitance associated with each node, and rise time of the clock are put together in the form of matrices. The model is solved for unknown voltages using the direct method of solution in Matlab.

Author’s Contributions: The author performed modeling and simulation, and wrote the paper. The proposed model is applicable to any number of power distribution TSVs according to logic load requirements. The pattern of power distribution TSV pairs can also be changed according to on-chip area constraints.

Paper 5. This paper is based on extended analysis of the model proposed in paper 4. The logic dies are interconnected using a single power distribution TSV pair in the center of each die for a 1mmx1mm grid in this paper and TSVs are placed along the periphery of the chips in paper 4. The behavior of core switching noise is investigated by adding dies in the vertical stack. The behavior of core switching noise is also investigated by increasing the clock frequency when only two dies are interconnected through a single power distribution TSV pair. The analysis is useful for early design tradeoffs for the design of a three-dimensional (3D) power distribution network.

Author’s Contributions: The author performed modeling and simulation, and wrote the paper.

Paper 6. Extensive transient simulations for on-chip power distribution networks are required to analyze power delivery fluctuations caused by dynamic IR and Ldi/dt drops. Speed and memory has become a bottleneck for simulation of power distribution networks in modern VLSI design where clock frequency is of the order of GHz. The traditional tools are very slow and require a lot of memory resources for simulation. The problem is further aggravated for huge networks like power distribution networks within a stack of ICs inter-connected through TSVs. This type of 3D power distribution network may contain billions of nodes at a time. This paper presents a fast transient simulation algorithm for the 3D power distribution bus. The proposed algorithm uses a combination of mathematical techniques and
visual C++ to reach a fast solution of the three-dimensional (3D) power distribution bus for voltage at each node. The branches at each node of the 3D power distribution bus are converted to a combination of resistance in series with the current source, using the trapezoidal rule. The star network obtained this way is converted to the corresponding π network for each node of the 3D power distribution bus. This network is reduced to a single π network using C++ reducing code. The reduced network is solved using nodal analysis i.e. the network is converted to a matrix equation containing unknown voltages and currents. The system of matrix is solved for unknowns using the direct method of solution in Matlab. Now, π network is back-solved for each node using C++ back solution code. The proposed algorithm has +/- 1 to 2% deviation compared to Ansoft Nexxim4.1. The proposed algorithm is several times faster than Ansoft Nexxim and also requires significantly less memory than Ansoft Nexxim. The speed is improved and less memory is required because of a significant reduction in the number of nodes compared to Ansoft Nexxim as shown by Table 5.5. The advantage in speed compared to Nexxim is also gained by converting the inductance and capacitance of each branch to the corresponding resistive element. The accuracy is achieved by means of an efficient algorithm using C++.

Author’s Contributions: The author conceived the idea, performed modeling and simulation, and wrote the paper. The model overcomes the speed and memory bottlenecks which is a prime concern because of the large size of the 3D power distribution network.

Paper 7. The core switching noise and LC resonance vary by varying different circuit parameters of the power distribution TSV pair for a three-dimensional (3D) power distribution network. Variation of circuit parameters such as effective resistance, effective inductance, and the damping factor of the power distribution TSV pair affect the core switching noise and LC resonance. The variation of LC resonance vs. time is explored in this paper. The analysis shows that the magnitude of the transient term reduces by increasing the effective resistance of the TSV pair. Specifically, the peak of the transient term reduces by adding decoupling capacitance. Reducing effective inductance of the TSV pair not only reduces the peak of the transient term but also reduces the lasting time of the transient term.

The analysis shows 4-5% variation compared to an equivalent Ansoft Nexxim model. The analysis is useful for choosing the design of power
distribution TSV pair to control the magnitude and lasting time of the transient term as a result of the switching logic load.

Author’s Contributions: The author conceived the idea, performed modeling and simulation, and wrote the paper.

Paper 8. The three-dimensional (3D)-DRAM-Over-Logic system is one of the promising applications of three-dimensional (3D) integration technology. It can be a vibrant technique to overcome memory wall and bandwidth wall problems. This paper considers a three-dimensional (3D) stack containing two DRAM dies over a single processor die. Placing decoupling capacitance on DRAM dies saves useful area on the processor die to increase the integration density in order to make full use of the bandwidth offered by TSVs in the vertical direction. The assumption is that decoupling capacitors are placed on each DRAM die and connected to power distribution TSV pairs passing through the DRAM stack.

The proposed mathematical model determines the optimum value of the decoupling capacitance on each DRAM die along with the optimum values of the effective resistance of the interconnecting power distribution TSV pairs. The proposed model has +/- 1.1% deviation compared to an equivalent Ansoft Nexxim model. The proposed model is flexible and useful for efficient budgeting of the amount of decoupling capacitance placed on DRAM dies.

Author’s Contributions: The author conceived the idea, performed modeling and simulation, and wrote the paper. To the author’s knowledge, so far this is the first mathematical model for efficient budgeting of decoupling capacitance and physical design of the power distribution TSV pairs for a 3D-DRAM-Over-Logic System.

Book Chapter. We presented an overview and state of the art of switching noise for three-dimensional (3D) stacked integrated circuits in the form of a book chapter.

Author’s Contributions: It is written by author along with Prof. Hannu Tenhunen. The chapter provides brief state of the art knowledge about switching noise in 3D stacked power distribution networks.
List of Publications not included in the thesis


List of Figures

Fig. 1.1 A vision of future three-dimensional (3D) hyper-integration of information technology, nanotechnology, and biotechnology systems, a new paradigm for future technologies [5]. ................................................................. 3

Fig. 2.1 The memory chip stacking by Samsung [38]. ........................................... 17

Fig. 2.2 Die-to-Die (D2D) stacking approaches: (a) Die-to-Die wire bonding, (b) Die-to-Die TSV bonding [24]. ................................................................................. 18

Fig. 2.3 Die-to-Wafer (D2W) stacking [41]................................................................ 19

Fig. 2.4 (a) Wafer-to-Wafer stacking. (b) A defective die may stack to good dies thereby reducing the yield in Wafer-to-Wafer stacking ............................................. 19

Fig. 2.5 Face-to-Face (F2F) bonding [44]............................................................... 20

Fig. 2.6 Face-to-Back (F2B) bonding [44]................................................................ 21

Fig. 2.7 Back-to-Back (B2B) bonding [44]............................................................ 21

Fig. 2.8 Via-first TSV technology [44].................................................................. 22

Fig. 2.9 Via-last TSV technology [44].................................................................... 22

Fig. 2.10 Dies inter-connected through bonding wires [11]................................. 23

Fig. 2.11 The dies interconnected through micro-bumps [11]............................... 24

Fig. 2.12 Cross section of three-dimensional (3D) power distribution network having three planes and a TSV pair................................................................. 24

Fig. 2.13 Top view of a 3D power distribution network [47].................................... 25

Fig. 2.14 3D integration of nano photonics and CMOS using Cu-Nails (TSVs) [48].................................................................................................................. 25

Fig. 2.15 Electro-static coupling through on-chip capacitors [12]............................ 26

Fig. 2.16 Electro-magnetic coupling through on-chip inductors [12]...................... 26

Fig. 3.1 Equivalent electrical model of a decoupling capacitance, where ESR is effective series resistance and ESL is effective series inductance of decoupling capacitance ........................................................................................................... 27

Fig. 3.2 V-curve showing impedance of the decoupling capacitance vs. frequency [49] .............................................................................................................. 28
Fig. 3.3 (a) Frequency response of the impedance of a power distribution network without decoupling capacitance. (b) Frequency response of the impedance of a power distribution network with decoupling capacitance [49].

Fig. 3.4 Distinctive peak of output impedance of a power distribution network due to anti-resonance [49].

Fig. 3.5 Anti-resonance of two equal value parallel decoupling capacitors i.e. $f_{ar1} < f_{ar2}$ as $f_{ar1}$ occurs when $L_1 > L_2$, and $f_{ar2}$ occurs when $L_1 < L_2$ [52].

Fig. 4.1 Three dies forming a three-dimensional (3D) system with face-to-back bonding through micro-connects. Power is supplied through vertical TSV pairs from the package substrate [29].

Fig. 4.2 3X3 orthogonal global supply grids of two neighboring dies in a 3D stack inter-connected through TSVs where red color indicates the power grid and black color indicates the corresponding ground grid.

Fig. 4.3 Physical model of a power distribution TSV pair with the logic load and decoupling capacitance.

Fig. 4.4 Schematic presentation of the on-chip current definition for CMOS [63].

Fig. 4.5 The shape of the on-chip logic switching current where $t_s$ is switching time.

Fig. 5.1 Block diagram of a typical power distribution network for an integrated circuit [22].

Fig. 5.2 Hierarchical Model [79].

Fig. 5.3 4X4 RLC power grid model of a typical System-on-Chip.

Fig. 5.4 Power Distribution node [86].

Fig. 5.5 Chip-Package-Board hierarchy of a typical three-dimensional (3D) power distribution network.

Fig. 5.6 Electrical model of decoupling capacitance placed on DRAM dies in a 3D-DRAM-Over-Logic System. $V_{dd}$ = supply voltage, $R_{p}^{eff}$ = effective resistance of package, $C_{dec1}$ = decoupling capacitance on top DRAM die, $V_{dec1}$ = voltage across $C_{dec1}$, $R_{TSV1}^{eff}$ = 2$R_{T1}$ = effective resistance of first TSV pair, $C_{dec2}$ =
decoupling capacitance on bottom DRAM die, $V_{\text{dec}2} = \text{voltage across } C_{\text{dec}2}$, $R_{\text{TSV}2}^{\text{eff}} = 2R_{\text{TSV}2}$ = effective resistance of second TSV pair, $V_L = \text{voltage across logic load}$, and $I_e = \text{equivalent current drawn by logic load}$..........................51

**Fig. 5.7** Value of $V_{\text{dec}2}$ vs. value of $C_{\text{dec}2}$ for different values of $C_{\text{dec}1}$ when $R_{\text{TSV}2}^{\text{eff}} = 50\, \text{m}\Omega$, $I_{\text{max}} = 3\, \text{mA}$, and $t_r = 100\, \text{p}\, \text{S}$.................................................................55

**Fig. 5.8** Value of $V_{\text{dec}2}$ vs. value of $R_{\text{TSV}}^{\text{eff}}$ for different values of $C_{\text{dec}1}$ when $R_{\text{TSV}2}^{\text{eff}} = 50\, \text{m}\Omega$, $C_{\text{dec}2} = 1\, \text{pF}$, $I_{\text{max}} = 5\, \text{mA}$, and $t_r = 50\, \text{p}\, \text{S}$.................................................................55

**Fig. 5.9** Value of $V_L$ vs. value of $C_{\text{dec}2}$ for different values of $t_r$ when $R_{\text{TSV}2}^{\text{eff}} = 50\, \text{m}\Omega$, $R_{\text{TSV}1} = 100\, \text{m}\Omega$, $C_{\text{dec}1} = 1\, \text{pF}$, and $I_{\text{max}} = 5\, \text{mA}$. Assume 10% ripple in load voltage..................................................................................................56

**Fig. 5.10** Electrical model of 3D power distribution bus having $n$ nodes........57

**Fig. 5.11** Equivalent bus consisting of trapezoidal Norton companion models at each node..................................................................................................................................57

**Fig. 5.12** Star to $\pi$ conversion model at a node.................................................................58

**Fig. 5.13** Equivalent $\pi$ model of a 3D bus after applying reduction algorithm.................................................................58

**Fig. 5.14** 3X3 orthogonal global supply grids of two neighboring dies in a 3D stack inter-connected through TSVs where red color indicates the power grid and black color indicates the corresponding ground grid..................................................61

**Fig. 5.15** Vertical power distribution TSV chains.................................................................63

**Fig. 5.16** Electrical model of a vertical chain of $n$ power distribution TSV pairs, interconnecting $n$ chips in a 3D stack of ICs.................................................................67

**Fig. 5.17** Equivalent electrical model of a power distribution TSV pair with logic load and decoupling capacitance..........................................................................................69

**Fig. 5.18** Clock and corresponding current and noise in time domain........70

**Fig. 5.19** Peak-to-peak ground noise comparison of model from equation (5.28) to the equivalent model in Ansoft Nexxim4.1, as a function of rise time for $t_{sw}^p = 3.3\, \text{m}\, \text{A}$, $L_{\text{TSV}}^{\text{eff}} = 50\, \text{pH}$, $R_{\text{TSV}}^{\text{eff}} = 0.25\, \Omega$, $R_{\text{dec}} = 75\, \text{m}\, \Omega$, $C_{\text{dec}} = 5\, \text{pF}$, and $t = t_1$ ........................................................................................................72
Fig. 5.20 Part of switching current supplied by decoupling capacitance as a function of rise time for $I_{sw} = 3.3mA$, $L_{TSV}^{eff} = 50pH$, $R_{TSV}^{eff} = 0.25\Omega$, $R_{dec}^{eff} = 75m\Omega$, $C_{dec} = 5pF$, and $t = t_s$ ..............................73

Fig. 5.21 Part of switching current drawn from power supply as a function of rise time for $I_{sw} = 3.3mA$, $L_{TSV}^{eff} = 50pH$, $R_{TSV}^{eff} = 0.25\Omega$, $R_{dec}^{eff} = 75m\Omega$, $C_{dec} = 5pF$, and $t = t_s$ ..............................74

Fig. 5.22 Peak-to-peak ground noise for different values of $R_{TSV}^{eff}$ for $I_{sw} = 3.3mA$, $L_{TSV}^{eff} = 50pH$, $R_{dec}^{eff} = 75m\Omega$, $C_{dec} = 5pF$, and $t = t_s$ .........................................................77

Fig. 5.23 Peak-to-peak ground noise for different values of $R_{dec}^{eff}$ for $I_{sw} = 3.3mA$, $L_{TSV}^{eff} = 50pH$, $R_{TSV}^{eff} = 0.25\Omega$, $C_{dec} = 5pF$, and $t = t_s$ .........................................................78

Fig. 5.24 Peak-to-peak ground noise reduction comparison by doubling the value of decoupling capacitance and by halving the value of TSV effective parasitic inductance.................................................................79

Fig. 5.25 Vertical chain of $n$ power distribution TSVs pairs interconnecting $n$ dies in a 3D stack. The power TSVs are in the red dotted box and the corresponding ground TSVs are in the black dotted box.................................80

Fig. 5.26 Electrical model for a vertical chain of $n$ power distribution TSV pairs.............................................................................................................................81

Fig. 5.27 The equivalent electrical model of a single power distribution TSV pair along with logic load and decoupling capacitance, where $I_{L}(t)$ is the current drawn by the logic load, $Z_{out}^{eff}$ is the effective output impedance across the load.................................................................82

Fig. 5.28 Plot obtained using (5.37), showing the magnitude of the normalized output impedance vs. normalized frequency for different values of $U_{L}$ for $U_{A} = Q_{TSV} = 1$ ..............................................................................................................85

Fig. 5.29 The plot obtained through (5.41), showing the magnitude of the normalized output impedance vs. the normalized frequency for different values of $Q_{TSV}$, for critically damped system.........................................................86

Fig. 5.30 The plot obtained through (5.42), showing the magnitude of the normalized decoupling capacitance vs. $U_{R}$ and $U_{L}$ ........................................87
List of Tables

Table 1.1 Classification of circuit size from 1963-2010 [1].................................1

Table 1.2 Intel Desktop Processor Trends for the last ten years [2]..........................................................4

Table 1.3 ITRS 2011 predictions about high volume microprocessors for the next decade [1]........................................4

Table 1.4 Comparison amongst different design approaches [14]..........................6

Table 1.5 Comparison amongst three-Dimensional (3D) assembly technologies [14]....................................................6

Table 1.6 Comparison of size and interconnect density between 3D IC and 3D SOP [16]........................................................................................................7

Table 1.7 Design challenges roadmap for 3D integration technology [1].........................8

Table 1.8 3D Interconnect Hierarchy Roadmap [1]................................................9

Table 1.9 Emerging Global Interconnect Level 3D-SIC/3D-SOC Roadmap for TSVs [1]........................................................................................................9

Table 1.10 Emerging Intermediate Interconnect Level 3D-SIC Roadmap for TSVs[1]........................................................................................................10

Table 4.1 TSV dimensions used for analysis.........................................................43

Table 5.1 Optimization of decoupling capacitance on both the DRAM dies based on the effective resistance of the TSV pairs, assuming 10% ripple to be allowed in the voltage across the load with the minimum value of $V_{dc,2}$ to be equal to 0.92V..................................................53

Table 5.2 Dependence of decoupling capacitance placed over DRAM dies on power distribution TSV pair’s resistance. Assume 10% ripple is allowed in voltage across load..................................................................................54

Table 5.3 Comparison of the proposed algorithm with Ansoft Nexxim for CPU time (Sec).................................................................58

Table 5.4 Comparison of the proposed algorithm with Ansoft Nexxim for required memory (Mb).................................................................60
Table 5.5 Comparison of the proposed algorithm with Nexxim for node reduction

Table 5.6 Simulation time comparison of proposed model with Ansoft Nexxim 4.1 by increasing the number of TSV pairs in vertical chain, for $R_{TSV}=0.25 \, \text{ohm}$, $L_{TSV}=50 \, \text{pH}$, $C_{Load}=500 \, \text{pF}$, and $C_d=5 \, \text{pF}$. Nodes are numbered from bottom to top.

Table 5.7 Memory requirement comparison of proposed model with Ansoft Nexxim 4.1 by increasing the number of TSV pairs in vertical chain, for $R_{TSV}=0.25 \, \text{ohm}$, $L_{TSV}=50 \, \text{pH}$, $C_{Load}=500 \, \text{pF}$, and $C_d=5 \, \text{pF}$. Nodes are numbered from bottom to top.

Table 5.8 Peak-to-peak ground noise comparison of (5.28) model with equivalent model obtained through Nexxim 4.1 for worst case rise time by varying values of effective inductance of TSV pair for $I_{sw}^p=3.3 \, \text{mA}$, $R_{TSV}^{eff}=0.25 \, \text{Ω}$, $R_{dec}^{eff}=0.075 \, \text{Ω}$, and $C_{dec}=5 \, \text{pF}$.

Table 5.9 Peak-to-peak ground noise comparison of (5.28) model with equivalent model obtained through Nexxim 4.1 for worst case rise time by varying values of effective resistance of TSV pair and effective series resistance of decoupling capacitor for $I_{sw}^p=3.3 \, \text{mA}$, $L_{TSV}^{eff}=50 \, \text{pH}$, and $C_{dec}=10 \, \text{pF}$.

Table 5.10 Peak-to-peak ground noise comparison of (5.28) model with equivalent model obtained through Nexxim 4.1 for worst case rise time by varying values of decoupling capacitance for $I_{sw}^p=3.3 \, \text{mA}$, $R_{TSV}^{eff}=0.25 \, \text{Ω}$, $R_{dec}^{eff}=0.075 \, \text{Ω}$, and $L_{TSV}^{eff}=50 \, \text{pH}$.

Table 5.11 Peak-to-peak ground noise reduction for different values of the rise time by increasing the value of decoupling capacitance for $I_{sw}^p=3.3 \, \text{mA}$, $L_{TSV}^{eff}=50 \, \text{pH}$, $R_{TSV}^{eff}=0.25 \, \text{Ω}$, $R_{dec}^{eff}=75 \, \text{mΩ}$, $C_{dec}=5 \, \text{pF}$ and $t=t_i$.

Table 5.12 Peak-to-peak ground noise reduction for different values of the rise time by decreasing the value of TSV effective inductance for $I_{sw}^p=3.3 \, \text{mA}$, $C_{dec}=5 \, \text{pF}$, $R_{TSV}^{eff}=0.25 \, \text{Ω}$, $R_{dec}^{eff}=75 \, \text{mΩ}$, $C_{dec}=5 \, \text{pF}$, and $t=t_i$. 
Abbreviations

3D: Three-Dimensional.

2D: Two-Dimensional.

IR drop: Voltage drop caused by resistance of the circuit due to current load.

Ldi/dt noise: Noise caused by inductance of the circuit due to switching of load.

TSV: Through-Silicon-Via.

I/O: Input/Output.

GHz: Gigahertz.

BGA: Ball Grid Array.

CAD: Computer-aided Design.

DRAM: Dynamic Random Access Memory.

ps: pico-second.

um: micro-meter.

nm: nano-meter.

PDN: Power Distribution Network.

IC: Integrated Circuit.

PCB: Printed Circuit Board.

P/G: Power/Ground.

ITRS: International Technology Road Map for Semiconductors.

ULSI: Ultra Large Scale Integration.

HCI: Hot Carrier Injection.

NBTI: Negative Bias Temperature Instability.

CMOS: Complementary Metal Oxide Semiconductor.

VLSI: Very Large Scale Integration.
**NMOS:** N-channel Metal Oxide Semiconductor.

**PMOS:** P-channel Metal Oxide Semiconductor.

**AC:** Alternating Current.

**RLC:** Resistance, Inductance, and Capacitance.

**ESR:** Effective Series Resistance.

**ESL:** Effective Series Inductance.

**CPU:** Central Processing Unit.

**nH:** nano-Henry.

**DC-DC:** Direct Current to Direct Current.

**D2D:** Die-to-Die.

**KGD:** Known Good Dies.

**D2W:** Die-to-Wafer.

**W2W:** Wafer-to-Wafer.

**F2B:** Face-to-Back.

**B2B:** Back-to-Back.

**F2F:** Face-to-Face.

**FEOL:** Front End of Line.

**BEOL:** Back End of Line.

**LC:** Inductance, and Capacitance.

**MHz:** Megahertz.
# Table of Contents

Abstract ....................................................................................................................... iii  
Acknowledgements ................................................................................................... v  
List of publications included in the thesis .............................................................. vii  
Summary of publications included in the thesis .................................................. ix  
List of publications not included in the thesis ..................................................... xv  
List of Figures ......................................................................................................... xvii  
List of Tables ............................................................................................................ xxi  
Abbreviations ........................................................................................................ xxiii  

1. Overview ............................................................................................................... 1  
   1.1 Motivation for Three-Dimensional Integration ............................................ 4  
   1.2 Challenges to Three-Dimensional Integration .............................................. 7  
   1.3 Road Map for 3D Interconnects ...................................................................... 9  
   1.4 Thesis Contribution and Objectives ............................................................. 10  
      1.4.1 Contribution......................................................................................................10  
      1.4.2 Objectives...........................................................................................................13  
   1.5 Overview of the Thesis Structure .................................................................. 14  

2. Structures for 3D Integration .............................................................................. 17  
   2.1 Three-Dimensional Stacking Approaches ................................................... 18  
      2.1.1 Die-to-Die (D2D) Stacking ................................................................. 18  
      2.1.2 Die-to-Wafer (D2W) Stacking ............................................................ 18  
      2.1.3 Wafer-to-Wafer (W2W) Stacking ..................................................... 18  

2.2 Three-Dimensional Bonding Approaches ............................................. 19
   2.2.1 Face-to-Face (F2F) Bonding ...................................................... 20
   2.2.2 Face-to-Back (F2B) Bonding ...................................................... 20
   2.2.3 Back-to-Back (B2B) Bonding ...................................................... 20
2.3 Three-Dimensional TSV Approaches ................................................. 22
2.4 Three-Dimensional Vertical Interconnect Approaches ....................... 23

3. On-Chip Decoupling Capacitance .......................................................... 27
   3.1 Impedance of the Decoupling Capacitance ........................................ 27
   3.2 Impact of Decoupling Capacitance on Target Impedance .................... 29
   3.3 Anti-resonance and parallel decoupling capacitors ............................ 31
   3.4 Effective On-Chip Decoupling Capacitance ....................................... 33

4. Core Switching Noise for 3D PDN .......................................................... 35
   4.1 Core Switching Noise ................................................................. 37
   4.2 Modeling of the Core Switching Noise .............................................. 39
   4.3 Toward a Robust 3D PDN Design ..................................................... 42

5. Modeling of On-Chip Power Distribution Networks ................................. 45
   5.1 Modeling Approaches ................................................................. 46
      5.1.1 Finite Difference Time Domain (FDTD) Method ......................... 46
      5.1.2 Latency Insertion Method (LIM) .............................................. 46
       5.1.3 Hierarchical Modeling Method ............................................... 46
      5.1.4 Distributed RLC Model for Power Grid ................................. 47
      5.1.5 Compact Physical Modeling for GSI Power Distribution Grids ....... 48
5.1.6 Combined Chip-Package-Board Modeling .......................................... 49
5.2 Previous Work and Thesis Contribution .............................................. 49

6. Summary and Future Work ................................................................. 91
6.1 Summary ......................................................................................... 91
6.2 Future Directions ............................................................................. 92

References ............................................................................................ 95
Chapter 1

Overview

The small scale commercial production of integrated circuits started in the early 1960's. In 1965, Gordon Moore predicted the transistor density would double every 18 to 24 months, which became known as Moore’s Law. After that the transistor density increased due to increased demand for small size in hand-held devices, System-on-Chip (SoC), and the defense industry. The modern day integrated circuits may have billions of transistors, as shown in Table 1.1 [1].

<table>
<thead>
<tr>
<th>Year</th>
<th>Integration Level</th>
<th>Transistor Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1963</td>
<td>Small Scale Integration (SSI)</td>
<td>&lt;100</td>
</tr>
<tr>
<td>1970</td>
<td>Medium Scale Integration (MSI)</td>
<td>100-300</td>
</tr>
<tr>
<td>1975</td>
<td>Large Scale Integration (LSI)</td>
<td>300-30000</td>
</tr>
<tr>
<td>1980</td>
<td>Very Large Scale Integration (VLSI)</td>
<td>30000-1 million</td>
</tr>
<tr>
<td>1990</td>
<td>Ultra Large Scale Integration (ULSI)</td>
<td>&gt;1 million</td>
</tr>
<tr>
<td>2010</td>
<td>Giga Scale Integration (GSI)</td>
<td>&gt;1 billion</td>
</tr>
</tbody>
</table>

The functionality comes by reducing the size of the transistors and also by increasing the transistor count per unit area of the processor die. More transistors are added to processors with each technology node as shown in Table 1.2 [2]. Speed or frequency of the processor is also increased with the passage of time in order to make full use of the functionality and resources. The transistor current density increases with each technology node. The capacitance of transistors reduces by decreasing the size. The power consumption should reduce due to reduction of voltage with each technology node but power consumption actually increases due to increase in frequency of the processor as shown in Table 1.2 [2].

There is a demand for increased chip size, and an increased number of I/O pins, which results in increased power dissipation due to increased speed and increased transistor count in future as shown in Table 1.3 [1]. System-on-Chip (SoC) was introduced in order to integrate all the components on a single substrate. SoC has incredible functionality having processor, digital logic, memory, embedded intelligence, and analog components on a single die. However, interconnect length is a critical issue for SoC due to increase in the chip dimensions. Mixed signal integration is another issue for present-day SoC as the noisy digital part cannot be placed in close proximity to the sensitive analog part. In addition to this, same technology node has to be followed for different components of SoC which may place stringent requirements on analog and RF components. Three-dimensional (3D)
integration with TSVs is, therefore, an obvious choice for mixed technology applications compared to SoC.

The inductance of on-chip power distribution grid causes ground bounce due to fast switching transistors. The simultaneous switching of I/O drivers also changes the voltage level of the power lines due to inherent inductance of the power distribution network. The overall effect of all these factors degrades the signal-to-noise ratio of integrated circuits. With the present advancements in semiconductor technology, the chip supply voltage has reached 0.9 V, maximum allowable power has reached 158 W, and frequency is of the order of GHz [1]. Complexities associated with all these developments are increased di/dt noise, reduced voltage margins, and increased IR-drop. On-chip inductance becomes significant at high frequencies and causes voltage fluctuations in the power distribution grid. These voltage fluctuations are in the form of overshoot or undershoot in power supply rails. An overshoot in the power rail (or undershoot in the ground rail) causes overstress in transistor gate-oxide or excessive charge/discharge in power distribution network nodes [3]. Conversely, undershoot in the power rail (or overshoot in the ground rail) reduces noise margins of the transistors causing poor driving capabilities [3]. The parasitic associated with long power distribution rails becomes a key issue under these circumstances.

Nowadays, there are many embedded or invisible devices in the environment. Some of these devices are context-aware i.e. they know about their state or situation. Some of the devices can be personalized or tailor-made to the user’s needs and can recognize the user. Sometimes the devices are adaptive i.e. can change response to the environment or an event. All this leads to More than Moore’s law which is meant to interact with the user and environment. More than Moore’s law is to serve the ambient intelligence requirements i.e. sensitivity and response to human needs and presence. There was a buzz for “More than Moore’s Law” with the introduction of System-on-Packag (SoP) technology. The real movement for “More than Moore’s Law” started with the arrival of three-dimensional (3D) integration technology which tends to focus on system integration. The More than Moore’s approach typically allows for non-digital functionalities such as sensors, actuators, power control, and radio frequency communication [4]. Three-dimensional (3D) integration technology is a strong candidate for More than Moore’s law. The anticipated applications of 3D integration include memory, portable devices, high performance computers, and high-density multifunctional heterogeneous integration of information technology, nanotechnology, and biotechnology systems as shown in Fig. 1.1 [5].

The purpose of 3D integration is either to partition a single chip into multiple strata to reduce global interconnect length or stacking of chips together using TSVs [6]. Increasing the number of strata from one to four reduces the length of the longest interconnect by 50% with 75% improvement in latency and 50% improvement in interconnect energy dissipation.
Using three-dimensional integration, the clock frequency can be increased by 3.9X and area and power can be reduced by 84% [6]. Energy dissipation of long on-chip wires can be reduced by 54% by using three-dimensional interconnects in a 45nm technology node [8]. Three-dimensional integration provides the potential for a tremendously increased level of integration per unit footprint compared to its two-dimensional (2D) counterpart [9].

Fig. 1.1 A vision of future three-dimensional (3D) hyper-integration of information technology, nanotechnology, and biotechnology systems, a new paradigm for future technologies [5].
### Table 1.2 Intel Desktop Processor trends for the last ten years [2].

<table>
<thead>
<tr>
<th>Year</th>
<th>Model</th>
<th>Process</th>
<th>Clock</th>
<th>Transistor Count</th>
<th>Max. TDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>Pentium</td>
<td>0.8um</td>
<td>66 MHz</td>
<td>3.1 million</td>
<td>8W</td>
</tr>
<tr>
<td>1995</td>
<td>Pentium Pro</td>
<td>0.6um</td>
<td>200 MHz</td>
<td>5.5 million</td>
<td>15.5W</td>
</tr>
<tr>
<td>1997</td>
<td>Pentium II</td>
<td>0.35um</td>
<td>300 MHz</td>
<td>7.5 million</td>
<td>43W</td>
</tr>
<tr>
<td>1999</td>
<td>Pentium III</td>
<td>0.25um</td>
<td>600 MHz</td>
<td>9.5 million</td>
<td>42.8W</td>
</tr>
<tr>
<td>2000</td>
<td>Pentium IV</td>
<td>0.18um</td>
<td>2 GHz</td>
<td>42 million</td>
<td>71.8W</td>
</tr>
<tr>
<td>2005</td>
<td>Pentium D</td>
<td>90nm</td>
<td>3.2 GHz</td>
<td>230 million</td>
<td>130W</td>
</tr>
<tr>
<td>2007</td>
<td>Core 2 Duo</td>
<td>65nm</td>
<td>2.33 GHz</td>
<td>410 million</td>
<td>65W</td>
</tr>
<tr>
<td>2008</td>
<td>Core 2 Quad</td>
<td>45nm</td>
<td>2.83 GHz</td>
<td>820 million</td>
<td>95W</td>
</tr>
<tr>
<td>2010</td>
<td>Six-Core Core i7-970</td>
<td>32nm</td>
<td>3.2 GHz</td>
<td>1170 million</td>
<td>130W</td>
</tr>
<tr>
<td>2011</td>
<td>10-Core Xeon</td>
<td>32nm</td>
<td>2.4 GHz</td>
<td>2600 million</td>
<td>130W</td>
</tr>
<tr>
<td>2012</td>
<td>Ivy Bridge Core i5-3570 (3D transistors)</td>
<td>22nm</td>
<td>3.4GHz</td>
<td>-</td>
<td>77W</td>
</tr>
</tbody>
</table>

### Table 1.3 ITRS 2011 predictions about high volume microprocessors for the next decade [1].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Node (nm)</td>
<td>24</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>15.3</td>
<td>14</td>
<td>12.8</td>
<td>11.7</td>
<td>10.6</td>
<td>9.7</td>
<td>8.9</td>
</tr>
<tr>
<td>Transistors/Chip (billions)</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Chip Size (mm²)</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>1260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
</tr>
<tr>
<td>On-Chip Clock (GHz)</td>
<td>3.74</td>
<td>3.89</td>
<td>4.05</td>
<td>4.21</td>
<td>4.38</td>
<td>4.55</td>
<td>4.73</td>
<td>4.92</td>
<td>5.12</td>
<td>5.33</td>
<td>5.54</td>
<td>5.75</td>
</tr>
<tr>
<td>Total No. of Pads</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
<td>3072</td>
</tr>
<tr>
<td>P/G Pads (% of total)</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
<td>66.7</td>
</tr>
<tr>
<td>Vdd (V)</td>
<td>0.9</td>
<td>0.87</td>
<td>0.85</td>
<td>0.82</td>
<td>0.8</td>
<td>0.77</td>
<td>0.75</td>
<td>0.73</td>
<td>0.71</td>
<td>0.68</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>Max. Allowable Power (W)</td>
<td>161</td>
<td>158</td>
<td>149</td>
<td>152</td>
<td>143</td>
<td>130</td>
<td>130</td>
<td>136</td>
<td>133</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
</tbody>
</table>

### 1.1 Motivation for Three-Dimensional (3D) Integration

The migration from two-dimensional (2D) to three-dimensional (3D) integrated circuits brings higher bandwidth, reduced power consumption, and higher integration density [5].
The improved performance, reduced timing, and small form factors are also key drivers for three-dimensional (3D) integration and through-silicon-via (TSV) interconnect technology [5]. The main application areas for three-dimensional (3D) ICs with TSVs are networking, graphics, wireless communication, and computing like multi-core CPUs, GPUs, packet buffers/routers, smart phones, tablets, notebooks, cameras, and DVD players. Three-Dimensional (3D) integration technology has the following features:

- Small form factor and comparatively mechanically robust due to TSVs [5].
- Improved packing density compared to two-dimensional (2D) integration [5].
- Reduced power and latency due to small height of TSVs compared to the length of long interconnecting wires for example in System-on-Chip (SoC) and System-on-Package (SoP) [10].
- Improved power integrity and signal integrity compared to two-dimensional (2D) integrated circuits due to small interconnect length and reduced capacitance, and inductance [10].
- Heterogeneous integration i.e. hybrid technologies like memory, logic, and analog together using different technology nodes [10].
- Improved performance due increased bandwidth in the vertical direction, and small latencies due to short vertical interconnects compared to board latencies.
- Miniaturization in the footprint of package [5][10].
- More options for power distribution in the vertical direction i.e. through inductive or capacitive coupling [11][12][13].
- Attractive for applications like three-dimensional (3D)-Network-on-Chip (3DNoC), and Memory-over-Logic system [5].
- Suitable for fabrication of on-chip inductor or capacitor [5].
- Suitable for fabrication of on-chip voltage regulator module (VRM) for more than one voltage level necessary for mixed technology applications [3].

Three-dimensional (3D) integration is a very attractive option compared to existing design approaches when it comes to size, performance, and flexibility of design but CAD tools and design methodologies are still in the evolution process as shown in Table 1.4 [14]. The
through-silicon via (TSV) has an edge compared to the rest of the three-dimensional assembly technologies due to its small size, high density, and low resistance, as shown in Table 1.5 [14]. The worldwide academic and research activities currently focus on innovation of 3D technology, simulation, design, and product prototypes [5]. Three-dimensional integration using TSVs is one of the future IC packaging technologies that can eliminate the copper wire between separate dies by stacking the dies on top of each other [15]. Through-silicon vias with heights comparable to the substrate thickness, can pass through the substrate and can be placed anywhere in the die, thus providing extra I/O flexibility compared to copper wires, which can only be placed along the periphery of the chip [14].

### Table 1.4 Comparison amongst different design approaches [14].

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Single Chip</th>
<th>System-on-Chip (SoC)</th>
<th>Three-Dimensional (3D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modular flexibility</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>System performance</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Physical dimensions</td>
<td>Large</td>
<td>Medium</td>
<td>Small</td>
</tr>
<tr>
<td>Fabrication complexity</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Design tools</td>
<td>Available</td>
<td>Available</td>
<td>Not deployed yet</td>
</tr>
</tbody>
</table>

### Table 1.5 Comparison amongst three-dimensional (3D) assembly technologies [14].

<table>
<thead>
<tr>
<th>Technology</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wire bonding</td>
<td>-Flexible connections</td>
<td>-Low density</td>
</tr>
<tr>
<td></td>
<td>-High reliability</td>
<td>-Long thin wire</td>
</tr>
<tr>
<td></td>
<td>-Mature processing</td>
<td>-Large pad area</td>
</tr>
<tr>
<td></td>
<td>-Cost effective</td>
<td>-Poor signal integrity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Poor power integrity</td>
</tr>
<tr>
<td>Metal bumps</td>
<td>-Short length</td>
<td>-Large solder balls</td>
</tr>
<tr>
<td></td>
<td>-Low resistance</td>
<td>-May short circuit with each other in the long run</td>
</tr>
<tr>
<td></td>
<td>-More connections</td>
<td></td>
</tr>
<tr>
<td>Through-silicon vias</td>
<td>-Small height</td>
<td>-Complex fabrication</td>
</tr>
<tr>
<td>(TSVs)</td>
<td>-Small footprint</td>
<td>-Capacitive coupling to substrate, devices, and TSVs in vicinity</td>
</tr>
<tr>
<td></td>
<td>-High density</td>
<td>-Mechanical stresses to thin substrate and devices</td>
</tr>
<tr>
<td></td>
<td>-Low resistance</td>
<td></td>
</tr>
<tr>
<td>Contactless i.e. inductive or capacitive coupling</td>
<td>-Small electric path length</td>
<td>-Low reliability</td>
</tr>
<tr>
<td></td>
<td>-Easy to fabricate</td>
<td>-Cross talk and coupling issues</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Size of inductor for inductive coupling</td>
</tr>
</tbody>
</table>
Multiple dies containing digital logic gates are stacked together in the 3D IC approach. On the other hand, the System-on-Package (SoP) approach contains both active and passive components. A thin film is incorporated to embed the passive components in a package rather than on the board in the system. RF, optical, and digital components are integrated together using IC-package-system co-design. The three-dimensional IC approach has a considerable edge in feature size compared to three-dimensional (3D) System-on-Package (SoP) as shown in table 1.6 [16].

<table>
<thead>
<tr>
<th>Feature</th>
<th>3D IC</th>
<th>3D SoP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smallest via size</td>
<td>0.14um</td>
<td>1um</td>
</tr>
<tr>
<td>Largest via size</td>
<td>10um</td>
<td>90um</td>
</tr>
<tr>
<td>Smallest via pitch</td>
<td>0.4um</td>
<td>10um</td>
</tr>
<tr>
<td>Largest via pitch</td>
<td>200um</td>
<td>200um</td>
</tr>
<tr>
<td>Interconnect density</td>
<td>$10^5-10^8$/cm$^2$</td>
<td>$10^6-10^8$/cm$^2$</td>
</tr>
<tr>
<td>Minimum silicon thickness</td>
<td>2um</td>
<td>4um</td>
</tr>
<tr>
<td>Maximum silicon thickness</td>
<td>70um</td>
<td>50-300um</td>
</tr>
</tbody>
</table>

1.2 Challenges to Three-Dimensional Integration

Three-dimensional (3D) integration technology is emerging quickly but with a lot of design, test, and verification challenges. The standard definitions are lacking, and power integrity, thermal integrity, signal integrity, floor planning, architectural analysis, and chip/package co-design capabilities are needed. Some of these capabilities are available whereas the rest are under construction. Three-dimensional (3D) integration technology is facing the following challenges.

- Testing issues, for example, pre-bond testing or post-bond testing. Pre-bond testing means testing each die before bonding. Post-bond testing means testing each die after it has been bonded to the stack [17]. Pre-bond testing involves identifying a known good die (KGD) for stacking whereas post-bond testing entails identifying a damaged die during the bonding process. The challenge for pre-bond testing is that probe equipment may damage the fine pitched micro bumps of the die. The challenge for post-bond testing is that it is not cost effective due to de-bonding of defective dies from the stack.

- Unavailability of CAD and EDA tools for 3D ICs impedes the design of power distribution network which is not overly designed [18].
• Reduced footprint size of 3D ICs degrades the power integrity of the system. The reason is that reduced area leads to higher current density and allows a smaller number of I/O pads for high computing systems [18].

• Switching noise is a key challenge when multiple high performance microprocessor chips are stacked together using TSVs and flip chip technology. Sometimes a several hundred ampere current is required to be delivered using a limited footprint when several dies switch together using the same footprint [19]. In this case, small micro bumps and narrow TSVs exhibit large parasitic inductance while passing switching current.

• Signal integrity is an issue due to the coupling effects of TSVs [20]. TSV to device coupling, TSV to TSV coupling, landing pad to metal coupling, and landing pad to landing pad coupling are potential coupling sources in this case.

• Thermal integrity is a challenging task for multiple high performance dies stacked together [19]. Hundreds of amperes of current pass through a limited footprint when multiple dice switch simultaneously. Heat dissipation is a challenging task because of the limited surface area of 3D ICs.

• Mechanical stress developed by TSV is a critical issue, specifically for thin wafers [21]. Thermo-mechanical stress is developed due to mismatch in coefficient of thermal expansion of TSV material and substrate material. The mechanical stress may affect the device performance or develop cracks in the substrate or the TSV itself.

Table 1.7 shows the latest ITRS predictions for product and design challenges to future 3D integration technology [1].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3D Technologies</strong></td>
<td>Homogeneous stack of silicon using interposers.</td>
<td>Tight integration of memory and logic.</td>
<td>Heterogeneous 3D, monolithic 3D IC.</td>
</tr>
<tr>
<td><strong>Product</strong></td>
<td>DRAM stack with high yield and small size.</td>
<td>Mobile memory-on-logic with significant power saving, and bandwidth enhancement.</td>
<td>Highly integrated and optimized system with no memory wall and cost issues.</td>
</tr>
<tr>
<td><strong>Design Challenges</strong></td>
<td>Power integrity using TSVs, Heat removal, stress caused by TSVs, Standards and formats for chip-package co-design for thermal and power integrity, cost, and yield.</td>
<td>-Power integrity and IR drop with TSVs to 10mV accuracy. -Thermal, stress and switching noise driven transients.</td>
<td>-More than 100A current delivery with 10mV accuracy. -Complex tradeoffs for heterogeneous system of more than ten dies.</td>
</tr>
</tbody>
</table>
1.3 Road Map for 3D Interconnects

There are different levels of three-dimensional (3D) interconnects i.e. packaging, bonding, global, intermediate, and local interconnects like conventional two-dimensional (2D) interconnects. Through-silicon-vias (TSVs) are used at all levels except the packaging level and local level interconnections. There are different ways of fabricating through-silicon via depending on the interconnect level at which TSVs are introduced, as shown in Table 1.8 [1]. TSVs have different roadmaps when they are introduced to the global interconnect level and intermediate interconnect level for 3D SIC/3D-SoC, as shown in Table 1.9 , and Table 1.10 respectively [1].

<table>
<thead>
<tr>
<th>Level</th>
<th>Name</th>
<th>Supply Chain</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Package</td>
<td>3D Packaging</td>
<td>Assembly and PCB</td>
<td>-Wire bonded die stacks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Package-on-Package (PoP) Stacks</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-No TSVs</td>
</tr>
<tr>
<td>Bond Pad</td>
<td>3D-Wafer-Level Packaging</td>
<td>Wafer Level Packaging (WLP)</td>
<td>-Via last process i.e. 3D interconnects are processed after the IC fabrication</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-TSV density follows the bond-pad density</td>
</tr>
<tr>
<td>Global</td>
<td>3D Stacked Integrated Circuit (3D SIC/SOC)</td>
<td>Wafer Fabrication</td>
<td>-Stacking large circuit blocks on different layers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-TSV pitch 4-16um</td>
</tr>
<tr>
<td>Intermediate</td>
<td>3D SIC</td>
<td>Wafer Fabrication</td>
<td>-Stacking smaller circuit blocks on different layers i.e. wafer to wafer stacking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-TSV pitch 1-4 um</td>
</tr>
<tr>
<td>Local</td>
<td>3D Integrated Circuit (3D-IC)</td>
<td>Wafer Fabrication</td>
<td>-Stacking transistor layers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Follows intermediate level interconnect density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1.8 3D Interconnect Hierarchy Roadmap [1].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Package</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Bond Pad</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Global</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Local</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 1.9 Emerging Global Interconnect Level 3D-SIC/3D-SoC Roadmap for TSVs [1].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. Diameter (um)</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Min. Diameter (um)</td>
</tr>
<tr>
<td>Min. Pitch (um)</td>
</tr>
<tr>
<td>Min. Height (um)</td>
</tr>
<tr>
<td>Max. Aspect Ratio (AR) (height/diameter)</td>
</tr>
<tr>
<td>No. of Dies/Stack</td>
</tr>
</tbody>
</table>
Table 1.10 Emerging Intermediate Interconnect Level 3D-SIC Roadmap for TSVs [1].

<table>
<thead>
<tr>
<th>Intermediate Level</th>
<th>2011-2014</th>
<th>2015-2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min. Diameter (um)</td>
<td>1-2</td>
<td>0.8-1.5</td>
</tr>
<tr>
<td>Min. Pitch (um)</td>
<td>2-4</td>
<td>1.6-3</td>
</tr>
<tr>
<td>Min. Height (um)</td>
<td>6-10</td>
<td>6-10</td>
</tr>
<tr>
<td>Max. Aspect Ratio (AR) (height/diameter)</td>
<td>5:1-10:1</td>
<td>10:1-20:1</td>
</tr>
<tr>
<td>No. of Dies/Stack</td>
<td>2-5</td>
<td>8-16 (DRAM)</td>
</tr>
</tbody>
</table>

1.4 Thesis Contribution and Objectives

The power integrity of the integrated circuits is the most critical issue [22][23]. Power integrity means that required voltage should be guaranteed for the desired operation of each and every transistor of the integrated chip. The present trend towards voltage scaling, higher speed, and higher integration density increases the drive for power integrity. Power integrity mainly depends on resistive or IR-drop and Ldi/dt noise. The inductance of the power distribution grid is prominent compared to resistance because of high speed and scaling of the on-chip interconnects. The switching of I/O drivers and core switching logic cause voltage fluctuations in the supply rails which result in simultaneous switching noise. The supply to a three-dimensional (3D) stack of dies is distributed through the board to the bottom of the stack. Similarly, the clock and other signals should also be distributed through the bottom of the stack. Consequently, there is limited footprint available for power and ground pins.

The vertical interconnects in 3D PDN, for example through-silicon vias, solder bumps, and pads used for power distribution have additional resistance and inductance. The load on power supply increases with the addition of each chip in the vertical stack. All these make the power integrity for 3D ICs more complex than the 2D ICs. The noise caused by switching of the internal logic load is a more critical issue than I/O driver noise for a three-dimensional (3D) stack of dies for two reasons. The first is that the speed of internal logic switching is considerably higher than the I/O driver speed for high computing 3D ICs. The second reason is that the size of the board is very small owing to small dimensions of 3D compared to its 2D counterpart. The problem of core switching noise is further aggravated for three-dimensional (3D) stacked processors where logic dies are interconnected using TSVs.

1.4.1 Contribution

Simulation of on-chip power distribution networks is a critical problem for traditional simulation tools due to the large size of the power distribution network. The traditional simulation tools fall short of speed and memory resources even during simulation of a large
2D power distribution network. The size of a 3D power distribution network is several times that of a 2D power distribution network. There may be billions of nodes in a 3D stacked power distribution network. It is very difficult to trace the nodes of such a huge power distribution network in a SPICE simulation. The fast transient simulation algorithm proposed in the thesis addresses the same problem for 3D stacked power distribution networks. The proposed algorithm has +/- 1 to 2% deviation compared to Ansoft Nexxim4.1. The proposed algorithm is several times faster than Ansoft Nexxim and also requires significantly less memory than Ansoft Nexxim. The speed is improved and less memory is required because of a significant reduction in the number of nodes compared to Ansoft Nexxim as shown by chapter 5.

The power distribution TSV pair is the fundamental part of a 3D PDN. On-chip logic load is assigned to power distribution TSV pairs. Each power distribution TSV pair is assumed to have logic load and decoupling capacitance as shown in chapter 4. A mathematical model is proposed for the optimal design of the power distribution TSV pair to maintain a certain level of switching noise based on the worst case rise time. The model has +/- 5% deviation compared to Ansoft Nexxim, as shown by chapter 5. This model is simple to implement using Matlab. The proposed model provides guidelines for the physical design of a power distribution TSV pair based on the on-chip load requirements. In addition to this, it is useful for optimal adjustment of power distribution TSV pair design and associated decoupling capacitance for specific load requirements.

A power TSV is accompanied by a ground TSV for the return current path in 3D PDN as shown in chapter 4. Power distribution TSVs are placed one above the other within a vertical stack of dies. In this way power distribution TSV pairs make vertical chains, as shown in chapter 5. The supply is distributed only through the base on a printed circuit board to such a vertical chain. The core switching current follows a hierarchical current path through the vertical chain. Such a vertical chain of power distribution TSV pairs should be analyzed for core switching noise in order to ensure required voltage level at each of the logic load. A fast and accurate mathematical model is proposed for the estimation of core switching noise along a vertical chain of power distribution TSV pairs, as shown in chapter 5. The proposed mathematical model is based on a physical model. The model is accurate with 2-3% deviation compared to Ansoft Nexxim. In addition to this, the model is 3-4 times faster than Ansoft Nexxim while requiring half of the memory. The proposed model is very useful for early estimation of the core switching noise along a chain of power distribution TSV pairs from bottom to top in a three-dimensional power distribution network.

The optimum design of decoupling capacitance based on optimum output impedance is required to control core switching noise for each stage of the vertical chain of power
distribution TSV pairs. The proposed design guidelines are useful for the selection of the value of decoupling capacitance along with effective series resistance (ESR), and effective series inductance (ESL) for each stage of the chain of power distribution TSV pairs. First of all the target impedance for each stage of the vertical chain of power distribution TSV pairs is determined according to allowed ripple and load at each stage. An equation is given to estimate the output impedance. The output impedance should be kept less than or equal to the target impedance. There are separate equations for the estimation of ESR and ESL for the decoupling capacitance, as shown in chapter 5. Finally, an equation is given for the estimation of the value of decoupling capacitance. The proposed technique is applicable to a large three-dimensional (3D) power distribution network containing multiple tiers, and multiple vertical power distribution chains.

The 3D-DRAM-Over-Logic system is a promising application of 3D integration technology. The issue is how much decoupling capacitance should be placed on each DRAM die without overdesigning and at the same time saving useful areas of the 3D-DRAM stack. This issue is comprehensively addressed by the proposed model of decoupling capacitance for a 3D-DRAM-Over-Logic system, as shown in chapter 5. Power distribution TSV pairs have to pass through the DRAM stack in order to supply the processor stack in a 3D-DRAM-Over-Logic System. The physical size of the power distribution TSV pairs is also important because TSVs block useful areas of the DRAM stack. Similarly, the foot-print size of the power distribution TSV pairs occupies useful area on the processor die. The resistance of a TSV depends on its physical size. The area occupied by decoupling capacitance on a DRAM die depends on the amount of decoupling capacitance i.e. a small amount of decoupling capacitance occupies a small area on the DRAM die. The mathematical model for decoupling capacitance for the power integrity of a 3D-DRAM-Over-Logic-System answers the above questions. The optimum budgeting of the decoupling capacitance is based on the resistance of the power distribution TSVs in the proposed model. It is of utmost importance to maintain the required voltage across logic load during switching of the load.

The author developed three different mathematical models for the estimation of core switching noise in 3D PDN, as shown in chapter 5. A mathematical model for the time domain analysis of ground noise based on worst case switching is also proposed. Design guidelines for the decoupling capacitance for 3D PDN are proposed. In addition to this, a mathematical model for the optimum budgeting of decoupling capacitance for 3D-DRAM-Over-Logic system is also proposed. The model gives the tradeoff between the amount of decoupling capacitance placed on DRAM dies and the value of the resistance of power distribution TSV pair. Techniques for fast, accurate, and memory efficient modeling of the core switching noise for 3D PDN, decoupling capacitance design guidelines for 3D PDN,
and optimum decoupling capacitance budgeting for a 3D-DRAM-Over-Logic system are the main outcomes of the research presented in this thesis.

1.4.2 Objectives
The core switching noise is a critical issue for 2D power distribution networks owing to ultra large scale integration and circuit speeds of the order of GHz [23]. The problem becomes more serious for high computing 3D stacked power distribution networks because of limited footprint size. The issue of core switching noise for 3D stacked power distribution networks is the prime focus of this thesis. Actual behavior of on-chip logic load is not known at the time of 3D stacked power distribution network design. Development of design tools for the power integrity of 3D integrated circuits is still in progress. This thesis provides fast and accurate models for the design of 3D stacked power distribution networks based on core switching noise. The objective of the proposed models is to provide estimation and early design tradeoffs for the design of 3D stacked power distribution networks.

Output impedance of the power distribution network should be well below the target impedance within the operating frequency range in order to guarantee the power integrity of the system. It is important to control the output impedance of 3D stacked power distribution network at each stage of the 3D stacked dies. The guidelines for the design of decoupling capacitance proposed in this thesis fulfill the above stated objective.

Optimum budgeting of on-chip decoupling capacitance is necessary because of large chip area required for decoupling capacitance. 3D-DRAM-Over-Logic is a promising application of 3D integration technology. Decoupling capacitance should be placed on DRAM dies instead of processor dies in order to save the useful area of processor dies. There should be optimum budgeting of decoupling capacitance placed on DRAM dies. The tradeoff between resistance of power distribution TSV pairs and amount of decoupling capacitance is critically important to establishing the power integrity of the 3D-DRAM-Over-Logic system. The mathematical model proposed in this thesis for the decoupling capacitance of the 3D-DRAM-Over-Logic system fulfills the above stated objective.

This thesis introduces the issue of core switching noise for on-chip 3D stacked power distribution networks. Different modeling techniques for the estimation of core switching noise for 3D stacked power distribution networks are proposed. Different structures used for 3D integration are introduced and briefly discussed. On-chip decoupling capacitance is briefly explained in conjunction with resonance spikes. Different prevailing modeling techniques for on-chip power distribution networks including the research results are briefly discussed.
1.5 Overview of the Thesis Structure

Here we present a brief summary of each of the chapters that follow in order to provide a quick glance at the rest of the thesis.

**Chapter 2.** State of the art 3D structures are presented in this chapter. There are different approaches and patterns for vertical stacking of ICs. Similarly, there are different ways of bonding the integrated circuits depending on which side should be facing the other. So far, the TSV is the most viable option for vertical interconnection. There are different methods used for growing TSVs in the substrate. The other bonding methods for 3D, for example bonding wire and micro bumps were already in use before the introduction of TSVs. The chapter starts with the description of different stacking approaches. Different bonding techniques are used for vertical integration. Bonding is different from true vertical integration. Active devices are fabricated in the vertical direction and from a single package containing different dies in true vertical integration. The chapter discusses bonding methods. Different TSV techniques are also described. Finally, different vertical interconnect implementation techniques are discussed. Then different bonding approaches are described, for example face-to-face, face-to-back, and back-to-back methods. Different TSVs methods are also described, for example via-first, via-last, and via-middle approaches. Finally, vertical interconnect approaches are discussed, for example bonding wire, micro bump, and TSV. Wireless interconnect approaches are also described, for example, capacitive coupling and inductive coupling.

**Chapter 3.** On-chip decoupling capacitance is discussed in detail in this chapter. The basic function of decoupling along with its impedance characteristics under self resonance condition are briefly discussed. The impact of decoupling capacitance on the output impedance of a power distribution network is also discussed. In addition to this the use of decoupling capacitance to maintain the impedance of the power distribution network below the target impedance is also described.

**Chapter 4.** The power integrity problem evolves with the shift from a two-dimensional to a three-dimensional power distribution network. Core switching noise is the determining factor for the power integrity of 3D high computing systems. The core switching noise along with its impact on a power distribution network is discussed in this chapter. Modeling of on-chip logic load for core switching noise is also described. Finally, the approach of the thesis towards robust 3D PDN design is briefly described.

**Chapter 5.** Present-day on-chip power distribution networks may have billions of nodes. It is very complex to simulate such huge networks using SPICE because of the complexities associated with tracking each node. In addition to this, SPICE requires a huge amount of
computation power and takes a very long time to complete the simulation. The simulation and modeling techniques for on-chip power distribution networks have always been in the process of evolution. The problem has been further aggravated with the introduction of 3D stacked integrated circuits. The simulation and modeling tools for 3D power distribution networks have not yet matured. Common modeling and simulation approaches for on-chip power distribution networks are presented in this chapter. Most importantly research results are discussed in connection with each other in order to elaborate the thesis contribution. In addition to this, it is shown how the thesis contribution is positioned in connection with the previous work in this chapter.

**Chapter 6.** Finally a brief summary is presented along with future directions in this chapter.
Chapter 2
Structures for 3D Integration

Despite the recent buzz about 3D integration in industry and academia, the concept is not new as 3D packaging using wire bonded dies has been around for years, for example, many modern cell phone devices use wire bonded 3D stacking [24]. The theoretical studies even in the 1980s and 1990s predicted reduced signal delay and also reduced power consumption using 3D ICs [25]. However, 3D integration was unable to receive due attention at that time owing to the smooth scaling of 2D technology. In addition to this there was no capability to thin the chips to less than 50um and to insulate the deep cuts etched through the thinned chips at that time [26]. However, 3D integration is being reconsidered today given the myriad problems in deep-submicron billion-transistor circuits. The current studies have shown that 3D integration technology is essential and probable rather than desirable and possible.

The through-silicon via (TSV), was invented more than fifty years ago but it was not intended for 3D integration at that time [27]. The technology of stacking chips through TSVs was proposed in 1998 [28]. The TSV is the heart of modern 3D integration, and stacking bare dies using TSVs is a very promising technique today, as shown in Fig. 2.2 (b) [24]. Power distribution through a 3D stack of dies using TSVs is a viable option [20][29]-[35]. The major candidates for using 3D ICs with TSVs would be those that require speed, bandwidth, and power optimization; for example, 3D integration has been used in image sensors [36].

Samsung introduced a 3D stack of memory chips, as shown in Fig. 2.1 [37]. The total thickness of the eight-chip memory stack is 560um, which is still less than the thickness of an ordinary chip [37]. A 3D stack of memory over processor may become the mainstream application of 3D integration, but processor dies without 3D stacking may not be able to fully avail the bandwidth that a TSV-based 3D stacked memory would provide. Recently,
IMEC with its 3D integration partners demonstrated a commercial DRAM chip on top of a logic IC for next-generation low-power mobile applications [38]. Another tailor-made application for 3D integration is for embedded memory in multi-core processors [39]. The high performance 3D integrated microprocessor has both power and performance benefits compared to its 2D counterpart [40].

2.1 Three-Dimensional Stacking Approaches
There are three common approaches for stacking in the vertical direction i.e. stacking of bare dies together, stacking the dies to wafer, and stacking of wafers together. Each of the approaches has certain advantages and drawbacks depending on the application.

2.1.1 Die-to-Die (D2D) Stacking
The dies are cut from different similar/dissimilar wafers. Normally, the dies are tested individually before stacking together. The dies can be stacked using wire bonding, solder bumps or TSVs, as shown in Fig. 2.2 [24]. The dies can be mixed; for example, a faster memory die can be matched to the faster logic die. Die-to-Die stacking can be a good option for heterogeneous integration where the wafer sizes may be significantly different. But alignment of the dies can be a major problem in this approach. This approach is not cost effective compared to the wafer-to-wafer stacking. However, known good dies (KGD) can be stacked together using this approach.

![Fig. 2.2 Die-to-Die (D2D) stacking approaches: (a) Die-to-Die wire bonding. (b) Die-to-Die TSV bonding [24].](image)

2.1.2 Die-to-Wafer (D2W) Stacking
Die-to-wafer stacking is almost similar to die-to-die stacking except that dies are stacked to a substrate as shown in Fig. 2.3 [41]. Die-to-wafer stacking has the advantage of flexible use of known good dies [42]. Known good dies are stacked over a substrate wafer; for example, using die-cavity technology [41]. It is suitable for common die sizes.

2.1.3 Wafer-to-Wafer (W2W) Stacking
The wafers are stacked together in Wafer-to-Wafer stacking as shown in Fig. 2.4 (a). Wafer-to-Wafer stacking has the advantage of low cost [35]. Some good dies may be forced to
bond to bad dies as shown in Fig. 2.4 (b). The yield can be a big issue for wafer-to-wafer stacking [43]. This approach is suitable for common wafer size.

Fig. 2.3 Die-to-Wafer (D2W) stacking [41].

Fig. 2.4 (a) Wafer-to-Wafer stacking. (b) A defective die may stack to good dies, thereby reducing the yield in Wafer-to-Wafer stacking.

2.2 Three-Dimensional Bonding Approaches

The dies are bonded in a 3D stack using different approaches, such as Face-to-Face (F2F) bonding, Face-to-Back (F2B) bonding, and Back-to-Back (B2B) bonding. Two dies face each other in the case of F2F bonding whereas the face of one die is to be joined to the back of the other die in the case of F2B bonding.
2.2.1 Face-to-Face (F2F) Bonding
The active layers of two dies are facing each other in Face-to-Face bonding, as shown in Fig. 2.5 [44]. Although Face-to-Face bonding allows only two dies to be bonded, but can provide higher TSV density due to similarity of synthesizing on die interconnects [45]. F2F bonding technique can be very useful for Wafer-to-Wafer (W2W) bonding in order to avoid any through hole through the delicate wafer.

![Back-side metal](image)

*Fig. 2.5 Face-to-Face (F2F) bonding [44].*

2.2.2 Face-to-Back (F2B) Bonding
The active layer of one die faces the back or substrate side of the other die in Face-to-Back bonding, as shown in Fig. 2.6 [44]. However, any number of dies can be stacked using the face-to-back approach but TSV density is less than its face-to-face counterpart as TSV must be etched through the back of the die and the etching process has low resolution compared to the face-to-face bonding approach [44]. The face-to-back bonding technique is suitable for die-to-die or die-to-wafer bonding.

2.2.3 Back-to-Back (B2B) Bonding
The substrate sides of both the dies are facing each other in Back-to-Back bonding, as shown in Fig. 2.7 [44]. It is rarely used because of the long vertical interconnect between a pair of neighboring dies.
Fig. 2.6 Face-to-Back (F2B) bonding [44].

Fig. 2.7 Back-to-Back (B2B) bonding [44].
2.3 Three-Dimensional TSV Approaches

The introduction of TSVs in the bonding process is a key factor. The TSV process has three approaches based on the stage of introduction of TSVs during the fabrication process. If TSVs are introduced quite early in the process even before devices are built, it is a called via-first process [44]. If TSVs are introduced after bonding and thinning then it is called a via-last process [44]. FEOL is fabrication of CMOS at high temperature. BEOL is adding insulator layers and making connections between CMOS and TSV. If TSVs are introduced in between the FEOL and BEOL processes then it is called via middle process.

In the via-first approach, the TSV is etched and filled first and then the transistor is fabricated at high temperature. In the next process phase, metal contacts are made, the back is thinned using a carrier, and the carrier is debonded at the end. Devices and TSVs are fabricated before deposition of metal layers and bonding of dies in the via-first approach, as shown in Fig. 2.6 [44] and Fig. 2.8 [44]. In via first technology, the TSVs are surrounded laterally by other TSVs and vertically by wires [44].

![Fig. 2.8 Via-first TSV technology [44].](image-url)

![Fig. 2.9 Via-last TSV technology [44].](image-url)
In the via-last approach, FEOL takes place at high temperature, then BEOL takes place at relatively low temperature. Then thinning takes place, after this TSV etching and filling, and finally de-bonding of the carrier takes place.

Fig. 2.6 shows the comparison of commonly used via-first and via-last approaches [44]. Devices and metal layers are fabricated first, TSVs are fabricated through all the metal layers from substrate to the topmost layer, and then dies are bonded in via-last technique, as shown in Fig. 2.7 [44] and Fig. 2.8 [44]. The TSVs in the via-last approach are surrounded laterally by other TSVs and laterally as well as vertically by wires [44].

In the via-middle approach the vias are made between FEOL (fabrication of CMOS at high temperature) and BEOL and then TSV is etched and filled [46].

2.4 Three-Dimensional Vertical Interconnect Approaches

Vertical interconnects like wire bonding and solder bumps are already in use for example in the mobile phone industry. TSV as a vertical interconnect is considered to be a key enabling technology for modern 3D integration technology for More than Moore’s applications. Wireless interconnects are also under consideration due to extra parasitic introduced by TSVs in the vertical direction.

Wire bonding has traditionally been used for 3D stacking of dies. Now there is a shift from wire bonding mainly for two reasons. The wires have high inductance at the frequencies of the order of GHz. The wires have limited I/O connections as wires can only be connected to the periphery of the chip. Two dies inter-connected through bonding wires are shown in Fig. 2.10 [11].

Wire bonds are replaced by solder bumps in order to provide low resistance and inductance at high frequency. An array of small solder bumps is adhered to I/O pads of the chip where I/Os are distributed across the whole surface of the chip. In addition to this the
solder bumps provide more I/Os than the wire bonds directly under the chip. A 3D stack of five dies interconnected through micro-bumps is shown in Fig. 2.11 [11].

![Micro-bumps](image)

**Fig. 2.11** The dies interconnected through micro-bumps [11].

Power distribution using TSVs is a viable option owing to low TSV parasitics compared to the long global interconnects. Fig. 2.12 shows the cross section of a three-tier power distribution network with a power distribution TSV pair. Through-Silicon-Vias (TSVs) pass through the substrate and active circuit layers. Fig. 2.13 shows top view of a 3D PDN with power distribution TSV pairs [47]. The logic load on each chip is divided into cells where the cell is defined as a part of the logic load supplied by a TSV pair in its vicinity. Each tier has paired and orthogonal power distribution buses placed at a constant distance from each other as shown in Fig. 5.14. The interconnection between two neighboring tiers is made through the power distribution TSV pairs which, is shown in Fig. 2.12.

![Cross section](image)

**Fig. 2.12** Cross section of a 3D PDN having three planes and a TSV pair.

Nano-photonics allow for thousands of components on a chip. Some of the components may need tuning or monitoring, which requires electronics functions on a similar scale. The integration of photonics and electronics is, therefore, required in a way that minimizes the limitations set to both the electronics and photonics chips. The tight integration of the nano-photonics circuits with the electronics circuits is made through Cu-nail TSVs, as shown in
Fig. 2.14 [48]. The potential of nano-photonics circuits can be tapped using this technique; for example, in telecommunications, sensing, or biomedical applications.

Fig. 2.13 Top view of a 3D power distribution network [47].

Fig. 2.14 3D integration of nano-photonics and CMOS using Cu-Nails (TSVs) [48].
Despite its benefits, TSV introduces additional complexities to the fabrication process; therefore, inductive and capacitive coupling links have recently been explored instead of TSVs [11]. The most common wireless interconnects are capacitive and inductive links [11]. The metal solenoids and capacitor plates are easy to fabricate using the top metal layer of the integrated circuit. The signals are linked through the electric field in the capacitive coupling, as shown in Fig. 2.15 [12]. The signals are linked through the magnetic field in the inductive coupling, as shown in Fig. 2.16 [12]. The capacitive coupling is favorable for Face-to-Face bonding because of the shorter range compared to the inductive coupling [12]. Inductive coupling has been used for memory-over-processor systems for a comparatively long communication distance of 120um in [13].

![Fig. 2.15 Electro-static coupling through on-chip capacitors [12].](image1)

![Fig. 2.16 Electro-magnetic coupling through on-chip inductors [12].](image2)
Decoupling capacitance can store the charge and release it when required. In other words it is like a charge reservoir. Decoupling capacitance is placed between the supply and ground rails of a power distribution network. It is normally placed close to the load in order to enable it to supply the switching current easily. Decoupling capacitance is discharged during switching and it is charged again through the power supply. The charge stored on it is given by \( Q = CV \), where \( Q \) is the charge stored on the decoupling capacitance, \( C \) is the capacity of the decoupling capacitance, and \( V \) is the supply voltage. Decoupling capacitance should be able to provide sufficient charge and discharge at high rates in order to supply load at high frequency. Decoupling capacitors are used to control \( \text{Ldi/dt} \) noise and have a significant effect on speed, cost, and power of integrated circuits (ICs). Decoupling capacitors provide sufficient charge to switching loads for short and periodic intervals of time. Physically a decoupling capacitance serves as an intermediate storage of charge which is closer to the switching circuit.

3.1 Impedance of the Decoupling Capacitance:

Ideally a decoupling capacitance should have no series resistance and series inductance but practically it always has series resistance and inductance, as shown in Fig 3.1. The impedance of decoupling capacitance is frequency dependent.

The capacitive reactance of the decoupling capacitance is given as follows:

\[
X_c = \frac{1}{\omega C}
\]

The inductive reactance of decoupling capacitance is given as follows:

\[
X_L = \omega L
\]

Fig. 3.1 Equivalent electrical model of a decoupling capacitance, where ESR is effective series resistance and ESL is effective series inductance of decoupling capacitance.
The impedance vs. frequency variation of the decoupling capacitance is shown in Fig. 3.2 [49]. The impedance decreases by increasing the frequency in the beginning when the frequency is less than the resonance frequency. It is because of the fact that capacitive reactance is dominant before resonance frequency. The impedance increases beyond resonance frequency. It is because of the fact that inductive reactance due to ESL becomes dominant at frequencies higher than the resonance frequency. The capacitive reactance becomes equal to inductive reactance at resonance frequency. It is a change over point where impedance of the decoupling capacitance changes its nature from capacitive to inductive. The impedance of the decoupling capacitance is equal to ESR at this point i.e.

\[
\begin{align*}
X_L &= X_C \\
\omega_{res}L &= \frac{1}{\omega_{res}C}
\end{align*}
\]

\[\Rightarrow \omega_{res} = \frac{1}{\sqrt{LC}}\]

Where L is effective series inductance (ESL) of decoupling capacitance, C is the amount of decoupling capacitance, and \(\omega_{res}\) is resonance frequency of decoupling capacitance.

Fig. 3.2 V-curve showing impedance of the decoupling capacitance vs. frequency [49].

The impedance of decoupling capacitance has minimum value at resonance frequency and is decided by the value of the effective series resistance (ESR) of the decoupling capacitance. The effective series resistance (ESR) depends on the resistance of the metal lead, and the resistance of conductive plates. In the case of on-chip metal oxide semiconductor (MOS) decoupling capacitor, the effective series inductance (ESL) depends on the area of the current loop i.e. effective series inductance is larger when the area of the current loop is larger. The capacitive curve moves down and to the right when the amount of decoupling capacitance is increased i.e. capacitive reactance is inversely proportional to the amount of
decoupling capacitance. The inductive curve remains unchanged if the effective series inductance of the decoupling capacitance is fixed. When similar types of decoupling capacitors with different values are placed, then only the capacitive curve moves and the inductive curve remains fixed at its position. The impedance of the decoupling capacitance for a chip or package is reduced by increasing the amount of decoupling capacitance if it is operated below resonance frequency. The V-curve moves up if the effective series resistance is increased and moves down if the effective series resistance is decreased. The effective series resistance and inductance is drastically reduced if the decoupling capacitance has a number of similar decoupling capacitors connected in parallel. This is actually an effective technique to control the effective series resistance (ESR) and effective series inductance (ESL) of the decoupling capacitance. This technique also indirectly increases the resonance frequency of the decoupling capacitance by lowering the effective series inductance of the decoupling capacitance. Decoupling capacitors with low effective series resistance (ESR), and low effective series inductance (ESL) can effectively maintain a low impedance profile for on-chip power distribution networks.

3.2 Impact of Decoupling Capacitance on Target Impedance

Gates switching at GHz frequency in modern high speed integrated circuits (ICs), create transient currents in the power distribution network. Decoupling capacitors can reduce the impedance of a power distribution network by reducing the current loop, thereby shunting high frequency currents. The impedance of a power distribution network can be reduced over a significant range of frequencies by using decoupling capacitance. However, using more decoupling capacitance renders a power distribution network susceptible to resonance as decoupling capacitance lowers the resonance frequency of the circuit. The power distribution network should exhibit minimum impedance as seen from the terminals of the current load in order to ensure power integrity. The impedance of the power distribution network of an integrated circuit (IC) should therefore be maintained below a certain level known as the target impedance. The impedance of PDN equal to the target impedance should be maintained within the operating frequency range in order to ensure the power integrity, where the target impedance is given as follows:

\[
Z_{\text{target}} = \frac{V_{dd} \cdot V_{\text{ripple}}(\%) }{I_r} 
\]

where \( V_{dd} \) is nominal supply voltage, \( V_{\text{ripple}} \) is maximum allowable ripple in the supply voltage, and \( I_r \) is maximum transient current on a clock edge. The power distribution system should be designed so that the impedance presented to the active circuits is less than or equal to the target impedance. By putting \( I_r = \frac{P}{V_{dd}} \) in (3.1) we get:

\[
Z_{\text{target}} = \frac{V_{dd}^2 \cdot V_{\text{ripple}}(\%)}{P} 
\]
where $P$ is peak transient power over a clock edge. By putting $P = CV_{dd}$ in (3.2) we get:

$$Z_{\text{target}} = \frac{V_{dd} \cdot V_{\text{ripple}}(\%)}{C}$$

The above equation shows that target impedance decreases by adding more decoupling capacitance for a fixed supply voltage and fixed ripple. The decoupling capacitor is placed across the power and ground rails to meet the target impedance for a specified frequency range [50]. Practically, the decoupling capacitance is not a pure capacitance at high frequency because of the intrinsic effective series inductance and effective series resistance. The impedance of the decoupling capacitance is inductive above the resonance frequency and decoupling capacitance is not as effective as desired above this frequency. Fig. 3.3 (a) shows that the impedance of a power distribution network is resistive at low frequency, whereas, it increases linearly with the frequency for higher frequencies due to the dominance of the inductive reactance of the network [49]. There is a maximum frequency $\omega_{\text{max}}$ at which the network impedance exceeds the target impedance.

![Figure 3.3](image_url)

**Fig. 3.3** (a) Frequency response of the impedance of a power distribution network without decoupling capacitance. (b) Frequency response of the impedance of a power distribution network with decoupling capacitance [49].
Fig. 3.3 (b) shows that the impedance of the network shoots up at the resonance frequency by using decoupling capacitance compared to the no decoupling capacitance case [49]. It is because of the parallel resonance caused by the LC tank circuit which in turn causes the maximum impedance. However, above this frequency, the impedance starts increasing linearly with frequency because of the dominance of the inductive reactance of the network at high frequency. Fig. 3.3 shows that the target impedance is reached at a higher frequency when using decoupling capacitance compared to the no decoupling capacitance case. Decoupling capacitance is used to increase the frequency at which the impedance of the power distribution network exceeds the target impedance. The impedance of a decoupling capacitance is equal to the effective series resistance of a capacitor at the resonance frequency. A logic gate that is not switching connects its output load capacitance to either the positive supply or ground [51]. Most of the time these output loads serve as symbiotic bypass capacitors that help maintain the supply voltage during current transients [51]. The method of calculating the symbiotic bypass capacitance is also given by [51]. The symbiotic bypass capacitance enhances the strength of the intentional on-chip decoupling capacitance. However, too large decoupling capacitance reduces the resonance frequency of the power distribution network. There should be a tradeoff between the resonance frequency and amount of decoupling capacitance, depending on specific design requirements.

3.3 Anti-resonance and parallel decoupling capacitors

Multiple decoupling capacitors are used in parallel in order to lower the impedance of a power distribution network. But a parallel bank of decoupling capacitors is always prone to anti-resonance. The impedance of the circuit increases drastically during anti-resonance. Anti-resonance causes failure of the power distribution network, thereby pushing the impedance of the power distribution network above maximum tolerable target impedance, as shown in Fig. 3.4 [49].

![Anti-resonance Spike](image)

Fig. 3.4 Distinctive peak of output impedance of a power distribution network due to anti-resonance [49].
Anti-resonance occurs in the overall parallel bank of capacitors due to each decoupling capacitance having a different self-resonance frequency than the other. Anti-resonance occurs between two parallel capacitors when one is still capacitive while the other has become inductive. Anti-resonance between two parallel decoupling capacitors $C_1$ and $C_2$ is shown in Fig. 3.5 [52]. The impedance of $C_1$ is inductive whereas the impedance of $C_2$ is capacitive in the frequency range from $f_1$ to $f_2$. The impedance peak of the LC tank circuit is in between $f_1$ and $f_2$ or $f_2$ and $f_3$. The magnitude of the anti-resonance spike depends on the effective series resistance (ESR) and effective series inductance (ESL). Anti-resonance frequency depends on the effective series inductance (ESL) of decoupling capacitors. Larger effective series resistance (ESR) of the decoupling capacitors lowers the peak of the anti-resonance spike by acting as a damping element. Decreasing the effective series inductance (ESL) of decoupling capacitors has the same effect on the magnitude of the anti-resonance spike as that achieved by increasing the amount of effective series resistance (ESR). The location of the anti-resonance spike depends on the ratio of the effective series inductance (ESL) of the decoupling capacitors. Assume $C_1=C_2$, then anti-resonance occurs between $f_1$ and $f_2$ provided the ESL of $C_1$ is greater than the ESL of $C_2$. When the ESL of $C_1$ is lower than the ESL of $C_2$, anti-resonance will occur between $f_2$ and $f_3$ i.e. higher than the self-resonance frequency of $C_2$.

![Fig. 3.5](image)

**Fig. 3.5** Anti-resonance of two equal value parallel decoupling capacitors i.e. $f_{ar1}<f_{ar2}$ as $f_{ar1}$ occurs when $L_1 > L_2$, and $f_{ar2}$ occurs when $L_1 < L_2$ [52].

The quality factor $Q=L/R$ of decoupling capacitors plays an important role in determining the magnitude and frequency of anti-resonance spikes i.e. the magnitude of the anti-resonance spike increases by increasing the quality factor of decoupling capacitors. The frequency for the anti-resonance spike decreases by increasing the quality factor $Q$ of decoupling capacitors. There are two options to reduce the quality factor i.e. increase the amount of effective series resistance (ESR) or decrease the effective series inductance (ESL).
of on-chip decoupling capacitors. Decreasing the ESL of on-chip decoupling capacitors is better option as increasing the ESR may increase the resistance of the power distribution network above the required level. The magnitude of anti-resonance spikes in a 3D power distribution network can be lowered by using decoupling capacitors having lowest ESR values along with power distribution TSV pairs with the lowest self-inductance. Large variation in the magnitude of parallel decoupling capacitors results in sharp anti-resonance spikes [52]. Similar decoupling capacitors should, therefore, be connected in parallel in order to avoid on-chip anti-resonance.

3.4 Effective On-Chip Decoupling Capacitance

Different on-chip capacitances contribute to the overall on-chip decoupling capacitance. On-chip decoupling capacitance has two components i.e. intrinsic decoupling capacitance and intentional decoupling capacitance.

Intrinsic decoupling capacitance is the inherent capacitance of the on-chip interconnects, and on-chip devices together. The input capacitance of the part of logic which is idle during a clock edge is known as symbiotic bypass capacitance [51].

The term on-chip decoupling capacitance mainly refers to intentional decoupling capacitance. Intentional decoupling capacitance is placed on chip during the design process. It is a major part of the on-chip decoupling capacitance in order to increase the overall decoupling capacitance to a certain level. Intentional decoupling capacitance may utilize more than 20% of the die area in high speed integrated circuits [49]. The area occupied by the on-chip decoupling capacitor is directly proportional to the magnitude of the on-chip decoupling capacitance [53]. The on-chip area occupied by a single large decoupling capacitance is many times that of the area occupied by a typical circuit. Intentional decoupling capacitance is built as a series of small parallel decoupling capacitors [49].

The type of decoupling capacitance depends on the design requirements of the on-chip power distribution network. Normally, three types of decoupling capacitors are used on-chip i.e. a polysilicon-insulator-polysilicon (PIP) capacitor, where an oxide insulator is used in between the polysilicon insulator, a metal-oxide-semiconductor (MOS), where the gate capacitance of MOS is used for decoupling purposes, and a metal-insulator-metal (MIM), where two metal electrodes are separated by a deposited dielectric layer [54].
Chapter 4
Core Switching Noise for 3D PDN

Modeling of the core switching noise and associated power integrity issues for a 3D stack of logic dies interconnected through TSVs is the prime focus of this dissertation. Fig. 4.1 shows a 3D stack of three dies interconnected through TSVs [29]. 3D integration is an emerging technology that is expected to lead to more than Moore’s law. The worldwide academic and research activities currently focus on innovation of technology, simulation, design, and product prototypes [5]. Three-dimensional integration using TSVs is one of the future IC packaging technologies that can eliminate the copper wire between separate dies by stacking the dies on top of each other [15]. TSVs with heights comparable to the substrate thickness can pass through the substrate and can be placed anywhere in the die, providing extra I/O flexibility compared to the copper wires, which can only be placed along the periphery of the chip.

Fig. 4.1 Three dies forming a three-dimensional (3D) system with face-to-back bonding through micro-connects. Power is supplied through vertical TSV pairs from the package substrate [29].

A 3D power distribution network interconnecting global supply grids of two neighboring dies is shown in Fig. 4.2. The grid contains orthogonal layers with paired power and ground lines. The grids are interconnected through corresponding TSVs at each power or ground node. The logic load is assumed to be connected between the power node and a
nearby ground node on each grid. A power distribution TSV pair along with the logic load and decoupling capacitance is shown in Fig. 4.3.

**Fig. 4.2** 3X3 orthogonal global supply grids of two neighboring dies in a 3D stack interconnected through TSVs where red color indicates the power grid and black color indicates the corresponding ground grid.

**Fig. 4.3** Physical model of a power distribution TSV pair with the logic load and decoupling capacitance.
The 3D integration increases the integration density by increasing the number of on-chip devices per unit footprint. For a 3D chip with a footprint size of $1cm^2$, we may have thousands of P/G I/Os for each die and millions of wire segments on the P/G grids in each die [55]. Reliable power delivery is a critical issue in conventional 2D chips due to the large interconnect parasitic involved. This issue is potentially exacerbated in 3D integrated circuits due to smaller footprint size. In addition to this, multiple dies operating at the same time give rise to simultaneous switching noise. Simultaneous switching noise from one die can couple to the neighboring dies and to the substrate through the power and ground TSV pairs. The power transport through the intermediate stack levels adds additional IR drop and $L_{di/dt}$ noise. The enhanced packing densities facilitated by three-dimensional (3D) integration technology also increase the amount of the current per unit footprint of the chip compared to its 2D counterpart [56].

### 4.1 Core Switching Noise

Logic cells are connected between supply and ground TSVs for a 3D power distribution network. Each logic cell has an equivalent capacitance as a load to the power distribution TSV pair. On-chip logic cells switch either low to high or high to low at a clock edge in a synchronous digital system. When on-chip logic cells switch, either they draw current from the supply network or inject current into the ground network. If a lot of logic cells switch simultaneously, they may produce voltage variations within the supply network due to parasitic associated with the power distribution network. This voltage variation is nothing but core switching noise. It is called voltage surge if variation is above the nominal voltage and is called voltage sag if variation is below the nominal supply voltage [57]. The core switching noise depends on the on-chip power distribution network parasitic rather than the package parasitic because of the scaling of interconnect in modern high speed ULSI design. The core switching noise has become more on-chip centric as the package inductance is significantly less than the on-chip inductance at high frequency with the introduction of BGAs and TSVs in modern packaging. The core switching noise is a major part of the total simultaneous switching noise as the current drawn by the core logic load is generally much higher than the I/O driver’s current [58]. The core switching noise in a 3D stack of logic dies interconnected through TSVs is more significant than the 2D ICs due to extra parasitic introduced by vertical power distribution TSV chains. A large switching noise is introduced in a 3D power distribution network if various stacked dies switch simultaneously [19]. The number of power distribution TSVs for a 3D stack of dies is basically limited by the footprint of the die.

While the third dimension is attractive for many applications, it imposes some stringent requirements and bottlenecks on 3D power delivery. Large current requirements per package pin for 3D integration lead to significant complications in reliable power delivery.
A k-tier 3D chip could use k times as much current as a single 2D chip of the same footprint size under similar packaging technology [9]. The power distribution network impedance has not kept up with the scaling of the technology node due to limited wire resources, increased device density, and current demands whereas this situation is further worsened for 3D integration [5]. The increased IR drop and Ldi/dt noise in 3D chips may cause larger variation in operating speed leading to more timing violations [9]. The supply noise overshoot due to inductive parasitic may aggravate reliability issues such as oxide breakdown, hot carrier injection (HCI), and negative bias temperature instability (NBTI) [56]. The power delivery to a 3D stack of high power chips also presents many challenges and requires careful and appropriate resource allocation at the package level, die level, and interstratal interconnect level [19]. Any drop in the core supply voltage directly impacts the maximum operating frequency of the processor [19]. The voltage variations due to core switching noise are spread out to diverse nodes of the power distribution network thereby causing severe performance degradations in the form of propagation delays [59]. The value of the supply voltage may vary from period to period causing severe reliability issues in logic.

The thickness of the gate oxide in modern CMOS VLSI circuits is very thin due to scaling of the power supply voltage [60]. An excessive surge in power voltage or drop in ground voltage may cause a transistor gate oxide reliability issue due to the electrical over-stress.

The channels of CMOS devices are already very short in length due to scaling with the technology nodes. An excessive surge in supply voltage or drop in ground voltage may cause the carriers to inject into the substrate or the gate oxide due to over voltage thereby depleting the drain-channel junction. It is called hot carrier injection and occurs when the transistor is in saturation.

TSVs cause a significant coupling noise and timing problems even if the TSV count is significantly less than the gate count [61]. The core switching noise through power distribution TSVs may directly or through the substrate couple to I/O drivers, power supply network, signal links, clock lines, and analog components of the chip. In addition to this, the core switching noise may couple to neighboring dies through TSVs for a 3D stack of dies interconnected through TSVs as substrates of the different planes may essentially be biased through a common ground [62]. I/O voltage is more sensitive to transient currents produced by switching of the core logic [19]. The power distribution TSVs having significant capacitance due to larger size than the signal TSVs, cause significant noise coupling to the substrate. The coupling noise from a power distribution TSV may cause path delay in the signal line due to the Miller effect. The coupling noise through power distribution TSVs may cause charge sharing to dynamic logic, thereby flipping the signal unintentionally or may change the state of a sequential element in the static logic.
4.2 Modeling of the Core Switching Noise

On-chip core switching noise is driven by the logic switching current, as shown in Fig. 4.4 [63]. Clock edge current is maximum instantaneous current drawn by core logic switching at the rising or falling edge of the clock. The steepness of the clock edge normally increases with the clock frequency and depends on the implementation of the on-chip circuitry. The charge per cycle remains constant and is independent of the frequency. Charge per cycle is an important parameter for the design of on-chip decoupling capacitance [63]. Dynamic current is the time average current of the clock edge and is directly proportional to the clock frequency [63]. Assume that $i(t)$ is clock edge current, and $T$ is the clock period, then charge per clock cycle $Q_{cycle}$, and dynamic current $I_{dynamic}$ is respectively given as follows [63]:

\[ Q_{cycle} = \int_{0}^{T} i(t) dt \]  

\[ I_{dynamic} = \frac{1}{T} \int_{0}^{T} i(t) dt \]

![Schematic presentation of the on-chip current definition for CMOS [63].](image)

There may be different switching current patterns in order to characterize the on-chip core switching noise. The most important are impulse current and AC steady state current. An impulse current is drawn only once due to simultaneous switching of the core logic on the rising edge or falling edge of the clock cycle. It is equivalent to the clock edge current where the sudden voltage sag is created as a result of a large switching current. Most of the
current is drawn through the decoupling capacitance as compared to the power supply due to low inductance and its proximity to the logic load. The voltage noise propagates in the form of a damped sinusoid in the power distribution network. The major power distribution network properties can be extracted through the impulse response [63]. The other one is AC steady state current for which the circuit switches repetitively over many clock cycles and finally the voltage reaches a steady state value [63]. The maximum switching noise is at the frequency around the resonance frequency of the power distribution network. The transient switching noise is clock frequency dependent and usually higher than the AC steady state switching noise beyond the resonance frequency of the power distribution network. Assume the logic current to be a linear ramp having zero value when the time is zero and maximum value when the time reaches the rise time, as shown in Fig. 4.5.

![Figure 4.5](image)

**Fig. 4.5** The shape of the on-chip logic switching current where $t_s$ is switching time.

Simulation of large power distribution networks with a large number of nodes is a major problem for modern VLSI design [64]. A poorly designed power distribution network causes voltage fluctuations and triggers performance degradation. There has been limited focus on modeling the decoupling capacitance and switching current, which play a key role in transient and resonance behavior of on-chip power distribution grid. Hence, extensive transient simulations are required during the design process to ensure the design quality of the power delivery network [65]. In physical design, power/ground (P/G) networks are the special kind of analog circuits that are very large in terms of RLC components due to the fact that all transistors must get their supply from P/G networks [66]. Checking the integrity of the supply voltage using traditional circuit simulation is not practical for reasons of time and memory complexity [66]. The traditional simulation tools are slow and run short of memory for the simulation of a 3D power distribution network. The detailed model of a power distribution network with multiple grids, and a number of vias and TSVs is very complex and computationally inefficient and takes enormous computer resources to optimize the number and values of the decoupling capacitors at each stage. A comprehensive mathematical model for the estimation of the decoupling capacitance with
the associated ESR and ESL at each stage of a 3D stack of dies is required for early design tradeoffs. This issue is also comprehensively addressed in the thesis.

Core switching noise depends on the amount of logic load driven on the rising/falling edge of the clock, the sharpness of the clock edge (i.e. rise time), and the nature of the network between power supply and logic load. The core switching noise can be reduced in different ways like placing on-chip decoupling capacitance close to the load, placing integrated decoupling capacitance with lower values of ESL and ESR into the substrate, keeping the output impedance across load as close to the target impedance as possible, and determining the optimum value of the damping factor for the power distribution network between supply and load. The rise time increases with the speed of the circuit and logic load increases with the integration density of transistors with each technology node and the problem is exacerbated for 3D power distribution networks. The decoupling capacitors integrated into the silicon substrate can provide high capacitance at low cost [67]. The integrated decoupling capacitors have comparatively low effective series resistance and inductance at high frequencies. The effective series resistance and inductance can be further scaled down by inserting banks of small parallel decoupling capacitors in the substrate. The integrated decoupling capacitors provide noise immunity and improved power distribution for 3D chip stacks [68]. Integrated decoupling capacitors are attractive for high density 3D stacked power distribution networks.

Even a small TSV inductance for a three-dimensional (3D) stack of chips may cause high upper peaks of PDN impedance compared to the single chip [69]. The 3D integration provides a limited number of current paths for higher level dies as well as adding more impedance to the on-chip power grids [70]. The power distribution network impedance in 3D ICs is more complicated and hierarchically distributed. In addition to this, the impedance of a 3D power distribution network increases with the frequency due to TSV parasitic from lower to upper dies. The impedance of the 3D PDN should be as close to the target impedance as possible in order to control the switching noise. The overall impedance profile of a 3D power distribution network is changed by placing a decoupling capacitance. This issue is also addressed for a vertical chain of power distribution TSV pairs.

Adding on-chip decoupling capacitance may cause resonance oscillations in the power distribution network [71]. A 3D power distribution network has lower resonance frequency than its 2D counterpart [72]. On-chip decoupling capacitance should, therefore, be selected with a significant ESR in order to damp the resonance oscillations. The logic load on each die may have a resonance frequency as a result of the interaction between the inductance of the power distribution TSV pairs and decoupling capacitance across the logic load. Damping is only required in the frequency domain around the resonance frequency, rather than at all frequencies. Decoupling capacitance should be selected to have maximum ESR around the resonance frequency. The performance and reliability of a 3D power
distribution network also depends on the magnitude and duration of these oscillations. Peak-to-peak ground noise on a power distribution TSV pair is given as follows, assuming that a TSV pair forms an under-damped system with a damping factor less than unity [73].

\[
\Delta v_{pp} = \Delta v \left( 1 + e^{\pi \zeta / \sqrt{1 - \zeta^2}} \right)
\]

where

\[\Delta v_{pp} = \text{Peak-to-peak ground noise on TSV pair.}\]
\[\Delta v = \text{Peak ground noise on TSV pair.}\]
\[\zeta = \frac{R_{\text{total}}}{2 \sqrt{L_{\text{total}}}} \]
\[\zeta = \text{Damping factor.}\]

The worst condition occurs when \( \zeta = 0 \) i.e. \( \Delta v_p = 2 \Delta v_p \). The noise accumulation occurs with the arrival of the next clock cycle.

The resonance frequency is given as follows:

\[
\omega_{res} = \frac{1}{\sqrt{L_{\text{total}} C_{\text{total}}}}
\]

Adding decoupling capacitance increases the amount of total capacitance, thereby lowering the resonance frequency as shown by (4.5). If the clock frequency is fixed, there will be a few oscillations before the arrival of the next clock edge in the power distribution network. This means that a system with on-chip decoupling capacitance is more prone to noise accumulation and needs more damping to control these oscillations. A careful selection of the total inductance of the power distribution TSV pair along with the decoupling capacitance plays an important role in avoiding the noise accumulation. This issue is also addressed in the thesis.

### 4.3 Toward a Robust 3D PDN Design

The structure, size, and layout of the power distribution network should be decided before starting the design of any other part of the chip. The crux of the problem for the power distribution network comes through many unknowns till the very end of the design cycle.
Most of the commercial tools focus on post-layout verification and analysis of the on-chip power grid when the global power distribution network is designed and blocks have been put in place. The detailed information about power distribution network parasitic and currents drawn by transistors are normally known at this stage of the design. It is neither easy nor cost effective to fix a problem in the power distribution network at this stage. The complexity of the problem is further increased by 3D power distribution networks where proper simulation and verification tools are not available. The mathematical models proposed in this thesis address the same issue with significant ease. The proposed models are suitable for early design tradeoffs for 3D power distribution networks. These models provide quick early validity of 3D stacked power distribution design using limited resources. The models are flexible and can be used, for example, for different combinations and design parameters of TSVs. The robustness of design increases by using these models as we take into account the inductance and capacitance of the structures along with their resistance. Table 4.1 shows the physical dimensions of TSVs used for the analysis of models proposed in this thesis. These dimensions are quite analogous to ITRS predictions shown in Table 1.9.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSV diameter</td>
<td>5-10 um</td>
</tr>
<tr>
<td>TSV height</td>
<td>50-100 um</td>
</tr>
<tr>
<td>TSV Aspect Ratio</td>
<td>5:1-10:1</td>
</tr>
<tr>
<td>SiO₂ barrier thickness around TSV</td>
<td>0.2-1 um</td>
</tr>
<tr>
<td>TSV Spacing</td>
<td>30-100 um</td>
</tr>
</tbody>
</table>

A three step approach is followed throughout the modeling and analysis. A physical model of a 3D stacked power distribution network is proposed. Vertical power distribution TSV pairs are placed having certain spacing relative to each other and horizontal power distribution rails are also put in place. Resistance, inductance, and capacitance for TSVs and horizontal power distribution segments are extracted using Ansoft Q3D extractor. Copper TSVs pass through the silicon substrate. TSVs have a cylindrical shape with thin SiO₂ cladding around each TSV. The power and ground rails are placed in the xy-plane with respect to TSVs. Copper rails for P/G are embedded into the top surface of the substrate and connected to respective power or ground TSVs. The physical model along with RLC values is ready at this stage. Each physical model is converted to an equivalent electrical model in the second step. TSVs and horizontal P/G rails are represented by RLC models. Distributed RLC modeling is used for Ansoft simulations. Distributed RLC model with
Ansoft analysis is ready for comparison with the proposed model at this stage. The electrical model is converted to an equivalent mathematical model in the third stage. The mathematical models and algorithms are proposed using different mathematical and simplification techniques. Finally, the models are analyzed and compared to Ansoft Nexxim models for accuracy, speed, and memory requirements.
Chapter 5
Modeling of On-Chip Power Distribution Networks

A power distribution network has a complex hierarchy, from a voltage regulation module (VRM) down to device level in the integrated circuit. The block diagram of a typical power distribution network is shown in Fig. 5.1 [22].

Fig. 5.1 Block diagram of a typical power distribution network for an integrated circuit [22].

The function of voltage regulation module is to sustain a typical voltage level. The board has metallic tracks to supply power to the package. The supply enters the package through package power and ground pins. The supply from pins goes to the pads inside the package and finally to the die or chip through thin metallic bonding wires. The power inside the die is distributed through a hierarchy of orthogonal power distribution grids. The devices are located on the bottom layer. The supply enters through a global supply grid and reaches the devices through a hierarchy of layers from top to bottom.

The term power integrity means making sure that all the devices or transistors on chip have the proper voltage to operate at their intended performance level, and within performance margins. The power delivery network is a complex chain of interconnections from voltage regulator to integrated circuit to guarantee the power integrity of each and every device on the chip. Each and every interface from voltage regulator module to the device contributes to the power distribution noise accordingly.

The difficulty in designing an on-chip power distribution network stems from different factors. For example, the size of the network is very large, having RLC parasitic elements together with switching devices. In addition to this, the switching patterns of the on-chip devices are not well known at the time of starting the design. The inductance of the on-chip power distribution grid plays a critical role for high frequency design. It is difficult to control the behavior of the on-chip inductance of the power distribution grid for a frequency of the order of GHz. Different modeling approaches are used for the design, simulation, and analysis of on-chip power distribution networks. The modeling approaches depend on the type and hierarchy of the on-chip power distribution network.
5.1 Modeling Approaches

5.1.1 Finite Difference Time Domain (FDTD) Method
The Finite Difference Time Domain (FDTD) method solves Maxwell’s equations in the time domain [74][75]. It has been used to solve the circuit equations for multi-conductor transmission lines. The method is computationally efficient and accurate but does not take the frequency dependent skin effect into account. In the case of an on-chip power distribution grid, the skin effect can be neglected if the power distribution lines are thin. The SPICE-based simulator can create problems for on-chip power distribution grids with millions of passive elements. The reason is that SPICE-based tools are normally based on the Modified Nodal Analysis (MNA) method and the matrix inversion takes enormous memory space as well as computation time. The Finite Difference Time Domain (FDTD) method, on the other hand, solves the circuit equations based on the previous solutions of the nodes and branches of the circuit. The basic element of the Finite Difference Time Domain (FDTD) method is the \( \pi \) equivalent model of a small portion of the transmission line [76]. In order to achieve numerical and computational stability, the time step has to be set as \( \Delta t < \frac{1}{\sqrt{LC}} \) (where L is inductance of the branch and C is capacitance of a node) [77]. A finite time step introduces a latency to get an updated value of the node voltages and branch currents of the circuit.

5.1.2 Latency Insertion Method (LIM)
The latency Insertion Method (LIM) was introduced for the efficient transient simulation of large networks [77]. The method is similar to the Finite Difference Time Domain (FDTD) method. The time domain equations are solved for electric and magnetic field in the Finite Difference Time Domain (FDTD) method [74]. The circuit equations are solved for voltage and current using the finite difference technique in the Latency Insertion method (LIM) [78]. The latency Insertion method takes a network as a grid to be composed of nodes interconnected through branches. The branch is composed of a series combination of resistor, inductor, and voltage source. The node is composed of a parallel combination of conductance, current source, and capacitance to ground. The currents and voltages are computed in alternate time intervals.

5.1.3 Hierarchical Modeling Method
Efficient and accurate simulation techniques are required in order to capture the dynamic voltage fluctuations due to the switching of millions of transistors in modern on-chip power distribution grids. Interconnects of the circuit board and package incur additional switching noise. It is extremely difficult to handle the size of the power distribution
network if board, chip, and package are taken together. Therefore, a divide-and-conquer technique is applied, using a hierarchical modeling technique. Each of the board, package, and chip power distribution networks is taken as a separate entity. The results are put together after simulating each of them. The block diagram of the hierarchical model is shown by Fig. 5.2 [79]. The hierarchical modeling technique is used for transient simulation of the power distribution network in order to capture the dynamic voltage fluctuations in the grid [80].

![Hierarchical Model](image)

**Fig. 5.2 Hierarchical Model [79].**

### 5.1.4 Distributed RLC Model for Power Grid

The RLC model for the simulation of large on-chip power distribution networks is extensively used in the literature [81]-[84]. The primary sources of current surges on the power distribution network are input/output (I/O) drivers and internal logic circuitry. The power density for GHz operating range may climb up to 20W/cm² (for air cooled package device) [84]. The power supply should meet average and peak power demands for each part of an integrated circuit. The logic gates switching on or closer to a clock edge make a major contribution to the switching noise in PDN. The RLC model for transient simulation is necessary for taking into account the effect of wire capacitance, and symbiotic bypass capacitance [81]. It is easier to locate and size on-chip decoupling capacitance using the RLC distributed model [81]. The conceptual RLC model for a system-on-chip power grid is shown in Fig. 5.3. The RLC model is used for the analysis of models proposed in this thesis.
5.1.5 Compact Physical Modeling for GSI Power Distribution Grids

The compact physical modeling technique for IR-drop analysis of GSI circuits was first introduced in [85][86]. A power distribution network has orthogonal grids with the vias connecting similar polarity crossing points. Each node is connected to four other nodes through resistive elements as shown in Fig. 5.4 [86]. Partial differential equations for power grid voltages are derived and solved in the compact physical modeling method.
5.1.6 Combined Chip-Package-Board Modeling

The analysis of noise contributed by the chip, package, and board respectively is a part of the system level power integrity. The hierarchy of a typical system level power distribution network starts from the voltage regulator module (VRM) and ends up at the chip or die, as shown in Fig. 5.5. On-chip power is delivered through bonding wires and pads or the C4 bump redistribution layer (RDL) from the package to the complex hierarchy of the power distribution grid. An on-chip power distribution grid has inherently non-ideal impedance which causes fluctuation in voltage before it reaches the active devices. A typical hierarchy of a 3D chip-package-board power distribution network is shown in Fig. 5.5. There are four levels of hierarchy from board to a typical logic cell inside the die or chip for a 3D stacked PDN. Each level of hierarchy has specific decoupling capacitance to meet the load requirements of the next level.

![Fig. 5.5 Chip-Package-Board hierarchy of a typical 3D power distribution network.](image)

5.2 Previous Work and Thesis Contribution

3D integration technology is going through a process of evolution. Power integrity is one of the key challenges especially for a multi core processor design using 3D integration technology. The modeling of interconnect in 3D involves extracting the electrical equivalent to interconnects for a 3D stacked power distribution network. The major challenge comes from the capacitive and inductive coupling of different interconnects. The other challenge is to cover a large frequency range in order to address mixed modes like radio-frequency, analog, and digital components. Unlike conventional horizontal wire, TSV is surrounded by a finite resistive silicon substrate. The capacitance of a TSV depends on the coupling of TSVs to the substrate and nearby TSVs. The silicon substrate is a source of loss and noise coupling for vertical TSVs. TSV is represented by an RLC model.
The impedance characteristics of a power distribution network can capture its resistive, inductive, and capacitive characteristics. Target impedance is used as a design parameter to control the switching noise of a power distribution network. The effects of meshed type, on-chip power distribution network design on the impedance of TSV based 3D stacked PDN are investigated in [87]. The analysis is based on the interaction of on-chip capacitance and through-silicon-via (TSV) inductance. The model is based on frequency domain analysis. The model shows that impedance of 3D-stacked PDN depends on the physical design of the on-chip power distribution grid, the location of power distribution TSVs, and frequency. The impact of 3D power grid parameters like resistance, inductance, and capacitance on the impedance of the power distribution network are studied in [88]. The transient simulation time reduces from hours to less than hundreds of seconds by using parallel processing. The simulation time reduces by using multiprocessors in parallel but use of multiprocessors does not eliminate the use of physical resources. Bonding wires are compared to TSV for the power integrity of a 3D-stacked power distribution network in [89]. However, there is a need for comprehensive power integrity analysis for three-dimensional power distribution network. A physical model for IR-drop analysis of a three-dimensional (3D) power distribution network is proposed in [90]. However, IR-drop analysis alone may not be sufficient for modern high speed VLSI/ULSI designs where on-chip inductance and capacitance play a major role. The effects of power gating noise for three-dimensional multi-processor system-on-chip (3D MPSoC) have been explored in [91]. The core switching noise is critical for the power integrity of such a system, which is not discussed in this paper. A differential power delivery technique is proposed for a three-dimensional integrated circuit in [72]. Only Power delivery issues have been discussed in this paper. There should be comprehensive modeling techniques to overcome power integrity issues. Frequency dependent RLCG model for three-dimensional stacked integrated circuits using through-silicon-vias (TSVs) is proposed in [94]. The effect of substrate coupling between TSVs and grid layers has been taken into account. The physical model proposed in [93] provides guidelines for reducing power supply noise on die level along a three-dimensional stack of dies interconnected through TSVs but more detailed modeling is needed to go into the die for accurate placement of the on-die decoupling capacitance. The impedance matrices for different segments of the network are formed and then reconnected to find the overall impedance of the power distribution network. The impedance characteristics of a three-dimensional GPU-Over-DRAM system using the segmentation method are analyzed in [94]. The drawback of this approach is that combined impedance may have errors because each part is treated as a separate entity. A three-dimensional integrated circuit using TSVs has been reported in [95] and the use of a silicon interposer has been reported in [96]. There is a large amount of IR-drop in a thin-wired interposer in a three-dimensional board-interposer-chip system. A comprehensive co-simulation technique for IR-drop analysis in such a system has been proposed in [97]. But only IR-drop analysis
is not sufficient for modern high speed power distribution networks where on-chip inductance is significant compared to the package inductance.

Fig. 5.6 Electrical model of decoupling capacitance placed on DRAM dies in a 3D-DRAM-Over-Logic System. $V_{dd}$=supply voltage, $R_{eff}^{p}$ = effective resistance of package, $C_{dec1}$ = decoupling capacitance on top DRAM die, $V_{dec1}$ = voltage across $C_{dec1}$, $R_{eff}^{TSV1}$ = effective resistance of first TSV pair, $C_{dec2}$ = decoupling capacitance on bottom DRAM die, $V_{dec2}$ = voltage across $C_{dec2}$, $R_{eff}^{TSV2}$ = effective resistance of second TSV pair, $V_L$ = voltage across logic load, and $I_L = \text{equivalent current drawn by logic load}$.

The rate of decrease in DRAM access latencies could not kept up with the rate of decrease in microprocessor cycle time from one generation to the other, known as the memory wall problem [98]. Stacking DRAM dies directly on top of the processor die is the obvious choice to overcome the memory wall problem. IMEC along with its partners demonstrated the potential for 3D integration of a commercial DRAM over the processor chip using TSVs in 2011 [38]. On-chip decoupling capacitors are very popular for the power integrity but occupy useful die area. Placing decoupling capacitance on a 3D stacked DRAM die has been proposed in order to eliminate the area penalty of decoupling capacitor insertion on the processor die [98][99][100][101]. The issue is how much decoupling capacitance should be placed on each DRAM die without overdesigning and at the same time saving useful
areas of 3D-DRAM stack. This issue is comprehensively addressed by the proposed model of decoupling capacitance for 3D-DRAM-Over-Logic system. Fig. 5.6 shows the proposed electrical model for decoupling capacitors placed on DRAM dies in a 3D-DRAM-Over-Logic System.

Power distribution TSV pairs have to pass through the DRAM stack in order to supply the processor stack in a 3D-DRAM-Over-Logic System. The physical size of the power distribution TSV pairs is also important because TSVs block useful area of the DRAM stack. Similarly, the footprint size of the power distribution TSV pairs occupies useful area on the processor die. The resistance of a TSV depends on its physical size. The area occupied by decoupling capacitance on the DRAM die depends on the amount of decoupling capacitance i.e. a small amount of decoupling capacitance occupies a small area on the DRAM die. The mathematical model for decoupling capacitance for the power integrity of a 3D-DRAM-Over-Logic-System addresses the above issue. The optimum budgeting of the decoupling capacitance is based on the resistance of the power distribution TSVs in the proposed model. It is of utmost importance to maintain the required voltage across logic load during switching of the load.

The current is supplied by both decoupling capacitors during switching of the logic load. The load current is assumed to be a linear ramp as follows

\[ I_L(t) = \frac{t}{t_r} I_{max} \]  

where \( t_r \) is rise time and \( I_{max} \) is peak switching current.

The voltage across load during switching is given as follows:

\[ V_L(t) = \frac{\left(C - A R_{TSV}^{eff}\right) I_{max} + B}{A} \]  

where,

\[ A = 2(C_{dec1} + C_{dec2})^2 t_r \]  

\[ B = \left(2C_{dec1}^3 + 4C_{dec1}^2 C_{dec2} + 6C_{dec1} C_{dec2}^2 \right) t_r \]  

\[ C = 2C_{dec1} R_{TSV1}^{eff} \left(C_{dec1} C_{dec2} t_r - C_{dec1}^2 R_{TSV1}^{eff} e^{-\alpha} \right) + 2C_{dec1}^2 C_{dec2}^2 R_{TSV2}^{eff} \left(R_{TSV2}^{eff} t_r - \alpha \right) - 2C_{dec1} (C_{dec1} + C_{dec2}) t_r^2 \]  

\[ \alpha = \frac{C_{dec1} + C_{dec2}}{C_{dec1} C_{dec2} R_{TSV1}^{eff}} t_r \]
The load voltage is a function of the effective value of TSV resistance, the value of decoupling capacitance, the rise time of the clock edge, and the peak current drawn by the switching logic load as follows:

\[ V_L = f(C_{\text{dec}1}, R_{\text{TSV1}}^{\text{eff}}, C_{\text{dec}2}, R_{\text{TSV2}}^{\text{eff}}, t_r, I_{\text{max}}) \]

The proposed model has +/- 1.1% deviation compared to Ansoft Nexxim, as shown by Table 5.1. There are certain ratios of \( R_{\text{TSV1}}^{\text{eff}} / R_{\text{TSV2}}^{\text{eff}} \) for a given peak of load current for which the decoupling capacitance budget is comparatively less. For example, the total amount of decoupling capacitance is 7pF for 3mA peak current when the \( R_{\text{TSV1}}^{\text{eff}} / R_{\text{TSV2}}^{\text{eff}} \) ratio is kept within 4-10. The amount of decoupling capacitance \( C_{\text{dec}1} \) placed over the lower DRAM die depends on the value of \( R_{\text{TSV1}}^{\text{eff}} \) i.e. the value of \( C_{\text{dec}1} \) should be less if \( R_{\text{TSV1}}^{\text{eff}} \) is higher and should be more if \( R_{\text{TSV1}}^{\text{eff}} \) is lower. The underlying reason is that \( C_{\text{dec}1} \) adds almost parallel to \( C_{\text{dec}2} \) when the value of \( R_{\text{TSV1}}^{\text{eff}} \) is minimum. For example, there is no need to place a decoupling capacitance on the upper DRAM die when the value of \( R_{\text{TSV1}}^{\text{eff}} \) is 0.2 \( \Omega \) for 3mA peak current, and 0.1 \( \Omega \) for 5mA peak current.

Table 5.1 Optimization of decoupling capacitance on both the DRAM dies based on the effective resistance of the TSV pairs, assuming 10% ripple to be allowed in the voltage across the load with the minimum value of \( V_{\text{dec}2} \) to be equal to 0.92V.

<table>
<thead>
<tr>
<th>( R_{\text{TSV2}}^{\text{eff}} ) (m( \Omega ))</th>
<th>( R_{\text{TSV1}}^{\text{eff}} ) (( \Omega ))</th>
<th>( I_{\text{max}} ) (mA)</th>
<th>( C_{\text{dec}2} ) (pF)</th>
<th>( C_{\text{dec}1} ) (pF)</th>
<th>( V_L (V) )</th>
<th>Calculated</th>
<th>Nexxim</th>
<th>(% error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.2</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>0.902</td>
<td>0.90118</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.4</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>0.9021</td>
<td>0.90119</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>0.9029</td>
<td>0.9018</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.8</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0.9023</td>
<td>0.9005</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0.9035</td>
<td>0.8954</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.1</td>
<td>5</td>
<td>0</td>
<td>12</td>
<td>0.904</td>
<td>0.9033</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.6</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>0.9061</td>
<td>0.9044</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.8</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0.9028</td>
<td>0.9000</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>0.903</td>
<td>0.902</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.5</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>0.90255</td>
<td>0.892</td>
<td>1.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2 shows that the contribution of the decoupling capacitance \( C_{\text{dec}2} \) depends on the value of \( R_{\text{TSV2}}^{\text{eff}} \) i.e. the value of \( C_{\text{dec}2} \) has to be substantially increased for a higher value of \( R_{\text{TSV2}}^{\text{eff}} \), specifically when \( R_{\text{TSV2}}^{\text{eff}} = R_{\text{TSV1}}^{\text{eff}} \) there is no need to place decoupling capacitance on
In other words, any decoupling capacitance placed over the lower DRAM die will be ineffective when the value of $R_{TSV}^{\text{eff}}$ is comparable to the value of $R_{TSV1}^{\text{eff}}$. The underlying reason is that the load is directly connected to $R_{TSV}^{\text{eff}}$ and $C_{\text{dec}}$ has to be increased in order to overcome the drop across $R_{TSV}^{\text{eff}}$.

The decoupling capacitance budget depends on the relative values of resistance of power distribution TSVs. In addition to this the part of total decoupling capacitance placed over a DRAM die depends on the relative values of resistance of power distribution TSVs.

Table 5.2 Dependence of decoupling capacitance placed over DRAM dies on power distribution TSV pair resistance. Assume 10% ripple is allowed in voltage across load.

<table>
<thead>
<tr>
<th>$R_{TSV}^{\text{eff}}$ (mΩ)</th>
<th>$R_{TSV1}^{\text{eff}} = 50,\text{mΩ}$</th>
<th>$R_{TSV1}^{\text{eff}} = 100,\text{mΩ}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_{\text{dec2}}$ (pF)</td>
<td>$C_{\text{dec1}}$ (pF)</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>40</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>50</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 5.7 shows that there is an optimum point beyond which further increase in the value of both decoupling capacitors $C_{\text{dec2}}$ and $C_{\text{dec1}}$ makes no contribution towards the voltage across load. The power distribution network will be overdesigned if the total amount of decoupling capacitance is increased beyond this point. The proposed model is hence useful to avoid overdesign and save useful area of DRAM dies which can otherwise be utilized to increase the memory of the system.

Fig. 5.8 shows that increasing the value of $C_{\text{dec1}}$ improves the voltage across $C_{\text{dec2}}$ up to a certain extent beyond which $C_{\text{dec1}}$ saturates. Increasing the value of $C_{\text{dec1}}$ beyond this point does not improve the voltage across $C_{\text{dec2}}$. Decoupling capacitance on the bottom die becomes less effective for higher values of $R_{TSV1}^{\text{eff}}$. $C_{\text{dec2}}$ has to supply the major part of the switching current under this condition.

Fig. 5.9 shows that voltage across output load improves by adding more decoupling capacitance on the upper DRAM die if decoupling capacitance has enough time to fully charge and recharge i.e. a large amount of decoupling capacitance on the upper DRAM die will not be an overdesign for higher rise times.
Fig. 5.7 Value of $V_{dec2}$ vs. value of $C_{dec2}$ for different values of $C_{dec1}$ when $R_{TSV1}^{eff} = 50 \Omega$, $I_{\text{max}} = 3 \text{mA}$, and $t_r = 100 \text{pS}$.

Fig. 5.8 Value of $V_{dec2}$ vs. value of $R_{TSV1}^{eff}$ for different values of $C_{dec1}$ when $R_{TSV2}^{eff} = 50 \Omega$, $C_{dec2} = 1 \text{pF}$, $I_{\text{max}} = 5 \text{mA}$, and $t_r = 50 \text{pS}$.
The proposed model is applicable to a stack of processor and DRAM dies by adding a dedicated power distribution TSV pair for each processor die. The physical size of TSVs and the total amount of decoupling capacitance can be effectively optimized by using the proposed model.

The simulation speed and required memory resources are both bottlenecks for the simulation of on-chip power distribution networks. The complexities associated with these issues are further increased with the introduction of 3D power distribution grids. The size of the power distribution network increases each time with the addition of a chip in 3D stack. The literature so far does not address this issue specifically for 3D power distribution networks. A multi-processing technique is proposed in [88] for transient simulation of a 3D power distribution network. The simulation time reduces from hours to hundreds of seconds using multiprocessing. But using multi-processors in parallel requires a lot of memory and physical resources. Fast and memory-efficient algorithms for the simulation of a 3D power distribution network are needed in order to save time and resources. To the best of the author’s knowledge, so far there is no algorithm for the transient simulation of a 3D power distribution bus with TSVs in the literature. This issue is comprehensively addressed by proposing a fast transient simulation algorithm in this thesis. The algorithm presents a solution for a 3D power distribution bus having \( n \) nodes, as shown in Fig. 5.10. The bus is connected to the top and bottom buses through power distribution TSVs, and left
and right buses through horizontal power distribution grid segments. The approach is based on the Latency Insertion Method (LIM) using efficient equivalent circuit techniques [78]. The algorithm is several times faster and more memory-efficient than Ansoft Nexxim, as shown in Table 5.3 and Table 5.4 respectively.

Fig. 5.10 Electrical model of 3D power distribution bus having $n$ nodes.

The proposed algorithm uses a combination of mathematical techniques and visual C++ to achieve a fast solution for the 3D power distribution bus for voltage at each node. The branches at each node of the bus are converted to a combination of resistance in series with the current source, using the trapezoidal rule. The equivalent bus obtained this way is shown in Fig. 5.11. This bus is converted to the $\pi$ network at each node using C++ reducing code, as shown in Fig. 5.12. The reduced network is solved using nodal analysis i.e. the network is converted to a matrix equation containing unknown voltages and currents. The system of matrices is solved for unknowns using the direct method of solution in Matlab. The reduced $\pi$ model is shown in Fig. 5.13. Finally, a two-terminal $\pi$ model is back solved for each node using C++ back solution code. The proposed algorithm has +/- 1 to 2% deviation compared to Ansoft Nexxim4.1.

Fig. 5.11 Equivalent bus consisting of trapezoidal Norton companion models at each node.
Fig. 5.12 Star to π conversion model at a node.

Fig. 5.13 Equivalent π model of a 3D bus after applying reduction algorithm.

Table 5.3 Comparison of the proposed algorithm with Ansoft Nexxim for CPU time (Sec).

<table>
<thead>
<tr>
<th>P/G Grids</th>
<th>Nexxim</th>
<th>Reduced ckt</th>
<th>Back Solver</th>
<th>Speed up over Nexxim</th>
</tr>
</thead>
<tbody>
<tr>
<td>50<em>50</em>10</td>
<td>5.95</td>
<td>0.018</td>
<td>0.02</td>
<td>156.5</td>
</tr>
<tr>
<td>100<em>100</em>10</td>
<td>92.34</td>
<td>0.075</td>
<td>0.06</td>
<td>684</td>
</tr>
<tr>
<td>200<em>200</em>10</td>
<td>346.8</td>
<td>0.83</td>
<td>0.76</td>
<td>218.1</td>
</tr>
<tr>
<td>400<em>400</em>10</td>
<td>600.3</td>
<td>4.22</td>
<td>3.68</td>
<td>75.9</td>
</tr>
<tr>
<td>600<em>600</em>10</td>
<td>1000</td>
<td>10.7</td>
<td>8.1</td>
<td>53.19</td>
</tr>
<tr>
<td>800<em>800</em>10</td>
<td>1500</td>
<td>16.42</td>
<td>14.26</td>
<td>48.89</td>
</tr>
<tr>
<td>1000<em>1000</em>10</td>
<td>2500</td>
<td>25.56</td>
<td>23.04</td>
<td>51.44</td>
</tr>
</tbody>
</table>
The advantage in both speed and memory is gained through the significant reduction in
the number of nodes compared to Ansoft Nexxim, as shown in Table 5.5. The advantage in
speed and memory compared to Ansoft Nexxim is also as a result of converting the
inductance and capacitance of each branch to the corresponding resistive element. The
accuracy is due to efficient C++ code as follows:

Code for reducing the network to a two-terminal \( \pi \) network:

\[
\begin{align*}
\{ & R_\alpha = \infty ; & I_v = I_{\text{equiv}}_{j,p+1} ; \\
& I_\alpha = 0 ; & R_{vz} = R_{vw} + R_{wx} + R_{yz} + R_{yz} ; \\
& i_{\text{equiv}}_{n,p+1} = 0 ; & I_v = \frac{I_v R_v I_z R_z}{R_{vz}} ; \\
& R_s = \frac{\Delta t}{2C_2} ; & j = j+1 ; \\
& R_v = R_w = R_y = R_z = R_{\text{equiv}}_1 ; & R_v = R_{vz} ;
\end{align*}
\]

\[
\begin{align*}
R_x = R_{uv} ; & \\
I_v = I_w = I_y = I_z = I_{\text{equiv}}_{1,p+1} ; & R_{z} = R_{zu} ;
\end{align*}
\]

\[
\begin{align*}
I_s = I_{sv} ; & \\
j = 2 ; & R_{\alpha} = R_{wu} ;
\end{align*}
\]

\[
\begin{align*}
\text{while} (j <= n) & \\
\{ & R_v = R_{f_{\text{equiv}}} ; & I_v = I_{vz} ; \\
& j = 2 ; & I_u = I_{zu} + i_{\text{equiv}}_{j,p+1} ; \\
& \text{while} \ (j <= n) & I_\alpha = I_\alpha + I_{uv} ; \} \}
\end{align*}
\]

Code for the back solution of reduced \( \pi \) network:

\[
\begin{align*}
\{ & I_{1,p+1} = \left( \frac{V_{n,p+1}}{R_\beta} + \frac{V_{1,p+1}}{R_\alpha} \right) + I_{\beta} + I_\alpha + \frac{V_{1,p+1}}{R_\alpha} \} ; \\
& \text{for} (j = 2 ; j <= n ; j++) \\
\{ & V_{j,p+1} = V_{j,1,p+1} + R_{f_{\text{equiv}}}^j \left( I_j 1_{1,p+1} + I_{\text{equiv}}_{j,1,p+1} \right) ; \\
& I_{j,p+1} = I_j 1_{1,p+1} + \frac{V_{j,p+1}}{R_{f_{\text{equiv}}^j}} + i_{\text{equiv}}_{j,p+1} ; \} \}
\end{align*}
\]
Table 5.4 Comparison of the proposed algorithm with Ansoft Nexxim for required memory (Mb).

<table>
<thead>
<tr>
<th>P/G Grids</th>
<th>Nexxim</th>
<th>Proposed algorithm</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>50<em>50</em>10</td>
<td>100</td>
<td>5.5</td>
<td>94.5</td>
</tr>
<tr>
<td>100<em>100</em>10</td>
<td>400</td>
<td>7.75</td>
<td>392.25</td>
</tr>
<tr>
<td>200<em>200</em>10</td>
<td>700</td>
<td>8.25</td>
<td>691.75</td>
</tr>
<tr>
<td>400<em>400</em>10</td>
<td>1000</td>
<td>10</td>
<td>900</td>
</tr>
<tr>
<td>600<em>600</em>10</td>
<td>3000</td>
<td>235</td>
<td>2765</td>
</tr>
<tr>
<td>800<em>800</em>10</td>
<td>5000</td>
<td>450</td>
<td>4550</td>
</tr>
<tr>
<td>1000<em>1000</em>10</td>
<td>10000</td>
<td>675</td>
<td>9325</td>
</tr>
</tbody>
</table>

Table 5.5 Comparison of the proposed algorithm with Nexxim for node reduction.

<table>
<thead>
<tr>
<th>P/G Grids</th>
<th>Number of nodes in Nexxim</th>
<th>Number of nodes in proposed algorithm</th>
<th>Node reduction ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>50<em>50</em>10</td>
<td>25000</td>
<td>501</td>
<td>49.9</td>
</tr>
<tr>
<td>100<em>100</em>10</td>
<td>100000</td>
<td>1001</td>
<td>99.9</td>
</tr>
<tr>
<td>200<em>200</em>10</td>
<td>400000</td>
<td>2001</td>
<td>199.9</td>
</tr>
<tr>
<td>400<em>400</em>10</td>
<td>1600000</td>
<td>4001</td>
<td>399.9</td>
</tr>
<tr>
<td>600<em>600</em>10</td>
<td>3600000</td>
<td>6001</td>
<td>599.9</td>
</tr>
<tr>
<td>800<em>800</em>10</td>
<td>6400000</td>
<td>8001</td>
<td>799.9</td>
</tr>
<tr>
<td>1000<em>1000</em>10</td>
<td>10000000</td>
<td>10001</td>
<td>999.9</td>
</tr>
</tbody>
</table>

Table 5.3 shows that the proposed algorithm is more than 50 times faster than Ansoft Nexxim for a 1000x1000x10 power distribution grid. In addition to this the proposed algorithm requires more than 13 times less memory resources than Ansoft Nexxim for a 1000x1000x10 power distribution grid. Table 5.5 shows that the proposed algorithm reduces the number of nodes by 1000 times compared to Ansoft Nexxim for a 1000x1000x10 power distribution grid.

The core switching noise caused by the on-chip power distribution grid was overlooked in the past because of low inductance of the on-chip power distribution grid compared to the package inductance. On-chip inductance is more prominent than the package inductance in modern packaging because of the introduction of BGAs and TSVs in 3D integration technology. The enormous speed and integration density of modern 3D-stacked integrated circuits compounds the importance of the inherent inductance of the on-chip power distribution network as a factor in core switching noise. This problem was addressed for 2D
power distribution grid in [23]. A comprehensive mathematical model for the estimation of core switching noise at each node of a 3D stacked power distribution network is proposed in order to address the above mentioned issue for a 3D stacked power distribution network. The parasitic inductance and capacitance of the 3D stacked high speed power distribution network is taken into account in this model. It assumes that orthogonal global supply grids of two neighboring chips are interconnected using power distribution TSV pairs, as shown in Fig. 5.14.

![Fig. 5.14 3X3 orthogonal global supply grids of two neighboring dies in a 3D stack interconnected through TSVs where red color indicates the power grid and black color indicates the corresponding ground grid.](image)

The minimum voltage at an arbitrary node $j$ during switching of the logic load is given as follows

$$V_{j,\text{min}} = \frac{1}{\lambda_j} \left\{ \sum_{i=1 \atop i \neq j}^{n} X_{i,j} V_{i,\text{min}} + \sum_{k=n+1}^{2n} X_{k,j} V_{k,\text{min}} + \frac{1}{2} \left( \sum_{i=1 \atop i \neq j}^{n} C_{i,j} + \sum_{k=n+1}^{2n} C_{k,j} \right) V_{d,\text{min}} \right\}$$  \hspace{1cm}  \text{(5.8)}$$

where

$$\lambda_j = \sum_{i=1 \atop i \neq j}^{n} X_{i,j} + \sum_{k=n+1}^{2n} X_{k,j} + \frac{1}{2} \left( \sum_{i=1 \atop i \neq j}^{n} C_{i,j} + \sum_{k=n+1}^{2n} C_{k,j} \right) + C_{L_j}$$  \hspace{1cm}  \text{(5.9)}$$
\[ X_{i,j} = \frac{t_s^2}{(6L_{i,j} + 3R_{i,j}t_s)} , \quad X_{k,j} = \frac{t_s^2}{(6L_{k,j} + 3R_{k,j}t_s)} \]  \hspace{1cm} 5.10

\( V_{dd} \) = Supply voltage.

\( t_s \) = Switching time.

\( L_{i,j} \) = Parasitic inductance of line segment from node i to j.

\( R_{i,j} \) = Parasitic resistance of line segment from node i to j.

\( L_{k,j} \) = Parasitic inductance of TSV between node j and k.

\( R_{k,j} \) = Parasitic resistance of TSV between node j and k.

\( C_L \) = Equivalent capacitance of switching load on node j.

Equation (5.8) is written in \( nxn \) matrix form as follows, where the bottom chip has 1 to \( n \) nodes and the top chip connected to it through TSVs has \( n+1 \) to \( 2n \) nodes:

\[
\begin{bmatrix}
-1 & X_{1,i} & X_{1,i-1} & \cdots & X_{1,m} \\
X_{i,1} & -1 & X_{i,1} & \cdots & X_{i,m} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
X_{n,1} & X_{n,1} & X_{n,1} & \cdots & -1 \\
& \frac{1}{\lambda_1} & \frac{1}{\lambda_2} & \cdots & \frac{1}{\lambda_m} & -1
\end{bmatrix}
\begin{bmatrix}
V_{i,1} \\
V_{i,2} \\
\vdots \\
V_{i,m} \\
V_{m,1} \\
V_{m,2} \\
\vdots \\
V_{m,n}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\frac{1}{2\lambda_1} \left( \sum_{i=1}^{n} C_{i,i} + \sum_{i=1}^{m} C_{i,i} \right) \\
\frac{1}{2\lambda_1} \left( \sum_{i=1}^{n} C_{i,i} + \sum_{i=1}^{m} C_{i,i} \right) \\
\vdots \\
\frac{1}{2\lambda_1} \left( \sum_{i=1}^{n} C_{i,i} + \sum_{i=1}^{m} C_{i,i} \right) \\
\frac{1}{2\lambda_2} \left( \sum_{i=1}^{n} C_{i,i} + \sum_{i=1}^{m} C_{i,i} \right) \\
\frac{1}{2\lambda_2} \left( \sum_{i=1}^{n} C_{i,i} + \sum_{i=1}^{m} C_{i,i} \right) \\
\vdots \\
\frac{1}{2\lambda_m} \left( \sum_{i=1}^{n} C_{i,i} + \sum_{i=1}^{m} C_{i,i} \right)
\end{bmatrix}
\]
The system of equations in (5.11) is solved using Matlab for unknown voltages. The proposed mathematical model is applicable to any number and any pattern of power distribution TSV pairs between two chips. The proposed model is +/- 5% accurate compared to Ansoft Nexxim for the following parameters of the power distribution grid:

\[ V_{dd} = 1V, \text{ Grid Size } = 10 \times 10 \text{ for each chip (100 nodes).} \]

TSV diameter= 5um, TSV height=50um, and SiO\(_2\) barrier thickness=1um.
Horizontal line segment width=35um, and thickness=1.2um.
Grid pitch=100um, length=width=1mm, spacing=50um.

When multiple dies are stacked together the power distribution TSV pairs form vertical chains as shown in Fig. 5.15. Power distribution TSV pairs should be placed almost one above the other for the return current path. 3D power distribution network may have multiple vertical power distribution chains. Each of the vertical chains should be analyzed for core switching noise at each stage in order to ensure the power integrity of the vertical chain. To the author’s knowledge there is no compact mathematical model to address this issue in the literature. A comprehensive mathematical model is proposed here to address this issue using a vertical power distribution chain shown in Fig. 5.16.

**Fig. 5.15** Vertical power distribution TSV chains.
Consider the $m^{th}$ TSV in the vertical chain of Fig. 5.16. Let voltage at input of $m^{th}$ TSV be $V_{TSV_{m-1}}$ and voltage at output of the $m^{th}$ TSV be $V_{TSV_{m}}$. Let current through the $m^{th}$ TSV be $I_{m}^{TSV}$. Let $C_{tsv}$ be lumped capacitance of a TSV, $C_{d}$ be the decoupling capacitance associated with each TSV pair (consists of symbiotic bypass capacitance plus intentional decoupling capacitance), and $C_{L}$ is the equivalent capacitance of logic load connected to the $m^{th}$ node and $V_{dd}$ is input supply voltage to vertical chain of TSVs. The circuit is in static mode before switching of logic load and voltage at each node of the chain is equal to $V_{dd}$ and, therefore, the charge on each node is equal to $(C_{TSV} + C_{d})V_{dd}$. Let $\Delta V_{\max}^{TSV} = V_{dd} - V_{\min}^{TSV}$ be peak noise at the $m^{th}$ TSV caused by switching of logic load, where $V_{\min}^{TSV}$ be the minimum voltage at node $m$. When the load connected to the $m^{th}$ node switches from low to high, the difference of charge $Q$ is transferred to ground from supply $V_{dd}$, which is given by the following equation:

$$Q = (C_{TSV} + C_{d} + C_{L})V_{\min}^{TSV} - (C_{TSV} + C_{d})V_{dd}$$

Let $t_{s}$ be switching time of logic load and $m$ be the number of TSVs in the vertical chain with respect to the bottom TSV. Then the current through the $m^{th}$ TSV is given by the following equation:

$$I_{m}^{TSV} = \frac{1}{L_{TSV}^{eff}} \int_{0}^{t_{s}} \left( \frac{V_{\min}^{TSV_{n-1}} - V_{\min}^{TSV_{n}} - R_{TSV_{n}}^{eff} I_{m_{\max}}^{TSV}}{t_{s}} \right) dt$$

$$I_{m}^{TSV} = \frac{\left\{ V_{\min}^{TSV_{n-1}} - V_{\min}^{TSV_{n}} - R_{TSV_{n}}^{eff} I_{m_{\max}}^{TSV} \right\}}{2L_{TSV}^{eff} t_{s}}$$

where

$$R_{TSV_{n}}^{eff} = (n+1)R_{TSV}, \quad L_{TSV_{n}}^{eff} = (n+1)L_{TSV}, \quad I_{m_{\max}}^{TSV} = (n-m+1) I_{L_{c}}$$

If we inspect the chain closely, we find that the simultaneous switching current flowing through a TSV effectively passes through $(n+1)$ TSVs before reaching the ground. For example, if we consider an $nth$ power TSV, then the current it carries, passes through it as well as $n$ ground TSVs (a total of $n+1$ TSVs). Therefore, the term $(n+1)$ is multiplied by the resistance and inductance of a single TSV in order to find equivalent effective values of voltage drops during simultaneous switching.

There is a total of $(n-m+1)$ equal currents flowing through each TSV. Suppose $n=5$ and $m=3$ (we consider 3rd power TSV with respect to bottom), then $(n-m+1) = 3$ i.e. it carries
switching current of load connected to itself plus switching current of the other two TSVs above it. Therefore, the term \((n-m+1)\) is multiplied by a load current in order to find out the maximum current through a TSV.

\[
I_{TSV}^m = \left\{ V_{TSV_{m-1}}^{\min} - V_{TSV_m}^{\min} - (n-m+1)(n+1)R_{TSV} I_{t_s} \right\} t_s \left/ 2(n+1)L_{TSV} \right.
\]

5.13

There is switching current through the \(m^{th}\) TSV, caused by simultaneous switching of logic load connected to the chain of TSVs. The charge transferred through the \(m^{th}\) TSV to load during this switching is given by the following equation:

\[
Q = \int_0^t \left( I_{m}^{TSV} \right) dt
\]

\[
Q = \int_0^t \left( V_{TSV_{m-1}}^{\min} - V_{TSV_m}^{\min} - (n-m+1)(n+1)R_{TSV} I_{t_s} \right) \frac{t_s}{2(n+1)L_{TSV}} dt \quad \text{for } 0 \leq t \leq t_s
\]

\[
Q = \alpha V_{TSV_{m-1}}^{\min} - \alpha V_{TSV_m}^{\min} - \delta_m
\]

5.14

where

\[
\alpha = \frac{t_s^2}{4(n+1)L_{TSV}}
\]

\[
\delta_m = \frac{(n-m+1)(n+1)R_{TSV} I_{t_s} t_s^2}{4(n+1)L_{TSV}}
\]

Inserting (5.12) in (5.14) and after simplifying, we get:

\[
\frac{\alpha}{\beta_m} V_{TSV_{m-1}}^{\min} - V_{TSV_m}^{\min} = \gamma_m \quad \text{for } m=1,2,...,n
\]

5.15

where

\[
\beta_m = \left( \alpha + C_{TSV} + C_d + C_m \right)
\]

\[
\gamma_m = \frac{(C_{TSV} + C_d)V_{dd} - \delta_m}{\beta_m}
\]
Writing (5.15) in matrix form for a chain of \( n \) TSV pairs stacking \( n \) chips one above the other in a 3D stack gives minimum voltage at each TSV node as follows:

\[
\begin{bmatrix}
\alpha & -1 & 0 & 0 & \cdots & 0 & 0 \\
0 & \frac{\alpha}{\beta_1} & -1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{\alpha}{\beta_2} & -1 & \cdots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \cdots & \frac{\alpha}{\beta_{n-1}} & -1 \\
0 & 0 & 0 & 0 & \cdots & \frac{\alpha}{\beta_n} & -1 \\
\end{bmatrix}
\begin{bmatrix}
V_{T_{\text{dd}}} \\
V_{T_{\text{min}} T_{\text{TSV}_1}} \\
V_{T_{\text{min}} T_{\text{TSV}_2}} \\
\vdots \\
V_{T_{\text{min}} T_{\text{TSV}_{n-1}}} \\
V_{T_{\text{min}} T_{\text{TSV}_n}} \\
\end{bmatrix}
= 
\begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_3 \\
\vdots \\
\gamma_{n-1} \\
\gamma_n \\
\end{bmatrix}
\]

Minimum switching voltage at any of the \( n \) power TSVs of vertical chain is computed by solving the system of equations in (5.16) using Matlab. The peak switching noise on an arbitrary \( m \)th TSV of the chain is given as follows:

\[
\left(V_{T_{\text{TSV}_m}}^p\right)_{\text{noise}} = \left(V_{\text{dd}} - V_{T_{\text{min}} T_{\text{TSV}_m}}^\text{min}\right) \quad m=1, 2, \ldots, n
\]

Assume a copper TSV with a 0.2um thick barrier layer of SiO\(_2\) around it, and silicon substrate with a thickness equal to the height of the TSV. The spacing between power/ground TSV is 30um. RLC values for different physical sizes of TSVs are extracted using an Ansoft Q3D extractor by varying diameter and keeping the height constant or by varying the height and keeping the diameter constant. The diameter of TSV is 10um, the height of TSV is 100um, and the spacing between TSVs is 30um. The simulation is made on a Windows workstation (2.66GHz processor and 3.48GB RAM).

The model has -/+ 3% deviation compared to Ansoft Nexxim for 500pf load and 5pf decoupling capacitance when applied to a chain of 100 power distribution TSV pairs. The proposed model is more than four times faster than Ansoft Nexxim for a chain of hundred TSV pairs as shown in Table 5.6. Required memory for the proposed model is almost half that of Ansoft Nexxim for a chain of hundred power distribution TSV pairs as shown in Table 5.7. The speed of the proposed model gradually improves compared to Ansoft Nexxim by adding power distribution TSV pairs in a vertical chain, as shown in Table 5.6. Hence the proposed model is more suitable for large 3D stacked power distribution networks.
Fig. 5.16 Electrical model of a vertical chain of $n$ power distribution TSV pairs, interconnecting $n$ chips in a 3D stack of ICs.

The power is mainly distributed from bottom to top using TSVs in a 3D stack of dies. The power distribution TSVs are in pairs i.e. each power TSV should have a nearby ground TSV for the shortest current return path as shown in Fig. 4.3. The power distribution TSV pair should be designed to guarantee the power integrity of the on-chip logic load. Therefore, setting the design parameters of a power distribution TSV pair with decoupling capacitance, and load to meet the specific switching noise requirements is critically important for the overall power integrity of a 3D power distribution network. The following model addresses the same issue in the time domain.
Table 5.6 Simulation time comparison of proposed model with Ansoft Nexxim4.1 by increasing the number of TSV pairs in a vertical chain, for $R_{TSV}=0.25$ ohm, $L_{TSV}=50$ pH, $C_{Load}=500pF$, and $C_d=5pF$. Nodes are numbered from bottom to top.

<table>
<thead>
<tr>
<th>Number of TSV pairs in vertical chain</th>
<th>Simulation time by proposed model (mS)</th>
<th>Simulation time by Ansoft nexxim4.1 (mS)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>38</td>
<td>90</td>
<td>137</td>
</tr>
<tr>
<td>20</td>
<td>40</td>
<td>110</td>
<td>175</td>
</tr>
<tr>
<td>30</td>
<td>42</td>
<td>160</td>
<td>281</td>
</tr>
<tr>
<td>40</td>
<td>43</td>
<td>210</td>
<td>388</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>270</td>
<td>440</td>
</tr>
<tr>
<td>60</td>
<td>55</td>
<td>300</td>
<td>445</td>
</tr>
<tr>
<td>70</td>
<td>60</td>
<td>350</td>
<td>483</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>90</td>
<td>100</td>
<td>430</td>
<td>330</td>
</tr>
<tr>
<td>100</td>
<td>105</td>
<td>450</td>
<td>328</td>
</tr>
</tbody>
</table>

Table 5.7 Memory requirement comparison of the proposed model with Ansoft Nexxim4.1 by increasing the number of TSV pairs in the vertical chain, for $R_{TSV}=0.25$ ohm, $L_{TSV}=50$ pH, $C_{Load}=500pF$, and $C_d=5pF$. Nodes are numbered from bottom to top.

<table>
<thead>
<tr>
<th>Number of TSV pairs in vertical chain</th>
<th>Required Memory for proposed model (Mbits)</th>
<th>Required Memory for Ansoft nexxim4.1 (Mbits)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10.5</td>
<td>20.75</td>
<td>97</td>
</tr>
<tr>
<td>30</td>
<td>12.07</td>
<td>21.6</td>
<td>79</td>
</tr>
<tr>
<td>40</td>
<td>15.8</td>
<td>26.3</td>
<td>67</td>
</tr>
<tr>
<td>60</td>
<td>16.3</td>
<td>27.05</td>
<td>66</td>
</tr>
<tr>
<td>80</td>
<td>16.9</td>
<td>28.19</td>
<td>67</td>
</tr>
<tr>
<td>100</td>
<td>17.2</td>
<td>29.00</td>
<td>69</td>
</tr>
</tbody>
</table>

Assume decoupling capacitance is located so that impedance between load and decoupling capacitance is negligible. Also assume that impedance between decoupling capacitance and power TSV is negligible. An electrical model of a power distribution TSV pair with logic load, and decoupling capacitance is shown in Fig. 5.17.
Fig. 5.17 Equivalent electrical model of a power distribution TSV pair with logic load and decoupling capacitance, where parameters are defined as follows:

\[ R_{TSV}^{eff} = \text{effective resistance of TSV pair.} \]

\[ L_{TSV}^{eff} = \text{effective inductance of TSV pair.} \]

\[ I(t) = \text{current drawn through power supply } V_{dd}. \]

\[ C_{dec} = \text{Decoupling capacitance.} \]

\[ R_{dec}^{eff} = \text{Effective series resistance of decoupling capacitance.} \]

\[ L_{dec}^{eff} = \text{Effective series inductance of decoupling capacitance.} \]

\[ I_{dec}(t) = \text{Current through decoupling capacitance.} \]

\[ I_{sw}^{p}(t) = \text{Peak switching current drawn by load.} \]

\[ t_{sw} = \text{Switching current rise time.} \]

\[ \Delta v_{g}^{p} = \text{Peak ground noise.} \]

\[ \text{ESR} = \text{Effective series resistance of the capacitor.} \]

\[ \text{ESL} = \text{Effective series inductance of the capacitor.} \]
Fig. 5.18 shows variation of peak and average noise with load current in the time domain for a clock cycle in a synchronous digital circuit. The power distribution TSV along with decoupling capacitance should be designed to meet the peak noise requirements in order to guarantee the power integrity of the 3D power distribution network. There is a glitch in supply voltage corresponding to ground bounce during switching of the logic load. Assume the ground noise $\Delta V_g(t)$ as a ramp function as follows:

$$\frac{\Delta V_g(t)}{t} = \frac{v_g^p(t)}{t_n} \Rightarrow \Delta V_g(t) = \frac{t}{t_n} v_g^p(t) \tag{5.18}$$

where, $v_g^p(t) =$ peak ground noise and, $t_n =$ rise time of the noise spike.

Transient current supplied by decoupling capacitance to load during switching is given as follows:

$$I_{dec}(t) = I_{dec}(0)e^{-t/R_{dec}C_{dec}} + I_{dec}(\infty)\left\{ 1 - e^{-t/R_{dec}C_{dec}} \right\} \tag{5.19}$$

where $I_{dec}(0) =$ current through decoupling capacitance when time $t$ is zero and $I_{dec}(\infty) =$ current through decoupling capacitance when time $t$ reaches infinity. Decoupling capacitance has zero current at $t=0$; therefore, $I_{dec}(t) = 0$. Decoupling capacitance is charged to $V_{dd}$ before switching of the logic load. Suppose a ground noise spike of $\Delta V_g(t)$ is produced during switching. The voltage across decoupling capacitance after this switching reduces to $V_{dd} - 2\Delta V_g(t)$.

Let voltage across decoupling capacitance be $V_{dec}(t)$. Then discharging current through decoupling capacitance is given as follows

$$I_{dec}(t) = -C_{dec} \frac{\partial V_{dec}(t)}{\partial t} \tag{5.20}$$
\[ I_{\text{dc}}(\infty) = -C_{\text{dec}} \frac{\partial}{\partial t}\{V_{\text{dd}} - 2\Delta v_g(t)\} \]  
5.21

Inserting (5.18) in (5.21) and after simplifying, we get:
\[ I_{\text{dc}}(\infty) = \frac{2 C_{\text{dec}}}{t_n} v_g^p(t) \]  
5.22

Inserting (5.22) in (5.19) and simplifying, we get:
\[ I_{\text{dc}}(t) = \frac{2 C_{\text{dec}} v_g^p(t)}{t_n} \left( 1 - e^{-\frac{t}{R_{\text{dec}} C_{\text{dec}}}} \right) \]  
5.23

The part of the current drawn through the power supply due to switching of logic load is given by the following equation:
\[ I(t) = \frac{1}{L_{\text{TSV}}^{\text{eff}}} \int_{0}^{\infty} \Delta v_g(t) dt - \frac{R_{\text{TSV}}^{\text{eff}}}{L_{\text{TSV}}^{\text{eff}}} \int_{0}^{\infty} I(t) dt \]  
5.24

By inserting (5.18) in (5.24) and simplifying after taking the time derivative of both sides, we get:
\[ I(t) = \frac{v_g^p(t)}{R_{\text{TSV}}^{\text{eff}} t_n} \left( \frac{L_{\text{TSV}}^{\text{eff}}}{R_{\text{TSV}}^{\text{eff}}} \right) \frac{\partial I(t)}{\partial t} \]  
5.25

The transient current drawn through the power supply changes as follows:
\[ \frac{\partial I(t)}{\partial t} = \frac{1}{t_n} \frac{v_g^p(t)}{R_{\text{TSV}}^{\text{eff}}} \left\{ 1 - e^{-\frac{t R_{\text{TSV}}^{\text{eff}}}{L_{\text{TSV}}^{\text{eff}}}} \right\} \]  
5.26

Through all these background equations and simplifications the equation for peak ground noise as a function of rise time is given as follows
\[ v_g^{\text{pp}}(t_{\text{rise}}) = I_{\text{sw}}^p(t_{\text{rise}}) \left( \frac{X_c(t_{\text{rise}}) X_c(t_{\text{rise}})}{X_c(t_{\text{rise}}) + X_L(t_{\text{rise}})} \right) \]  
5.27

where
- \( I_{\text{sw}}^p \) = Peak value of the switching current.
- \( X_c \) = Capacitive reactance.
- \( X_L \) = Inductive reactance.

Hence, peak-to-peak ground noise is given as follows:
\[ v_g^{\text{pp}}(t_{\text{rise}}) = v_g^p(t_{\text{rise}}) \left\{ 1 + e^{-\pi \xi^2 / \sqrt{\xi^2 - 1}} \right\} \]  
5.28
where $v_g(t_{\text{rise}})$ = peak-to-peak ground noise, $v_g(t_{\text{rise}})$ = peak ground noise, and $\zeta$ = damping factor of the circuit given as follows

$$\zeta = \left( \frac{R_{TSV}^{\text{eff}} + R_{\text{dec}}^{\text{eff}}}{2} \right) \sqrt{\frac{C_{\text{dec}}}{L_{TSV}^{\text{eff}}}}$$

5.29

When $R_{\text{sw}}^{\text{eff}} \to 0$ and $R_{TSV}^{\text{eff}} \to 0$, then $\zeta = 0$ as shown by (5.29). The peak to peak ground noise is equal to twice the peak ground noise under this condition as shown by (5.28). The worst case noise accumulation takes place for the next switching instances. The power distribution network may be destabilized and become unpredictable. Hence, the power distribution TSV pair with decoupling capacitance needs more attention to damping during design to avoid the risk of noise accumulation. Equation (5.29) shows that damping factor of a power distribution TSV pair can be controlled by careful selection of resistance and inductance of the power distribution TSV pair, and the value of the associated decoupling capacitance.

**Fig. 5.19** Peak-to-peak ground noise comparison of model from equation (5.28) to the equivalent model in Ansoft Nexxim4.1, as a function of rise time for $I_{\text{sw}} = 3.3mA, L_{TSV}^{\text{eff}} = 50pH, R_{\text{sw}}^{\text{eff}} = 0.25\Omega, R_{TSV}^{\text{eff}} = 75m\Omega, C_{\text{dec}} = 5pF$, and $t = t_{\text{rise}}$. 
Peak ground noise is directly proportional to peak switching current, and the magnitude of peak noise oscillations can be reduced by reducing the amount of peak switching load for a clock edge. Equation (5.28) has -/+ 5% deviation compared to Ansoft Nexxim, as shown in Fig. 5.19. The part of the switching current supplied by decoupling capacitance is higher at lower rise times because of lower capacitive reactance, as shown in Fig. 5.20. Hence, decoupling capacitance is more effective at lower rise times or higher frequencies. The part of the switching current drawn through the power supply increases by increasing the rise time because of lower inductive reactance at lower frequencies, as shown in Fig. 5.21.

Maximum peak noise occurs when capacitive reactance is equal to inductive reactance, as shown by (5.27). There exists a rise time at which capacitive reactance is equal to the inductive reactance and peak noise has maximum value. This is worst case rise time ($t_w = t_{sww}$) and is analogous to resonance frequency in frequency domain analysis.
Worst case rise time is given as follows:

\[
t_w = 2 \sqrt{L_{TSV}^{\text{eff}} C_{\text{dec}}}
\]

**Fig. 5.21** The part of switching current drawn from power supply as a function of rise time for \( I_w = 3.3 \text{mA}, L_{TSV}^{\text{eff}} = 50 \mu\text{H}, R_{TSV}^{\text{eff}} = 0.25 \Omega, R_{\text{dec}}^{\text{eff}} = 75 \text{m}\Omega, C_{\text{dec}} = 5 \mu\text{F}, \) and \( t = t_w. \)

**Table 5.8** peak-to-peak ground noise comparison of (5.28) model with equivalent model obtained through Nexxim4.1 for worst case rise time by varying values of effective inductance of TSV pair for \( I_w = 3.3 \text{mA}, R_{TSV}^{\text{eff}} = 0.25 \Omega, R_{\text{dec}}^{\text{eff}} = 0.075 \text{m}\Omega, \) and \( C_{\text{dec}} = 5 \mu\text{F}. \)

<table>
<thead>
<tr>
<th>( L_{TSV}^{\text{eff}} ) (pH)</th>
<th>Max ground noise for ( t_w = 2 \sqrt{L_{TSV}^{\text{eff}} C_{\text{dec}}} ) (mV)</th>
<th>Max ground noise Nexxim (mV)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.38</td>
<td>2.43</td>
<td>2.1</td>
</tr>
<tr>
<td>20</td>
<td>2.64</td>
<td>2.69</td>
<td>1.8</td>
</tr>
<tr>
<td>50</td>
<td>3.07</td>
<td>3.11</td>
<td>1.3</td>
</tr>
</tbody>
</table>
Adding decoupling capacitance is an effective way of reducing the noise but the side-effect is that resonance frequency comes down or, in other words the worst case switching time goes up. Symbiotic bypass capacitance is the constituent part of the decoupling capacitance. For high density 3D integrated circuits, the symbiotic bypass capacitance is significant which may increase the worst case switching time or reduce the resonance frequency. The assumption that fast switching transients cause the worst case noise is no more valid for such high density circuits. Table 5.8 shows that peak-to-peak ground noise for worst case rise time increases by increasing the value of effective inductance of TSV pair whereas the % error decreases. Table 5.9 shows that peak-to-peak ground noise increases by increasing the value of effective series resistance of decoupling capacitor as well as the value of TSV effective resistance. The % error also increases by increasing the value of TSV effective resistance and effective series resistance of decoupling capacitance but is sufficiently small for small values of TSV effective resistance and effective series resistance of decoupling capacitance. Table 5.10 shows that noise decreases by increasing the value of decoupling capacitance whereas % error increases. This effect on peak-to-peak ground noise is further elaborated by Fig. 5.22 and Fig. 5.23 for different values of TSV effective resistance and effective series resistance of decoupling capacitance. Fig. 5.22 shows that peak-to-peak ground noise reduces by increasing TSV effective resistance up to a specific rise time equal to 100ps. This is because of additional damping. The noise increases beyond this rise time due to increased IR drop. The decoupling capacitance therefore, becomes more effective beyond this point. Fig. 5.23 shows that the noise decreases by increasing the effective series resistance of decoupling capacitance for higher rise times for example above 100-200ps in this case. This is due to increased damping. On the other hand the noise is increased by increasing the effective series resistance of decoupling capacitance for lower rise times, for example below 100ps in this case, where decoupling capacitance is less effective.

Table 5.11 and 5.12 show the effect of decoupling capacitance and TSV effective inductance respectively on peak-to-peak ground noise for different values of rise time. Table 5.11 shows that the maximum noise reduction is 50% at 10pS rise time by doubling the value of decoupling capacitance. It gradually reduces by increasing the rise time and finally it is minimum i.e. 5% at 300ps rise time. Table 5.12 shows that noise reduction is, minimum at 10ps rise time and it gradually increases by increasing the rise time. It is maximum and equal to 40% when rise time is 300ps and the value of TSV effective inductance is halved. This fact is further elaborated in Fig. 5.24 which shows a peak-to-peak ground noise reduction comparison by doubling the value of decoupling capacitance and halving the value of TSV effective inductance. It shows that doubling the value of decoupling capacitance is highly effective for noise reduction at lower rise times for example 100ps in this case. On the other hand, halving the value of TSV effective inductance is highly effective for noise reduction at higher rise times for example above
100ps in this case. The slope of curves in Fig. 5.24 shows that we get less noise reduction by increasing the value of decoupling capacitance at higher rise times and similarly noise reduction is less by reducing the value of TSV effective inductance for lower rise times. Therefore, the designer has to control TSV effective inductance while designing for lower frequencies while taking into account the value of decoupling capacitance when designing for higher frequencies.

Table 5.9 Peak-to-peak ground noise comparison of (5.28) model with equivalent model obtained through Nexxim4.1 for worst case rise time by varying values of effective resistance of TSV pair and effective series resistance of decoupling capacitor for $I_m = 3.3 \text{ mA}$, $L_{TSV} = 50 \text{ pH}$, and $C_{dec} = 10 \text{ pF}$.

<table>
<thead>
<tr>
<th>$R_{TSV}^{\text{eff}}$ (Ω)</th>
<th>$R_{dec}^{\text{eff}}$ (Ω)</th>
<th>$t_u = 2\sqrt{\frac{L_{TSV}^{\text{eff}}}{C_{dec}^{\text{eff}}}}$ (mV)</th>
<th>Max ground noise Nexxim (mV)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.075</td>
<td>2.48</td>
<td>2.54</td>
<td>2.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.075</td>
<td>3.22</td>
<td>3.3</td>
<td>2.5</td>
</tr>
<tr>
<td>0.75</td>
<td>0.075</td>
<td>3.63</td>
<td>3.74</td>
<td>3</td>
</tr>
<tr>
<td>0.75</td>
<td>0.375</td>
<td>3.76</td>
<td>3.89</td>
<td>3.7</td>
</tr>
<tr>
<td>0.75</td>
<td>0.75</td>
<td>3.86</td>
<td>4.02</td>
<td>4.1</td>
</tr>
</tbody>
</table>

Table 5.10 Peak-to-peak ground noise comparison of (5.28) model with equivalent model obtained through Nexxim4.1 for worst case rise time by varying values of decoupling capacitance for $I_m = 3.3 \text{ mA}$, $R_{TSV}^{\text{eff}} = 0.25 \Omega$, $R_{dec}^{\text{eff}} = 0.075 \Omega$, and $L_{TSV}^{\text{eff}} = 50 \text{ pH}$.

<table>
<thead>
<tr>
<th>$C_{dec}$ (pF)</th>
<th>$t_u = 2\sqrt{\frac{L_{TSV}^{\text{eff}}}{C_{dec}^{\text{eff}}}}$ (mV)</th>
<th>Max ground noise Nexxim (mV)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.07</td>
<td>3.11</td>
<td>1.3</td>
</tr>
<tr>
<td>7</td>
<td>2.78</td>
<td>2.83</td>
<td>1.8</td>
</tr>
<tr>
<td>10</td>
<td>2.48</td>
<td>2.54</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Table 5.11 Peak-to-peak ground noise reduction for different values of the rise time by increasing the value of decoupling capacitance for $I_{in} = 3.3mA$, $L_{TSV}^{eff} = 50pH$, $R_{TSV}^{eff} = 0.25\Omega$, $R_{ac}^{eff} = 75m\Omega$, $C_{dec} = 5pF$ and $t = t_s$.

<table>
<thead>
<tr>
<th>Rise time (pSec)</th>
<th>Peak-to-peak ground noise for $C_{dec} = 5pF$ (mV)</th>
<th>Peak-to-peak ground noise for $C_{dec} = 10pF$ (mV)</th>
<th>Reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5.49</td>
<td>2.75</td>
<td>50</td>
</tr>
<tr>
<td>15.8</td>
<td>8.13</td>
<td>4.4</td>
<td>45</td>
</tr>
<tr>
<td>70</td>
<td>7.89</td>
<td>6.41</td>
<td>20</td>
</tr>
<tr>
<td>150</td>
<td>4.84</td>
<td>4.43</td>
<td>8</td>
</tr>
<tr>
<td>300</td>
<td>3.05</td>
<td>2.9</td>
<td>5</td>
</tr>
</tbody>
</table>

Fig. 5.22 Peak-to-peak ground noise for different values of $R_{TSV}^{eff}$ for $I_{in} = 3.3mA$, $L_{TSV}^{eff} = 50pH$, $R_{ac}^{eff} = 75m\Omega$, $C_{dec} = 5pF$, and $t = t_s$. 
Fig. 5.23 Peak-to-peak ground noise for different values of \( R_{\text{dec}} \) for \( I_p = 3.3mA, L_{\text{TSV}}^{\text{eff}} = 50pH, R_{\text{TSV}}^\text{eff} = 0.25\Omega, C_{\text{dec}} = 5pF, \) and \( t = t_s \).

Table 5.12 Peak-to-peak ground noise reduction for different values of the rise time by decreasing the value of TSV effective inductance for \( I_p = 3.3mA, C_{\text{dec}} = 5pF, R_{\text{TSV}}^\text{eff} = 0.25\Omega, R_{\text{dec}}^\text{eff} = 75m\Omega, C_{\text{dec}} = 5pF, \) and \( t = t_s \).

<table>
<thead>
<tr>
<th>Rise time (pSec)</th>
<th>Peak-to-peak ground noise for ( L_{\text{TSV}}^\text{eff} = 50pH ) (mV)</th>
<th>Peak-to-peak ground noise for ( L_{\text{TSV}}^\text{eff} = 25pH ) (mV)</th>
<th>Reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5.49</td>
<td>4.84</td>
<td>12</td>
</tr>
<tr>
<td>15.8</td>
<td>10.1</td>
<td>7.5</td>
<td>25</td>
</tr>
<tr>
<td>70</td>
<td>7.89</td>
<td>4.58</td>
<td>42</td>
</tr>
<tr>
<td>150</td>
<td>4.84</td>
<td>2.9</td>
<td>39</td>
</tr>
<tr>
<td>300</td>
<td>3.0</td>
<td>1.8</td>
<td>40</td>
</tr>
</tbody>
</table>
Fig. 5.24 Peak-to-peak ground noise reduction comparison by doubling the value of decoupling capacitance and by halving the value of TSV effective parasitic inductance.

Power distribution TSV pairs appear in the form of vertical chains in a 3D stacked power distribution network as shown in Fig. 5.15. Consider a vertical chain of power distribution TSV pairs with logic load and decoupling capacitance as shown in Fig. 5.25. The logic load is represented by current source. Each power TSV has a corresponding ground TSV for the current return path. Decoupling capacitance is connected across the load at each stage.

The output impedance across the core logic at the top tier of a 3D power distribution system is a function of the impedance associated with the lower tiers. The impedance peaks (for MHz-GHz frequency range) are formed due to resonance between the decoupling capacitance and the TSV effective inductance at each stage of the 3D power distribution network. The more challenging task in a power distribution system is to minimize the voltage variations caused by the fast switching currents at low operating voltages [105]. These voltage fluctuations can be kept within the allowable limits by keeping the output impedance around the target impedance [105]. The impedance at each stage of a 3D power distribution network should be designed as a resonance free and low impedance path for a
targeted frequency range in order to eliminate the voltage drops caused by the above mentioned resonant peaks.

Fig. 5.25 Vertical chain of \( n \) power distribution TSVs pairs interconnecting \( n \) dies in a 3D stack. The power TSVs are in the red dotted box and the corresponding ground TSVs are in the black dotted box.

Adding decoupling capacitance is an effective way of reducing the impedance of a power delivery system at high frequency [103]. The location of decoupling capacitance significantly affects the design of a power distribution network in high speed integrated circuits [104]. On-chip decoupling capacitance should be close to the switching logic load in order to reduce the excessive supply voltage drop [105]. The decoupling capacitance should not be located far from the supply voltage so that it is unable to fully recharge between two
successive switching events which makes the allocation of decoupling capacitance a complex issue in 3D ICs [106].

\[ \text{Efficient and accurate selection of the decoupling capacitance at each stage of a 3D stack of dies is of critical importance for a reliable design in order to eliminate the noise peaks and to realize the output impedance closer to the target impedance for targeted frequency range. The decoupling capacitors behave as localized low impedance charge reservoirs to realize reduced output impedance. The detailed model of a 3D power distribution network with multiple grids, a number of vias and TSVs is very complex and computationally inefficient and takes enormous computer resources to optimize the number and values of the decoupling capacitors at each stage.} \]

\[ \text{A comprehensive mathematical model to estimate the amount of decoupling capacitance with the associated ESR, and ESL at each stage of a 3D stack of dies is required for early design tradeoffs. The following model addresses this issue for a vertical chain of power} \]
distribution TSV pairs. The proposed model gives guidelines for the design of decoupling capacitance for each stage of the vertical chain of power distribution TSV pairs. The guidelines are based on resonance free output impedance across load for each stage.

Fig. 5.26 shows an equivalent electrical model of the vertical chain of the power distribution TSV pairs with logic load, decoupling capacitance, and the output impedance at each stage. Consider the bottom power distribution TSV pair in isolation from the whole chain as shown in Fig. 5.27. Assume the decoupling capacitance to be very close to both the logic load and the power distribution TSV pair. Assume zero impedance between the logic load and the decoupling capacitance as well as the TSV pair and the decoupling capacitance. Effective values of the TSV resistance and the TSV inductance are given as $R_{TSV}^{eff} = 2R_{TSV}$ and $L_{TSV}^{eff} = 2(L_{TSV} - M_{TSV})$, where the self resistance of the TSV is $R_{TSV}$, the self inductance of the TSV is $L_{TSV}$, and the mutual inductance of the power and the ground TSV is $M_{TSV}$.

Assume a high speed design of the order of GHz. The effective impedance of the TSV pair is in parallel with the impedance of the decoupling capacitance in Fig. 5.27. The effective impedance $Z_{P}^{eff}$ across the current source looking towards the power distribution TSV pair is given as follows:

$$Z_{P}^{eff} = Z_{TSV}^{eff} \| Z_{dec} = \frac{(Z_{TSV}^{eff} Z_{dec})}{(Z_{TSV}^{eff} + Z_{dec})}$$

5.32

where, $Z_{TSV}^{eff} = (R_{TSV}^{eff} + j\omega L_{TSV}^{eff})$, and $Z_{dec} = \left\{ R_{dec} + j\omega L_{dec} - \frac{j}{\omega C_{dec}} \right\}$

![Fig. 5.27](image)

**Fig. 5.27** The equivalent electrical model of a single power distribution TSV pair along with logic load and decoupling capacitance, where $I_{L_i}(t)$ is the current drawn by the logic load, and $Z_{P}^{eff}$ is the effective output impedance across the load.
Inserting these values in (5.32) and normalizing the output impedance with \( R_{TSV}^{eff} \) and after simplifying, we get

\[
Z_{r,s}^{eff} = \frac{ \left[ U_R + Q_{TSV} \left\{ 1 - U_{L} \left( \frac{f}{f_r} \right)^2 \right\} \right] + jQ_{TSV} \left\{ \left( U_R + U_{L} \right) \frac{f}{f_r} - \frac{f}{f} \right\} }{(1 + U_R) + jQ_{TSV} \left\{ (1 + U_{L}) \frac{f}{f_r} - \frac{f}{f} \right\}}
\]

5.33

Where

\[
U_R = \frac{R_{dec}}{R_{TSV}^{eff}}
\]

5.34

\[
U_{L} = \frac{L_{dec}}{L_{TSV}^{eff}}
\]

5.35

\[
f_r = \frac{1}{2\pi \sqrt{L_{TSV}^{eff} C_{dec}}} = \text{Resonance frequency.}
\]

\[
f = \text{Clock frequency.}
\]

\[
\frac{f}{f_r} = \text{Normalized frequency.}
\]

The quality factor of the power distribution TSV pair is given as follows

\[
Q_{TSV} = 2\pi f_r \left( \frac{L_{TSV}^{eff}}{R_{TSV}^{eff}} \right)
\]

5.36

The magnitude of the normalized output impedance across the power distribution TSV pair using (5.32) is given as follows

\[
|Z_{r,s}^{eff}| = \sqrt{ \left[ U_R + Q_{TSV} \left\{ 1 - U_{L} \left( \frac{f}{f_r} \right)^2 \right\} \right]^2 + \left[ Q_{TSV} \left\{ \left( U_R + U_{L} \right) \frac{f}{f_r} - \frac{f}{f} \right\} \right]^2 }
\]

\[
(1 + U_R)^2 + \left[ Q_{TSV} \left\{ (1 + U_{L}) \frac{f}{f_r} - \frac{f}{f} \right\} \right]^2
\]

5.37

Fig. 5.28 shows the magnitude of the normalized output impedance vs. the normalized frequency for different values of \( U_{L} \) when \( Q_{TSV} = 1 \), and \( U_R = 1 \). We can see that the
normalized output impedance is equal to one for all the values of the normalized frequency for $U_{L_i} = 0$, i.e. $R_{\text{dev}} \approx 0$ (when ESL of the decoupling capacitance is equal to zero), which is the required flat output impedance condition for the design of the decoupling capacitance. The magnitude of the normalized output impedance decreases up to a certain value of the normalized frequency and then increases almost linearly with the frequency when $U_{L_i}$ is above 0.05. However the value of the normalized frequency for which the magnitude of the normalized output impedance becomes equal to one, also decreases by increasing the value of $U_{L_i}$ of the decoupling capacitance in comparison to the effective inductance of the TSV pair. This means that the resonance condition occurs at a lower frequency by increasing the value of $U_{L_i}$ and occurs at a higher frequency by decreasing $U_{L_i}$ of the decoupling capacitance.

The quality factor of the power distribution TSV pair as a function of $U_{R_i}$, and $U_{L_i}$ is given as follows

$$Q_{\text{TSV}} = \frac{1+U_{R_i}}{2\sqrt{1+U_{L_i}}}$$

5.38

The damping factor of power distribution TSV pair shown in Fig. 5.28 is given as follows

$$\xi = \left( \frac{R_{\text{dev}}^{\text{eff}} + R_{\text{dev}}}{2} \right) \sqrt{\frac{(L_{\text{TSV}}^{\text{eff}} + L_{\text{dev}})}{C_{\text{dev}}}}$$

5.39

$$\xi = \frac{\left( 1 + \frac{R_{\text{dev}}^{\text{eff}}}{R_{\text{dev}}^{\text{eff}}} \right)}{2 \left( \frac{R_{\text{TSV}}^{\text{eff}}}{L_{\text{TSV}}^{\text{eff}}} \right) \sqrt{\frac{L_{\text{dev}}}{C_{\text{dev}}}}} = \frac{(1+U_{R_i})}{2Q_{\text{TSV}} \sqrt{(1+U_{L_i})}}$$

5.40

The damping factor for the critically damped power distribution TSV pair shown in Fig. 5.28 is given as follows using (5.40):

$$\xi = 1, \text{ when } U_{R_i} = Q_{\text{TSV}} = 1 \text{ and } U_{L_i} = 0 \text{ in (5.40).}$$

The magnitude of the normalized output impedance for the critically damped power distribution TSV pair as a function of $U_{R_i}$, $U_{L_i}$ and $f/f_c$ is given as follows
Fig. 5.28 Plot obtained using (5.37), showing the magnitude of the normalized output impedance vs. normalized frequency for different values of $U_{L_1}$ for $U_r = Q_{TSV_1} = 1$.

Fig. 5.29 shows a plot of the normalized output impedance vs. the normalized frequency using (5.41). The figure shows that the peak of the normalized output impedance is shifted
to lower frequency by decreasing the quality factor or in other words by increasing the value of the decoupling capacitance for the critically damped system.

![Diagram](image)

**Fig. 5.29** The plot obtained through (5.41), showing the magnitude of the normalized output impedance vs. the normalized frequency for different values of $Q_{TSV_i}$, for critically damped system.

The normalized decoupling capacitance for the critically damped power distribution TSV pair is given as follows:

$$
\frac{C_{dec_i}}{I_{TSV_i}} = \frac{1 + U_{L_i}}{(1 + U_{R_i})^2} \left[ \frac{1}{\xi} \right]
$$

5.42
The normalized value of decoupling capacitance in (5.42) is a function of $U_{R_i}$ and $U_{L_i}$. The system is critically damped when $U_{L_i} = 0$ and $U_{R_i} = 1$. The value of decoupling capacitance $C_{dec_i}$ is equal to $\frac{L_{TSV_i}^{eff}}{(R_{TSV_i}^{eff})^2}$ under this condition.

It can be seen from Fig. 5.30 that for the given values of the $R_{eff}^{TSV_i}$, and the $L_{eff}^{TSV_i}$, the decoupling capacitance requirement decreases by increasing the value of the $U_{R_i}$ (i.e., ESR or $R_{dec_i}$), and by decreasing the value of the $U_{L_i}$ (i.e., ESL or $L_{dec_i}$).

![Normalized Decoupling Capacitance Plot](image)

**Fig. 5.30** The plot obtained through (5.42), showing the magnitude of the normalized decoupling capacitance vs. $U_{R_i}$ and $U_{L_i}$.

The target impedance must be met to control the current transients at the targeted frequency range [44]. Meeting the target impedance in the frequency domain keeps the noise below a specified limit in the time domain [44]. The decoupling capacitors placed in parallel with lower values of ESR, and ESL should be used to meet the target impedance.
We need to start with the decoupling capacitors having the lower ESR and lower ESL values in order to reduce the number of decoupling capacitors placed in parallel. Using this technique we can avoid the anti-resonance caused by different parallel decoupling capacitors. The voltage output across the current source of Fig. 5.28 as a result of the step output current is given as follows

\[ v(t) = V_{dd} - I_{L_1}(t)R_{TSVi} - I_{L_2}(t)L_{TSVi}C_{dec} e^{\frac{R_{TSVi}}{L_{TSVi}}t} \sin(\omega t - \theta) \]

where

\[ \omega_r = 2\pi f_r \]

\[ \theta = \tan^{-1}\left(\frac{X_{TSVi} - X_{C_{dec}}}{R_{TSVi}}\right) \]

The transient term in (5.41) shows peak deviation of the output voltage from its steady state value because of the step output current. The design of the decoupling capacitance for a particular load on the power distribution TSV pair is of utmost importance in controlling this deviation in the output voltage. The transient term dies out when the time reaches a certain value after it occurs but it may not die before the next clock cycle and may cause the noise accumulation for the subsequent clock cycles, and ultimately the design failure. When the effective inductance of the TSV pair is very low and the system is critically damped with the damping factor equal to one, the magnitude of the transient term shown by (5.43) also approaches zero. Assume that the effective TSV resistance is very low i.e. \( R_{TSVi} \to 0 \), which is quite possible if thousands of power distribution TSV pairs are used in parallel; then the voltage across the load as shown by (5.41) is only dependent on the transient term.

In order to keep the transient term within a prescribed limit for the given values of \( L_{TSVi} \) and \( I_{L_1}(t) \), the decoupling capacitor can be sized by using the term \( I_{L_2}(t)L_{TSVi}C_{dec} \) as mentioned in (5.43).

The damping factor of the power distribution TSV pair shown in Fig. 5.27 is unity i.e. \( \zeta = 1 \) when \( Z_{TSVi} = R_{TSVi}, Q_{TSVi} = 1, U_{R_i} = 1, \) and \( U_{L_i} = 0 \).

This means that the TSV effective inductance is bypassed by the decoupling capacitance and consequently the transient term has no effect if the decoupling capacitance is designed using the following equation and keeping the system critically damped.
The time domain voltage across the output load is
\[ v(t) = V_{dd} - I_{L} \cdot R_{TSV}^{eff} \] using (5.43), under the above-mentioned conditions. This confirms that the effective inductance \( L_{TSV}^{eff} \) of the power distribution TSV pair is bypassed when the value of the decoupling capacitance is equal to \( \frac{L_{TSV}^{eff}}{\sqrt{(R_{TSV}^{eff})^2}} \), ESR is equal to \( R_{TSV}^{eff} \), and ESL is equal to zero. As a result the impedance between the input source and the current source in the power distribution TSV pair shown in Fig. 5.27 is equal to \( R_{TSV}^{eff} \). On the other hand, for a nonzero value of \( U_{R} \), an optimally damped and almost flat transient voltage response can be obtained by selecting the values of \( U_{R}, U_{L}, \) and \( Q_{TSV} \) by using (5.38).

Hence, it is concluded that output impedance across the load at each stage of the vertical chain of power distribution TSV pairs shown in Fig. 5.26 is almost resistive if we choose the decoupling capacitance and the associated ESR and the ESL for each stage following the above mentioned guidelines. Therefore, the target impedance at each stage is defined such that it is greater than or equal to the sum of the equivalent resistances of that stage and all the preceding stages, assuming that the impedance for each of the preceding stages is estimated using the same guidelines. The guidelines should be applied in the following sequence to an arbitrary \( mth \) power distribution TSV pair of the vertical chain shown in Fig. 5.26:

1) Compute the target impedance across load for the \( mth \) TSV pair as follows

\[ Z_{T_m} = \frac{V_{dd} \times \text{Allowed ripple}}{\text{Load current at output of the stage } m} \]

2) Choose the output impedance as follows

\[ Z_{P_m}^{eff} = (R_{TSV_1}^{eff} + R_{TSV_2}^{eff} + \ldots \ldots + R_{TSV_m}^{eff}) \leq Z_{T_m} \]

3) Estimate the ESR by keeping

\[ R_{dec_m}^{eff} = U_{R_m} R_{TSV_m}^{eff} \]

4) Estimate the ESL by keeping

\[ L_{dec_m}^{eff} = U_{L_m} L_{TSV_m}^{eff} \]

5) Estimate the value of the decoupling capacitance as follows

\[ C_{dec_m} = \frac{L_{TSV_m}^{eff}}{(Q_{TSV_m} R_{TSV_m}^{eff})^2} \text{ for } U_{R_m} = 1 \text{ and } |Z_{P_m}^{eff}| = 1 \]
\[
C_{\text{dec}} = \frac{L_{\text{TSV}}^{\text{eff}}}{(Q_{\text{TSV}} R_{\text{TSV}}^{\text{eff}})^2}
\]

where

\[
Q_{\text{TSV}} = \frac{(1 + U_{R_e})}{2\sqrt{1 + U_{L_e}}} \text{ for } \xi = 1.
\]
Chapter 6
Summary and Future Work

6.1 Summary
Power distribution is one of the critical challenges for three-dimensional 3D integrated circuits. The integration density and speed of the on-chip logic has to be increased in order to efficiently utilize the bandwidth offered by 3D integration technology. The core logic switches at higher speed than I/O drivers in 3D high computation systems. The core switching noise is a limiting factor for 3D stacked power distribution networks. To the Author’s knowledge, the core switching noise has not yet been specifically addressed for 3D power distribution networks in the literature.

The size of a 3D power distribution network increases with the addition of each die in the vertical stack. Three-dimensional power distribution networks may expand to billions of nodes. However, simulation of 3D power distribution network cannot be handled using conventional simulation tools due to their slow simulation speed and limited memory resources. In addition to this, the simulation tools for 3D power distribution networks are still in the process of evolution. Fast and accurate simulation and modeling techniques for the analysis of 3D power distribution networks are required for early design tradeoffs. Three different mathematical models have been proposed in order to address the speed and memory issues for the simulation of 3D power distribution networks.

The power distribution TSV pair is the main constituent part of a 3D stacked power distribution network. Optimizing the design parameters of a power distribution TSV pair along with decoupling capacitance is necessary for the design of a 3D power distribution network. A mathematical model to estimate the peak ground noise across a power distribution TSV pair with decoupling capacitance and logic load is proposed. The model is based on time domain analysis. The model shows that peak ground noise is a function of the rise time and directly proportional to peak switching current. In addition to this the peak ground noise depends on the capacitive reactance and inductive reactance of the power distribution TSV pair. The model shows that peak ground noise for 3D PDN occurs at the worst case rise time in the time domain analysis which is analogous to resonance frequency in a frequency domain analysis. The worst case rise time depends on the inductance of the power distribution TSV and the amount of decoupling capacitance.
On-chip decoupling capacitance is used to suppress the core switching noise. The output impedance of the power distribution network should be closer to the target impedance for an optimum power distribution network design. The guidelines for the estimation of the decoupling capacitance to achieve the output impedance closer to the target impedance for each stage of a vertical of the power distribution TSV pairs have been proposed in this thesis.

The 3D-DRAM-Over-Logic system is a promising application for 3D integration technology. The decoupling capacitance placed on DRAM dies instead of the logic dies saves some useful area of logic dies for a 3D-RAM-Over-Logic system. A comprehensive mathematical technique to assign the optimum amount of decoupling capacitance to each of the DRAM dies is proposed in this thesis. The model shows a tradeoff between the resistance of interconnecting TSVs and the amount of decoupling capacitance on each DRAM die.

6.2 Future Directions

There is a great deal of research activity in the 3D integration paradigm. There are different challenges facing 3D integration; for example, power integrity and heat dissipation are both major challenges. Power integrity and heat dissipation are interlinked to each other. Power integrity should be addressed in conjunction with the heat dissipation. Therefore, thermal and power distribution co-design would be interesting for the future of 3D integration technology.

Chip, package, and board are physically and electrically interlinked to each other. The impedance of the power distribution network of package and board also affects on-chip switching noise. The power distribution network should address chip-package-board together for comprehensive power integrity. There is a great deal of literature about chip-package and board co-design for 2D power distribution network. For 3D stacked power distribution networks, however, chip, package, and board needs to be addressed together in future in order to guaranty overall power integrity.

Knowing that core switching current is much higher than the drivers current for high computing 3D power distribution networks, the effects of core switching noise on I/O driver’s noise need to be explored. The modeling of core switching noise and I/O driver’s noise along with the mutual effects would be an interesting future topic.
The resistance of power distribution TSVs may vary due to extra heat in a 3D stacked power distribution network. This also gives rise to thermal and mechanical stresses in a 3D stacked package. Power integrity for a 3D stacked power distribution network should also be viewed in conjunction with thermal and mechanical stresses in future.

The power distribution network for a 3D Network-on-Chip would be an interesting future topic where vertical through-silicon vias can be used for power distribution along with the signal transmission. The coupling of core switching noise to signal can be an issue in this case. It would be quite interesting to ensure the required amount of voltage for proper operation of intelligent networks associated with each node of 3D Network-on-Chip.

The convergence of biology, chemistry, and digital technology is possible using 3D integrated circuits. This type of small integrated circuits can be encapsulated into the human body in order to monitor health parameters. Power distribution network design for this type of three-dimensional integrated circuits can be an interesting future topic.
References


[46] 3D ICs with TSVs-Design Challenges and Requirements, Cadence White Paper. 


