Title
Modeling and Optimization for High-speed Links and 3D IC

Author
Yao, Wei

Publication Date
2012

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Modeling and Optimization for High-speed Links and 3D IC

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Wei Yao

2012

Abstract of the Dissertation

Modeling and Optimization for High-speed Links and 3D IC

by

Wei Yao

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2012

Professor Lei He, Chair

The advance of modern integrated circuit (IC) processes has supported increasing date rates on chip-to-chip communications in many consumer and professional applications, such as multimedia and optical networking. Serial links have successfully evolved and achieved the bit-rate of several tens of Gb/s per channel by applying new generations of IC process and advanced circuit techniques. However, as process technologies further scale down, severe process variations significantly impact the performance of high speed serial links and makes todays circuit designs have to be optimized not only for nominal performance but also for a reasonable yield. On the other hand, three-dimensional (3D) IC provides a smaller form factor, higher performance, and lower power consumption than conventional 2D integration by stacking multiple dies vertically. Through-silicon-via (TSV) enables the vertical connectivity between stacked dies or interposer and is a key technology for 3D IC. However, electrical signaling over TSVs presents a unique set of design challenges and thus requires accurate modeling and detailed signal and power integrity analysis.

In this research, the bottlenecks in TSV modeling, variation-aware circuit optimization and efficient performance evaluation for high bit-rate applications are

analyzed, and solutions are presented. A simple yet accurate pair-based model for multi-port TSV networks (e.g., coupled TSV array) is proposed by decomposing the network into a number of TSV pairs and then applying circuit models for each TSV pair. This methodology is first verified against full-wave electromagnetic (EM) simulation for up to 20GHz and subsequently employed for a variety of examples of signal and power integrity analysis. For high speed serial links, an optimization framework is proposed for the joint design time and post-silicon tuning optimization for digitally tuned analog circuits, and can be used to maximize the yield in serial link transmitter design and the phase-locked-loop (PLL) design subject to the area and power constraints. Moreover, an efficient mathematical method is proposed to capture the worst-case data-dependent jitter and noise without lengthy simulations. These modeling and optimization methodologies can be applied to accurately explore the chip-to-chip integration and signaling schemes at early design stage in today's and tomorrow's 3D IC and high speed serial link design.

The dissertation of Wei Yao is approved.

M.-C. Frank Chang

Sudhakar Pamarti

Rupak Majumdar

Lei He, Committee Chair

University of California, Los Angeles

2012

To My Family and Friends.

Table of Contents

1 Introduction **1**

2 Worst-Case Data-Dependent Jitter and Amplitude Noise for Differential Signaling **5**

2.1 Introduction 5

2.2 Preliminaries 8

2.2.1 RLGC Model for Transmission Line 8

2.2.2 Parallel Transmission Lines 9

2.2.3 Pre-emphasis Filter 12

2.3 Jitter and Amplitude Noise Model 13

2.3.1 Subcomponents of Jitter and Amplitude Noise 14

2.3.2 Modeling of Data-Dependent Jitter and Noise 15

2.4 Worst-case Timing Jitter 17

2.4.1 Problem Formulation 18

2.4.2 Relaxation Based Binary Search 19

2.4.3 Pseudo-Boolean Optimization 20

2.4.4 Proposed Heuristic 21

2.5 Worst-case Amplitude Noise 24

2.6 Experimental Results 25

2.6.1 Jitter and Amplitude Noise Model Validation 25

2.6.2 Worst-case Jitter and Amplitude Noise Calculation 27

2.7 Conclusion 29

3 Joint Design-Time and Post-Silicon Optimization for Digitally Tuned Analog Circuits 30

3.1 Introduction 30

3.2 Preliminaries on Digitally Tuned Analog Circuits 33

3.3 Problem Formulation 35

3.4 Optimization Framework 37

3.4.1 Algorithm Overview 38

3.4.2 Partitioning and Bound Estimation 40

3.4.3 Gradient Ascent Method 41

3.5 Transmitter Design in High-speed Serial Link 42

3.5.1 Design-time Optimization 44

3.5.2 Post-silicon Tuning and Joint Optimization 46

3.6 PLL Design in High-speed Serial Link 47

3.6.1 Design-time Optimization 48

3.6.2 Post-silicon Tuning and Joint Optimization 50

3.7 Experimental Results 52

3.7.1 Transmitter Design 52

3.7.2 PLL Design 56

3.8 Conclusions 57

4 Modeling and Application of Multi-Port TSV Networks in 3D IC 58

4.1 Introduction 58

4.2 Preliminary on TSV Modeling 60

4.3 Multi-port TSV Network Modeling 63

4.3.1 Framework Overview 64
4.3.2 Pair-based Equivalent Circuit Model 69
4.3.3 Model Validation via Simulation 71
4.4 Multi-port TSV Network Characteristics in Signal and Power Integrity 72
4.4.1 Crosstalk Analysis for Chip-to-chip TSV Networks in Silicon Interposer 73
4.4.2 Impedance Analysis for Power/Ground TSV Array in 3D Power Distributed Network (PDN) 77
4.5 TSV Array Bandwidth Analysis for 3D IC 79
4.6 Conclusions 82

5 Modeling of RDL Coplanar Waveguide on Silicon Interposer . 83
5.1 Introduction 83
5.2 CPW Equivalent Circuit Model: Shunt Admittance 84
5.3 CPW Equivalent Circuit Model: Series Impedance 88
5.4 Model Validation via Simulation 91
5.5 Conclusion 92

6 Power-Bandwidth Trade-off on TSV Array and TSV-RDL Junction Design Challenges on Silicon Interposer 94
6.1 Introduction 94
6.2 Signal Integrity Analysis on TSV Array 96
6.3 TSV Power and Bandwidth Trade-off Analysis 99
6.4 Design Challenges on Junction Structures between TSV and RDL Trace 103

6.5 Conclusions 106

7 Conclusion 107

A Mathematical Proof for Pair-based TSV Array Model 109

B 3D IC Technology and TSV Dimension in ITRS Roadmap . . 112

References 114

List of Figures

1.1 CMOS threshold voltage variation 1

1.2 Beyond-die Integration Options. 2

1.3 Structure of the dissertation, showing how the various topics relate to each other. 3

2.1 Eye diagram and eye mask 6

2.2 Transceiver block diagram for differential signaling 7

2.3 Section of differential microstrip line 8

2.4 Parallel transmission lines 11

2.5 Pre-emphasis filter at transmitter end for CML differential signaling 13

2.6 Eye diagram (left) without the pre-emphasis filter and (right) with applying a 4-taps pre-emphasis filter 14

2.7 Differential signaling with parallel transmission lines and termination 16

2.8 Time domain response of the channel. 18

2.9 Transient simulation comparison between (left) SPICE and (right) our model. The origin point is different. 26

2.10 Time domain response: SPICE simulation(left) and MATLAB simulation with our model(right). The origin point is different. 27

3.1 Examples of digitally tuned analog circuits: (a) CMOS current source and (b) capacitance array. 31

3.2 V_{th} variation model (a) and current mirror with V_{th} mismatch (b). 34

3.3 Post-silicon tuning through DAC 35

3.4 (a) Unit cell design. (b) Swing and delay vs. number of parallel-connected cells. 39

3.5 System diagram of a high-speed serial link. 43

3.6 Power and performance variation for 1000 die samples by Monte Carlo simulation: (a) without tuning and (b) with tuning. 46

3.7 Tunable and adaptive bandwidth PLL. [SLK00] 48

3.8 Output jitter sensitivity to the (a) loop damping factor ζ and (b) charge pump current ratio I_{CP2}/I_{CP1}. 49

3.9 Probability density for output jitter(%). (a) without tuning (b) with tuning circuit and optimized digital control. 51

3.10 Charge pump schematic [SLK00]. 52

3.11 BER distribution for four different designs. 53

3.12 Yield curves for our designs and design heuristic with respect to area (a), power (b) and V_{th} (c). 55

3.13 Yield for our algorithms and design heuristic w.r.t power (a) and V_{th} (b) in the PLL. 56

4.1 Multi-port TSV network example: A $M \times N$ TSV array 59

4.2 Structures and dimensional variables of a single TSV pair and its RLGC equivalent circuit model [XLS10]. 61

4.3 The charge distribution and the inductance comparison of a 3-TSV network show the proximity effect can be neglected when TSV pitch is larger than 6× TSV radius. The equivalent inductance is calculated at 20GHz based on our methodology and the TSV pair model in [XKS11]. The results are compared to Ansoft Q3D [ANS] with various TSV radius and pitch/radius ratios. 65

4.4 2x2 TSV differential pair differential S-parameter comparison. The TSV diameter is 25μm and pitch is 150μm 71

4.5 TSV Networks in silicon interposer for 2.5D chip-to-chip communication 72

4.6 4x4 TSV array RLGC comparison for (1) self terms of TSV_A and (2) mutual terms between TSV_A and TSV_1. 73

4.7 S Parameters of NEXT and FEXT between TSV_A and TSV_1 with different pitch sizes and between TSV_A and other TSVs with pitch $40\mu m$ 74

4.8 Normalized crosstalk voltages between TSV_A and TSV_1. The noise voltages are separated to capacitive and inductive couplings using (4.40) and (4.41). The pitch size is 40 μm. 75

4.9 Time-domain crosstalk voltage at near and far ends between TSV_A and TSV_1, with 1V signal strength and 30ps rise time. 77

4.10 Power/ground TSV array arrangements and system-level power distribution network (PDN) 78

4.11 P/G TSV array impedance v.s. (a) different P/G TSV array arrangement (b) different pitch size and (c) different diameter. . . . 78

4.12 3D PDN impedance comparison between different P/G TSV array arrangements. 79

4.13 3D IC TSV array bandwidth analysis considering different layers of stacking and different CMOS technology nodes. 80

4.14 TSV array date rate per TSV and power efficiency vs number of stacked silicon layers and technology nodes. 80

4.15 Eye diagram of (a) 1 layer of 4×4 TSV array at 24.13Gbps and (b) 12 layers of 4×4 TSV arrays at 2.41Gbps. Both cases are assumed using 16nm technology and with the same driver driving strength. 81

5.1 Coplanar waveguide on silicon interposer. 85

5.2 Geometries to calculate the coefficients in the $\boldsymbol{p_s}$ matrix. 87

5.3 Equivalent circuit model for the shunt admittance between signal and ground conductors of CPW. 88

5.4 Complex image of line current source in SiO_2-Silicon media. The current flows in the opposite direction in the image lines. 89

5.5 Comparisons of the series impedance for on-chip CPW structures between different silicon conductivity. 90

5.6 Shunt admittance obtained from our model is compared with simulated results using FEM: (left) capacitance (right) conductance. . 91

5.7 Derivations of capacitance and conductance for on-chip CPW using FEM and the model proposed. The x-axis is the ratio of trace spacing to trace width. The y-axis is the average deviation in percentage from 50 MHz to 50 GHz. 92

6.1 Concept of 3D IC using TSV and silicon interposer 95

6.2 (a) 4×4 TSV array and (b) an example of junction structure between TSVs and interposer RDL traces 95

6.3 An array of cylindrical Cu-filled TSVs and the differential S_{21} and S_{11} on one of the differential TSV pair (marked in red). 96

6.4 The differential impedance (Z_{diff}) and eye diagram comparison between $30\mu m$ and $90\mu m$ TSV array pitch. The TSV diameter is $10\mu m$ and height is $150\mu m$. 97

6.5 The differential impedance (Z_{diff}) comparison when changing the center-to-center distance between differential signal TSVs from $30\mu m$ to $90\mu m$ while keeping the signal to ground distance fixed at $30\mu m$. 98

6.6 Signal-to-Crosstalk-Ratio (SCR) comparison on the 4×4 TSV array when considering (a) single-ended signaling and (b) differential signaling. 99

6.7 Maximum achievable data rate per TSV in a 4×4 TSV array versus various TSV pitches and diameters. 100

6.8 Different types of drivers and terminations: (a) push-pull single-ended with resistive termination, (b) push-pull single-ended with pure capacitive loading, and (c) differential CML driver. $R_0 = 40\Omega$ and $C = 2pF$. 101

6.9 (a) Maximum achievable data rate per TSV versus input swing voltages and (b) corresponding power efficiency versus input swing voltages for various output drivers and terminations shown in Fig. 6.8. 101

6.10 Maximum achievable data rate per TSV using single-ended signaling versus driver input swing with different TSV array patterns: (a) G-S-S-G column-by-column and (b) distributed signal and ground. 102

6.11 Fanout-like junction structure between TSV and RDL and the full link S_{21} and S_{11} comparison for different TSV pitches. 104

6.12 G-S-G TSV S_{21} and S_{11} comparison for different TSV pitches. . . 104

6.13 Full link S_{21} and S_{11} comparison between two different junction structure between TSV and RDL. 105

List of Tables

2.1 Differential transmission line testbench design information: width(w), spacing(s), thickness(t), dielectric height(h), length(L) and characteristic impedance. 26

2.2 Jitter and amplitude noise model validation. 27

2.3 Worst-case jitter, amplitude noise and runtime comparison: PRBS denotes pseudo-random bit sequence. For worst-case jitter, relaxation-based binary search (BS) is used along with pseudo-Boolean optimizer (PBO) or our heuristic algorithm (Heuristic). For worst-case noise, LP denotes directly solving linear programming. 28

4.1 Mutual inductance and capacitance between TSV_A and TSV_1 at $10GHz$ with different pitch sizes. 76

B.1 High-density TSV projections in 2008 ITRS update [Ass08]. . . . 113

Vita

1998–2002	B.S., Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan.
2002–2004	M.S., Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan.
2006–present	Ph.D. program, Department of Electrical Engineering, University of California, Los Angeles, California, USA. Teaching Assistant, Digital Design Laboratory, Fall 2007, Winter 2008. Teaching Assistant, Logic Design for Digital Systems, Winter 2009, Fall 2010. Graduate Student Researcher, UCLA Design Automation Lab, 2006–present.
2004–2006	Engineer, Synopsys, Inc., Taipei, Taiwan. Worked on chip-level static and dynamic IR drop analysis and power EM analysis tool, PrimeRail®.
2009-2010	Engineer, Taiwan Semiconductor Manufacturing Company, Ltd., Hsinchu, Taiwan. Worked on ASIC design flows, including decap insertion, signal EM, ERC, and low-power design netlist check.
2011	Research Intern, Cisco Systems, Inc., San Jose, California, USA. Worked on SiP power integrity, TSV modeling and its signal integrity, and 3D IC design exploration.

Publications

Yiyu Shi, **Wei Yao**, Jinjun Xiong, and Lei He, "Incremental and On-demand Random Walk for Iterative Power Distribution Network Analysis," *Asia and South Pacific Design Automation Conference*, 2009.

Wei Yao, Yiyu Shi, Lei He and Sudhakar Pamarti, "Worst Case Timing Jitter and Amplitude Noise in Differential Signaling," *International Symposium on Quality Electronic Design*, 2009.

Wei Yao, Yiyu Shi, Lei He, and Sudhakar Pamarti, "Joint Design-Time and Post-Silicon Optimization for Digitally Tuned Analog Circuit," *International Conference on Computer Aided Design*, 2009.

Yiyu Shi, **Wei Yao**, Lei He, and Sudhakar Pamarti, "Joint Design-Time and Post-Silicon Optimization for Analog Circuits: Case Study using High-Speed Transmitter," *Semiconductor Research Corporation Techcon Conference*, 2009.

Wei Yao, Yiyu Shi, Lei He, and Sudhakar Pamarti, "Timing Jitter and Amplitude Noise Modeling and Estimation for Differential Signaling," *Semiconductor Research Corporation Techcon Conference*, 2009.

Lei He, Shauki Elassaad, Yiyu Shi, Yu Hu and **Wei Yao**, "System-in-Package: Electrical and Layout Perspectives," *Foundations and Trends in Electronic Design Automation*, 2011.

Wei Yao, Yiyu Shi, Lei He, and Sudhakar Pamarti, "Worst-Case Estimation for

Data-Dependent Timing Jitter and Amplitude Noise in High-Speed Differential Link," *IEEE Transactions on Very Large Scale Integration Systems*, 2012.

Wei Yao, Feng Shi, Lei He, Siming Pan, Brice Achkir, and Li Li, "Power-Bandwidth Trade-off on TSV Array in 3D IC and TSV-RDL Junction Design Challenges,", *Electrical Performance of Electronic Packaging and Systems Conference*, 2012.

Wei Yao, Siming Pan, Jun Fan, Brice Achkir, and Lei He, "Modeling and Application of Multi-Port TSV Networks in 3D IC,", reviewed by *IEEE Transactions on Computer-Aided Design of Integrated Circuits and System*, 2012.

CHAPTER 1

Introduction

High-speed I/O links for digital communication systems is dramatically developed and improved in this decade. Per-pin bit rate of multi-processor interconnections and processor-to-memory interfaces have been aggressively increased to scale aggregate I/O bandwidth. However, as process technologies are scaled down to finer feature size and circuit applications move to higher frequency bands, analog/mixed-signal design faces several new challenges. First, device models have become increasingly complex in order to capture the physical behavior of nano-scale transistors at high operation frequencies. Also, parasitic couplings become more important and more complex. Moreover, because of the sub-wavelength photo-lithography and random doping fluctuation, process variation become a critical issue and significantly impact the performance of mixed-signal circuit [Ass05], as shown in Fig. 1.1.

As the variation increases along with the technology scaling, tradition corner-

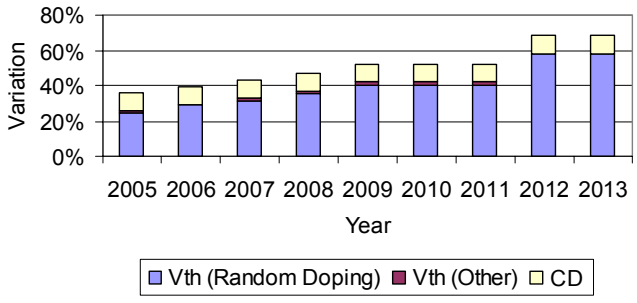

Figure 1.1: CMOS threshold voltage variation

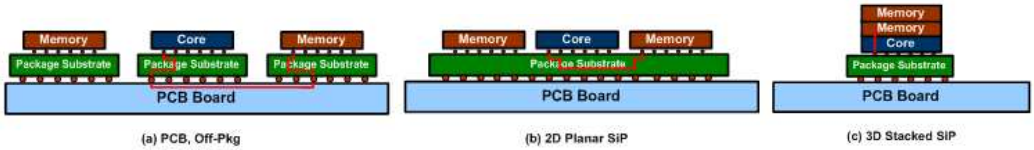

Figure 1.2: Beyond-die Integration Options.

based design methodology is becoming insufficient and may not be viable eventually. Statistical design, on one hand, is proposed to analyze the performance distribution from process variation and defines *parametric yield* as the probability of the design meeting a specified performance or power constraint [BKN03, MSO06]. Today's mixed-signal circuit designs, as a result, not only have to be optimized for nominal performance but also for robustness in order to maintain a reasonable yield.

On the other hand, three-dimensional integration to create multilayer chips (3D ICs) offers an exciting alternative to traditional scaling and provides high integration density, fast signal transmission, low power consumption, and heterogeneous integration opportunities in the "More-than-Moore" era [BSK01, BAC07], as shown in Fig. 1.2. Through-silicon-via (TSV) has been well regarded as a key component in 3D integration, connecting chips vertically with shortened electrical delay and providing extremely dense I/O connections. While TSV fabrication technologies have progressed [PBW08], it is vitally important to understand TSV electrical properties accurately and efficiently for 3D system-performance analysis and subsequent design optimization. At the same time, though TSV could potentially provide thousands of multi-Gb/s I/O and support for tens of Tb/s data bandwidth between local chips due to its short distance and superior signal integrity [LLF12, SCA10], detailed analysis considering various TSV array patterns and geometries with different signaling and termination techniques is still desperately required in order to find a balance between power and bandwidth under different design constraints.

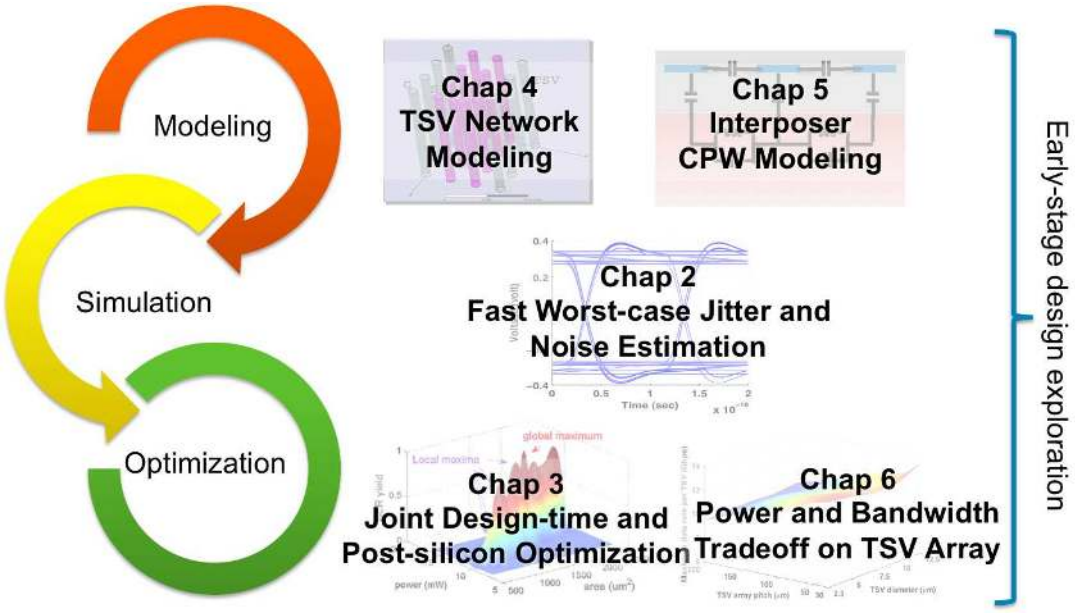

Figure 1.3: Structure of the dissertation, showing how the various topics relate to each other.

The main topics addressed in this dissertation and how they are related to each other are illustrated in Fig. 1.3. The reminder of this dissertation is organized as follows. In order to quickly evaluate the high-speed link's signal integrity performance, an efficient mathematical method is proposed to calculate the worst-case data-dependent jitter and noise for differential signaling. In Chapter 2, we first propose formula-based jitter and noise models considering the combined effect of ISI, crosstalk, and pre-emphasis filter. Moreover, using the formula-based models, we develop mathematical programming algorithms to directly find out the input patterns for worst-case jitter and worst-case amplitude noise. This performance metric can be used for efficient circuit evaluation instead of lengthy simulation and measurements. In Chapter 3, a joint design time and post-silicon optimization framework based on the branch-and-bound algorithm and gradient ascent method is proposed for digitally tuned analog circuits to maximize the parametric yield, subject to power and area constraints. We demonstrate our framework

with two examples in high-speed serial link, the transmitter design and the phase-locked-loop (PLL) design. Simulation results show the proposed joint design-time and post-silicon optimization can significantly improve the yield from the design heuristic in analog designers' perspective.

In Chapter 4, a simple yet accurate pair-based model for multi-port TSV networks (e.g., coupled TSV array) is proposed by decomposing the network into a number of TSV pairs and then applying circuit models for each TSV pair. The model is first verified against a commercial electromagnetic solver for up to 20GHz and subsequently employed for a variety of examples for signal and power integrity analysis. Chapter 5 derives a rigorous frequency-dependent circuit model for horizontal coplanar waveguide (CPW) on silicon interposer based on partial equivalent element circuit (PEEC) in layered metal-isolator-semiconductor (MIS) media. Chapter 6 studies the signal integrity issues of TSV-based 3D IC with high-speed signaling based on 3D electromagnetic field solver and SPICE simulations. Power and bandwidth trade-off on TSV array are also analyzed between different signaling and termination techniques, such as single-ended, differential and reduced-swing signaling. Beyond TSV, critical design challenges for the junction structure between TSVs and RDL traces are also revealed and analyzed in Chapter 6 as well.

CHAPTER 2

Worst-Case Data-Dependent Jitter and Amplitude Noise for Differential Signaling

2.1 Introduction

Differential signaling has been widely used in high-speed I/O interconnect standards like PCI-Express and Serial ATA. It has several advantages, such as a high transmission rate due to low signal swing, little electromagnetic interference (EMI), and common-mode noise immunity. Considerable signal integrity issues, however, still limit the link performance and become bottlenecks during system integration. Such issues include resistive losses, reflections, inductive ringing and crosstalk between differential pairs [SH03, BH05].

To evaluate the combined effect of these impairments on the overall system performance, the associated eye diagram [Hay00, Bre05] has been used as an effective measure. As shown in Fig. 2.1, the eye diagram is defined as the synchronized superposition of all possible realizations of the signal viewed within a particular signal interval. It provides a fast evaluation of system performance. The width of the eye opening defines the time interval over which the received signal can be sampled without error. The height of the eye opening with the amount of amplitude noise at a specified sampling time defines the signal-to-noise-ratio of the received signal [Hay00].

Consider the eye diagram shown in Fig. 2.1. The amounts of timing jitter and amplitude noise determine the width and height of the eye. Jitter is defined as

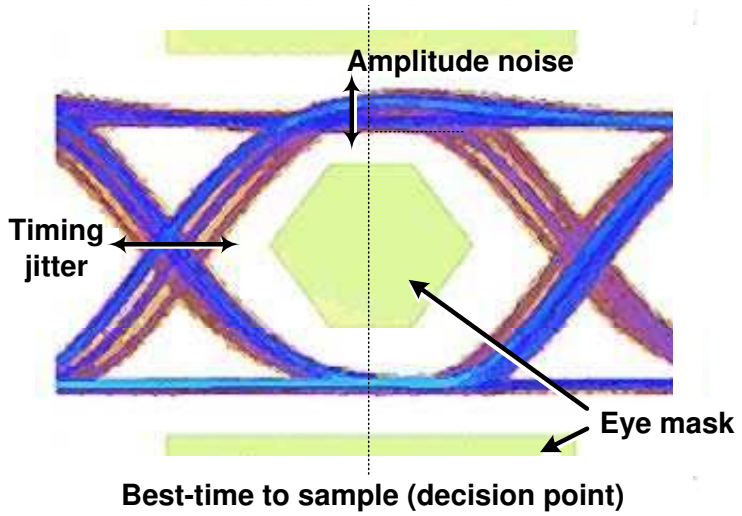

Figure 2.1: Eye diagram and eye mask

the deviation of the zero-crossing from its ideal occurrence time and decreases the eye's width [KRF05]. Amplitude noise, on the other hand, decreases the signal-to-noise ratio (SNR) and, accordingly, the eye's height. As a result, to determine the performance of the interconnect, we need to consider both timing jitter and amplitude noise simultaneously.

Specifically, inter-symbol interference (ISI) and crosstalk are two major factors that induce jitter and amplitude noise. ISI is defined as one symbol interfering with subsequent symbols and is caused by channel impairments such as attenuation, reflection, and group delay distortion. Crosstalk, on the other hand, is caused by electromagnetic coupling between transmission lines. To counteract ISI, a finite impulse response (FIR) pre-emphasis filter at the transmitter side is widely used to emphasize the signal prior to the impact of the channel [LWT04, TBP05, KLS06]. Pre-emphasized signal, on the other hand, also affects coupled electromagnetic energy and changes crosstalk behavior. As a result, for both ISI and crosstalk, it is important to take pre-emphasis filter into consideration.

Traditionally, the eye diagram is obtained through lengthy time domain simulation with pseudo-random bit sequences as the input data. In the litera-

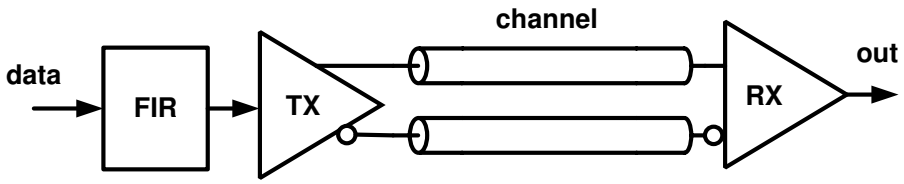

Figure 2.2: Transceiver block diagram for differential signaling

ture, several types of techniques were proposed to model the eye diagram and tried to efficiently predict the jitter and amplitude noise at the design phase [HST07, PKT07, BAH04, BA04, OFK04, SYZ08]. However, [HST07] considers reflection and attenuation with only one input pattern, and [PKT07] only considers a lossless transmission line. As a result, those models are far from accurate. [BAH04,BA04] and [OFK04] have a better model because they consider lossy transmission lines, but they still take only a few input patterns into account and use an inaccurate linear approximation. To efficiently handle the input patterns in adequate length, [SYZ08] uses the pull-up and pull-down transitions of the signal step response to predict the worst-case eye diagram. This waveform-based method, however, is very sensitive to the transmission environment and only applies to single transmission line without any crosstalk noise. Most importantly, all existing works fail to consider the effect of the pre-emphasis filter, which effectively reduces ISI, as shown in Fig. 2.6 in Section II, but at the same time affects crosstalk.

In this chapter, we first propose formula-based jitter and amplitude noise models that consider the combined effect of ISI, crosstalk, and the pre-emphasis filter for differential signaling. With the same set of input patterns, experiment shows our models achieve within 5% difference compared to SPICE simulation. Note that we apply a RLGC lossy transmission line model according to differential microstrip line geometry, and we represent the channel impairments and crosstalk through transmission line time domain response. By using these formula-based

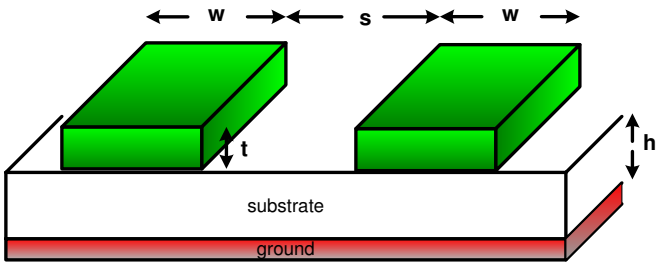

Figure 2.3: Section of differential microstrip line

models, we then develop algorithms to directly predict the input patterns that cause worst-case jitter and worst-case amplitude noise through pseudo-Boolean optimization and mathematically programming. Moreover, heuristic algorithm is proposed to further reduce runtime. Experiments show our algorithms obtain more reliable worst-case jitter and noise compared to pseudo-random bit sequences simulation and, meanwhile, reduce runtime by 25× when using general PBO solver and by 150× when using our proposed heuristic algorithm.

2.2 Preliminaries

We first review the RLGC model for parallel transmission lines and the frequency domain relationships between input and output ports. Next, an overview of the transmitter pre-emphasis filter is provided and its impact on eye diagram is also demonstrated.

2.2.1 RLGC Model for Transmission Line

A cross-section of the differential microstrip line is shown in Fig. 2.3. We assume the lines are homogeneous, uniform, and parallel to each other without any variation [PKT07]. The dielectric is assumed to be homogeneous with constant permittivity ε and permeability μ.

The distributed self and mutual inductances are computed with the method of images [YNM03]: the effect of the ground plane is replaced with the image currents. The rectangular shapes of conductors were changed into circular ones for geometry simplification and the following expressions were found for the per-unit-length self and mutual inductances [YNM03]:

$$l = \frac{\mu}{2\pi} ln \left(1 + \frac{2H_{eq}}{r_{eq}}\right) \tag{2.1}$$

$$m = \frac{\mu}{4\pi} ln \left(\frac{(s_{eq} + 2r_{eq})^2 + (r_{eq} + 2H_{eq})^2}{(s_{eq} + 2r_{eq})^2 + r_{eq}^2}\right), \tag{2.2}$$

where H_{eq}, r_{eq}, and s_{eq} are the equivalent height, thickness and spacing of the differential microstrip line. l is self inductances and m is mutual inductance. The distributed capacitances may be calculated as follows [Pau06]

$$c_p = \frac{\mu\epsilon m}{l^2 + m^2} \tag{2.3}$$

$$c = \frac{\mu\epsilon l}{l^2 + m^2}, \tag{2.4}$$

where c is the distributed capacitance between the conductor and the ground and c_p is the distributed parasitic capacitance between the conductor lines.

By using these analytical parasitics models, the RLGC per-unit-length model for the differential microstrip lines can be established. Note that other field-solver-based tools can also be used for RLGC extraction and our jitter and noise modeling and estimation algorithm still apply.

2.2.2 Parallel Transmission Lines

High-speed signal propagation on an interconnect can be influenced by several effects, such as delay, attenuation, reflection, slew rate limitation, and crosstalk. All of these effects, which are also known as transmission line effects [AN01], can be captured by distributed transmission line equations with accurate RLGC per-unit-length model. To analyze the three-wire differential signaling, as shown in Fig. 2.4, we first consider the general multiconductor transmission line system.

Transmission line characteristics are in general described by Telegrapher's equations and per-unit-length $\boldsymbol{R, L, G, C}$ matrices [AN01, TN92]:

$$\frac{\partial}{\partial x}\boldsymbol{v}(x,t) = -\boldsymbol{R}\boldsymbol{i}(x,t) - \boldsymbol{L}\frac{\partial}{\partial x}\boldsymbol{i}(x,t) \tag{2.5}$$

$$\frac{\partial}{\partial x}\boldsymbol{i}(x,t) = -\boldsymbol{G}\boldsymbol{v}(x,t) - \boldsymbol{C}\frac{\partial}{\partial x}\boldsymbol{v}(x,t), \tag{2.6}$$

where $\boldsymbol{v}$ and $\boldsymbol{i}$ are voltage and current vectors. $\boldsymbol{R, L, G, C}$ are the per-unit-length resistance, inductance, conductance, and capacitance matrix for the transmission line. Taking the Laplace transform of (2.5) and (2.6), we can get

$$\frac{\partial}{\partial x}\boldsymbol{V}(x,s) = -\boldsymbol{Z}\boldsymbol{I}(x,s) \tag{2.7}$$

$$\frac{\partial}{\partial x}\boldsymbol{I}(x,s) = -\boldsymbol{Y}\boldsymbol{V}(x,s), \tag{2.8}$$

where $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are the impedance and admittance matrices, given by

$$\boldsymbol{Z} = \boldsymbol{R} + s\boldsymbol{L}, \boldsymbol{Y} = \boldsymbol{G} + s\boldsymbol{C}. \tag{2.9}$$

Further derivation could be achieved through multiconductor transmission line decoupling [Pau06, AN01, TN92] for lossy identical transmission lines. First differentiating the partial differential equations given in equations (2.7) and (2.8) with respect to x, we get following two coupled equations:

$$\frac{\partial^2}{\partial x^2}\boldsymbol{V}(x,s) = -\boldsymbol{Z}\boldsymbol{Y}\boldsymbol{V}(x,s) \tag{2.10}$$

$$\frac{\partial^2}{\partial x^2}\boldsymbol{I}(x,s) = -\boldsymbol{Y}\boldsymbol{Z}\boldsymbol{I}(x,s). \tag{2.11}$$

Decoupling of equations (2.10) and (2.11) can be achieved by introducing a transformation matrix $\boldsymbol{W}$ relating to actual circuit voltage $\boldsymbol{V}$ and modal voltage $\tilde{\boldsymbol{V}}$ [AN01].

$$\boldsymbol{V}(x,s) = \boldsymbol{W}\tilde{\boldsymbol{V}}(x,s). \tag{2.12}$$

Using equation (2.12), we could rewritten equations (2.10) as

$$\frac{\partial^2}{\partial x^2}\tilde{\boldsymbol{V}}(x,s) = -(\boldsymbol{W}^{-1}\boldsymbol{Z}\boldsymbol{Y}\boldsymbol{W})\tilde{\boldsymbol{V}}(x,s). \tag{2.13}$$

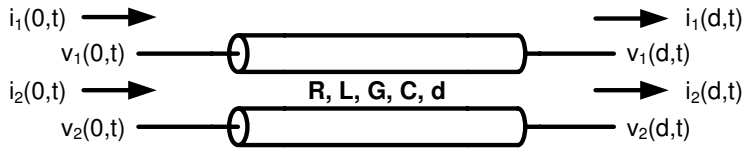

Figure 2.4: Parallel transmission lines

To successfully decouple the equations, the matrix product in parenthesis of equation (2.13) must be a diagonal matrix as

$$\boldsymbol{W^{-1}ZYW} = \begin{bmatrix} \gamma_1^2 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \gamma_N^2 \end{bmatrix}, \tag{2.14}$$

where the transformation matrix $\boldsymbol{W}$ corresponding to the eigenvectors of product $\boldsymbol{ZY}$. And the resulting diagonal matrix contains the eigenvalues $(\gamma_1, \cdots, \gamma_N)$ of product $\boldsymbol{ZY}$. The solution of equation (2.13) can then be written as

$$\tilde{\boldsymbol{V}}(x) = [\boldsymbol{E}(x)]C_1 + [\boldsymbol{E}(x)]^{-1}C_2, \tag{2.15}$$

and

$$\boldsymbol{V}(x) = \boldsymbol{W}[\boldsymbol{E}(x)]C_1 + \boldsymbol{W}[\boldsymbol{E}(x)]^{-1}C_2, \tag{2.16}$$

where $\boldsymbol{E}(x) = diag[e^{-\gamma_1 x}, \cdots, e^{-\gamma_N x}]$ and (C_1, C_2) are constants determined by boundary conditions.

Substituting equation (2.16) back to equation (2.7) , we have

$$\boldsymbol{I}(x) = \boldsymbol{W_i}[\boldsymbol{E}(x)]C_1 + \boldsymbol{W_i}[\boldsymbol{E}(x)]^{-1}C_2, \tag{2.17}$$

where

$$\boldsymbol{W_i} = \boldsymbol{Z^{-1}W\Gamma} \tag{2.18}$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \gamma_N \end{bmatrix}. \tag{2.19}$$

Combining equations (2.16) and (2.17) to eliminate constants (C_1, C_2), the voltage-current relationships between $x = 0$ and $x = d$, in frequency domain can be expressed as

$$\begin{aligned} \begin{bmatrix} \boldsymbol{I}(0) \\ -\boldsymbol{I}(d) \end{bmatrix} &= \begin{bmatrix} \boldsymbol{Y_{11}} & \boldsymbol{Y_{12}} \\ \boldsymbol{Y_{21}} & \boldsymbol{Y_{22}} \end{bmatrix} \begin{bmatrix} \boldsymbol{V(0)} \\ \boldsymbol{V(d)} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{W_i E_1 W^{-1}} & \boldsymbol{W_i E_2 W^{-1}} \\ \boldsymbol{W_i E_2 W^{-1}} & \boldsymbol{W_i E_1 W^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{V(0)} \\ \boldsymbol{V(d)} \end{bmatrix}, \end{aligned} \tag{2.20}$$

where

$$\boldsymbol{E_1} = diag\{\frac{1 + e^{-2\gamma_k d}}{1 - e^{-2\gamma_k d}}\} \tag{2.21}$$

$$\boldsymbol{E_2} = diag\{\frac{-2e^{-2\gamma_k d}}{1 - e^{-2\gamma_k d}}\}, \qquad k = 1, 2, \ldots, N \tag{2.22}$$

with $\boldsymbol{I}(0)$, $\boldsymbol{I}(d)$, $\boldsymbol{V}(0)$ and $\boldsymbol{V}(d)$ are Laplace transforms of $\boldsymbol{i}(0,t)$, $\boldsymbol{i}(d,t)$, $\boldsymbol{v}(0,t)$ and $\boldsymbol{v}(d,t)$, respectively. Here $Y_{11}, Y_{12}, Y_{21}, Y_{22}$ form the equivalent admittance matrix or Y-parameters of the transmission line. Please note that the admittance matrix can also be directly obtained from measured response or measured S-parameters.

2.2.3 Pre-emphasis Filter

Using a symbol-spaced finite impulse response (FIR) filter to pre-emphasize the signal at the transmitter end is a common way to counteract ISI. The filter can be expressed as

$$y(n) = \sum_{i=-N}^{M} W_i x(n-i), \tag{2.23}$$

where W_i is the coefficient for each filter tap and $x(n)$ and $y(n)$ are the corresponding filter input and output signal. A circuit implementation of the current-mode logic (CML) pre-emphasis driver is shown in Fig. 2.5. The coefficient of each tap is realized by the current source and requires a dedicated differential pair to drive

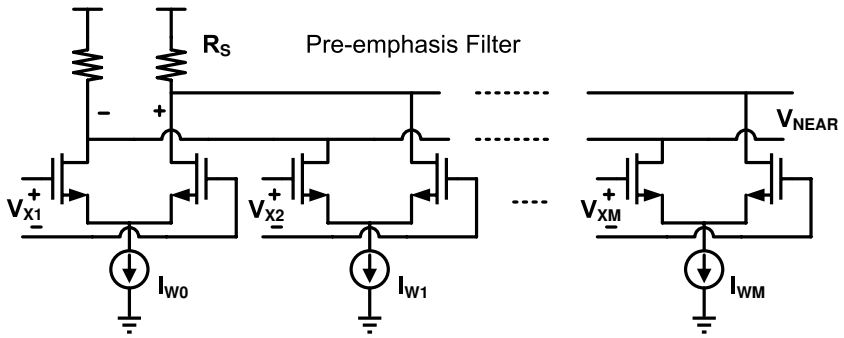

Figure 2.5: Pre-emphasis filter at transmitter end for CML differential signaling

the output. Normally, the number of taps ranges from 2 to 5 because of power and area constraints.

The coefficient of each tap is directly related to the channel characteristic mentioned in the previous sub-section and can be determined adaptively by the least-mean-square (LMS) algorithm [LWT04, TBP05]:

$$W_i^{k+1} = W_i^k + \mu \epsilon_k x_{k-i}, \tag{2.24}$$

where W is the tap coefficient and μ is the step size. ϵ_k is the error signal and is defined as the difference between the received signal value and the transmitted value. The convergence of errors drives the coefficients to their optimal value.

To demonstrate the effectiveness of the pre-emphasis filter, the eye diagram with and without the pre-emphasis filter is compared in Fig. 2.6. The SNR improvement can be clearly seen. As a result, jitter and amplitude noise models can't capture the actual link performance without considering the existence of pre-emphasis filter.

2.3 Jitter and Amplitude Noise Model

The jitter and amplitude noise are actual stochastic processes and can be divided into two categories: random and deterministic. The random part is usually described through a probability density function (PDF) or its root-mean-square

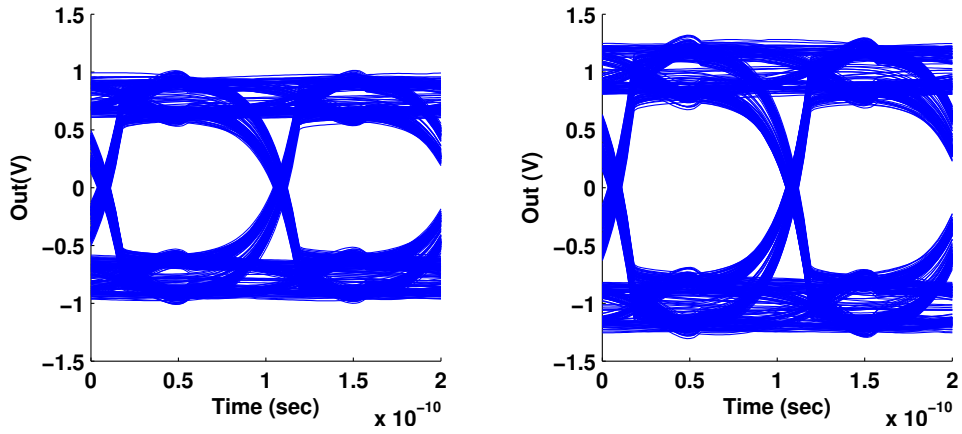

Figure 2.6: Eye diagram (left) without the pre-emphasis filter and (right) with applying a 4-taps pre-emphasis filter

(rms) value. On the other hand, the deterministic part is predictable and makes the dominant contribution to the shape of eye diagram [KHK05].

2.3.1 Subcomponents of Jitter and Amplitude Noise

Take timing jitter as an example, the total jitter (TJ) is subdivided into two categories: random jitter (RJ) and deterministic jitter (DJ). RJ is a random process and is usually assumed to have a Gaussian distribution because it is mainly contributed by thermal noise [OFK04]. In contrast, DJ is repeatable and predictable. The peak-to-peak value of deterministic jitter is bounded due to its predictable nature. Data-Dependent Jitter (DDJ), one of the most important sub-component of DJ, is dependent on the bit pattern transmitted on the link under test and is caused by duty-cycle distortion (DCD) and inter-symbol interference (ISI).

Typical crosstalk noise coupling from adjacent data-carrying links belongs to bounded uncorrelated jitter (BUJ). BUJ is bounded due to finite coupling strength, and uncorrelated because there is no correlation to the channels own data pattern. In this work, we consider the crosstalk from the adjacent differential link and, as a result, the jitter becomes part of DDJ since we exactly know

the transmitted data pattern on the adjacent link.

The PDF of data-dependent jitter and noise are always a series of pulses at the locations where a specific bit pattern experiences a cross over. Therefore, in order to get an accurate measure of the worst-case, a large number of bit patterns must be analyzed. As a result, it is critical to find out the worst-case input pattern without doing lengthy simulations. In order to efficiently find out the worst-case, we start with formula-based jitter and noise models in the following subsection.

2.3.2 Modeling of Data-Dependent Jitter and Noise

To start with, we model the CML transmitter shown in Fig. 2.5 as an independent voltage source, V_s, with matching conductance G_s. At the receiver end, G_L and C_L are used to model the loading conductance and parasitic capacitance of the CML receiver, as shown in Fig. 2.7. Therefore, the termination constraints become

$$V(0) = V_s - \frac{I(0)}{G_s}, \tag{2.25}$$

and

$$I(d) = (G_L + sC_L)V(d), \tag{2.26}$$

and we can derive the frequency domain transfer function using (2.20), (2.25) and (2.26). The result is as follows:

$$V(d) = \tilde{H}V_s(s) = (Y_{12} + (Gs + Y_{11})\tilde{Y})^{-1}Gs \cdot V_s(s), \tag{2.27}$$

where

$$\tilde{Y} = Y_{21}^{-1}(-Y_{22} - G_L - sC_L). \tag{2.28}$$

Note that G_s, G_L, and C_L are all 2×2 diagonal matrices. $\tilde{H}$ describes the complete two-port relationship and includes the effect of signal distortion, dispersion, reflection and all other channel impairments. The frequency domain relationship

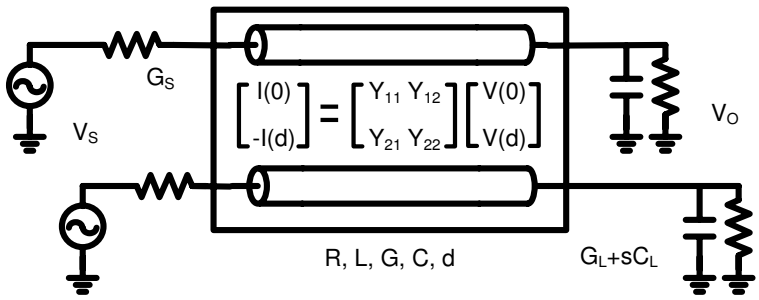

Figure 2.7: Differential signaling with parallel transmission lines and termination

between differential input and differential output now becomes

$$H(s) = \begin{bmatrix} 1 & -1 \end{bmatrix} \tilde{H} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}. \tag{2.29}$$

In order to find the time domain response, (2.29) is approximated into the following pole-residue form:

$$H(s) = c + \sum_{i=1}^{q} \frac{k_i}{s - p_i}, \tag{2.30}$$

by using a least-square-approximation-based method [BS98], where c is a constant, and p_i and k_i are the i_{th} out of q poles and residues of $H(s)$. In this way, the time domain step response $s(t)$ can be obtained through the inverse Laplace transform of $H(s)/s$ and we get

$$s(t) = c \cdot u(t) + \sum_{i=1}^{q} \frac{k_i}{p_i}(e^{p_i t} - 1)u(t), \tag{2.31}$$

where $u(t)$ is the unit step function.

The received signal $r(t)$ at the far-end of the transmission line now can be expressed as

$$r(t) = \sum_{i=-\infty}^{\infty} b_i p(t - iT), \tag{2.32}$$

where $p(t) = s(t) - s(t - T)$ is the time-domain response of a non-return-to-zero (NRZ) symbol, and b_i is the filter's output which can be evaluated as

$$b_i = \sum_{j=0}^{m-1} W_j a_{i+j}, \tag{2.33}$$

with W_j as the pre-emphasis filter coefficient and a_i as the input symbol pattern. m is the number of taps in the filter and is the same as shown in (2.23). Note here we assume the input symbol pattern a_i is uncorrelated with each other. But any linear data correlation, such as encoded data using block coding, can be taken into consideration and the same problem formulation can be applied when we combine the correlation with the pre-emphasis filter using one sigle linear expression similar to (2.33).

We define the reference time point t_0 as the time when the waveform, without interference from neighboring symbols, crosses a certain threshold V_{th} [BAH04, BA04]. In other words, t_0 can be solved with

$$p(t_0) = V_{th}, \quad 0 \leq t_0 < T. \tag{2.34}$$

Jitter is the deviation from such a time point. For a given input pattern, the jitter can be computed as

$$|t_1 - t_0|, \tag{2.35}$$

where $r(t_1) = V_{th}$. On the other hand, the amplitude noise is defined as the amplitude variation at the optimal sampling time, that is

$$|r(t_s) - p(t_s)|, \tag{2.36}$$

where

$$t_s = \arg\max_t \{p(t)\}. \tag{2.37}$$

2.4 Worst-case Timing Jitter

The data-dependent jitter and amplitude noise highly depend on the input pattern. In this section, we develop algorithms that, by using mathematical programming, can directly find out the input patterns for worst-case jitter and worst-case noise without doing lengthy simulations. To start with, the algorithm for worst-case jitter is proposed in this section.

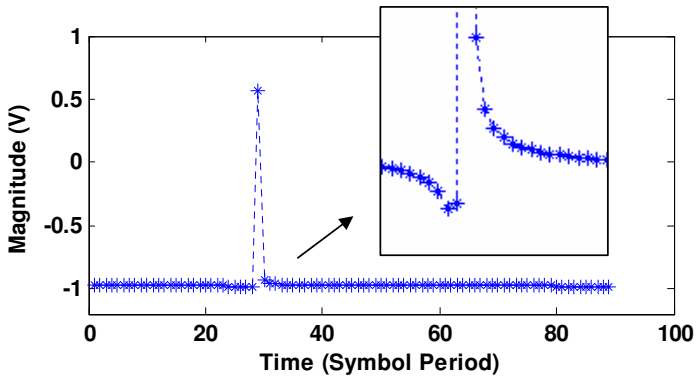

Figure 2.8: Time domain response of the channel.

2.4.1 Problem Formulation

The worst jitter is the sum of the maximal positive deviation $t_1 - t_0$ ($t_1 > t_0$) and the maximal negative deviation $t_0 - t_1$ ($t_0 > t_1$). For simplicity of presentation, we only discuss how to compute the maximal positive deviation. It should be understood that the same procedure can be applied to compute the maximal negative deviation as well. We can formulate the maximal positive deviation as the following integer non-convex programming problem (**P1**)

$$
\begin{aligned}
(\mathbf{P1}) \quad \max_{a_i} \quad & t_1 - t_0 && (2.38)\\
s.t. \quad & \sum_{i=-\infty}^{\infty} b_i r(t_1 - iT) = V_{th} && (2.39)\\
& t_0 \le t_1 < T && (2.40)\\
& b_i = \sum_{j=0}^{m-1} W_j a_{i+j} && (2.41)\\
& a_i \in \{0, 1\}, && (2.42)
\end{aligned}
$$

where t_0 is defined in (2.34) given zero-crossing threshold V_{th}. $r(t)$ is the time-domain response for one-bit transmission as shown in Fig. 2.8. a_i and b_i are pre-emphasis filter's input and output with W_j is the filter's coefficient for tap j.

Note that $r(t)$ attenuates quickly as time goes to infinity, as shown in Fig. 2.8.

Thus (2.39) can be well approximated by

$$\sum_{i=-N}^{N} b_i r(t_1 - iT) = V_{th}, \tag{2.43}$$

where N can be decided such that the error is within certain bound

$$|b_N r(t - NT)| < |\epsilon r(t)|, \forall 0 \leq t < T \tag{2.44}$$

and ϵ is in $[0, 1]$ and is specified by user. A larger ϵ reduces the problem complexity, but introduces more significant error.

2.4.2 Relaxation Based Binary Search

If we assign a set of values to t_1, then the problem becomes a non-linear feasibility problem and can be solved through an efficient heuristic method, i.e., for each value of t_1, we test whether a combination of the symbols a_i can be found such that (2.39) holds, and then pick the t_1 that maximizes $t_1 - t_0$ among all the feasible solutions. Such a problem structure enables us to use the binary search technique on t_1, which is bounded in $[t_0, T)$. However, the main difficulty lies in the fact that the feasible space for t_1 is not continuous. If we randomly assign values to t_1, the chance for it to be feasible is slim.

To overcome this difficulty, instead of finding a set of symbols that satisfies (2.39), we look for a nearby feasible value as an alternative, if possible. This is done by the following procedure. Suppose t_1 is assigned with value $\tilde{t}_1$. Then the corresponding feasibility problem would be

$$\sum_{j=-N}^{N} b_j r(\tilde{t}_1 - jT) = V_{th} \tag{2.45}$$

$$0 \leq \tilde{t}_1 < T \tag{2.46}$$

$$b_j = \sum_{i=0}^{m-1} W_i a_{i+j} \quad -N \leq j \leq N \tag{2.47}$$

$$a_j \in \{0, 1\} \quad -N \leq j \leq N + m - 1, \tag{2.48}$$

2.4.3 Pseudo-Boolean Optimization

We first show how to re-formulate the problem so that we can use a general pseudo-Boolean solver as the core of our binary search algorithm.

Once t_1 is assigned with value $\tilde{t}_1$ within binary search, the value of $r(\tilde{t}_1 - jT)$ in (2.45) can be calculated easily and become a known variable to us, denoted as r_j. Along with (2.47), as a result, the constraints in the original feasibility problem now becomes:

$$\sum_{j=-N}^{N+m-1} c_j a_j = d \tag{2.49}$$

$$a_j \in \{0,1\} \quad -N \leq j \leq N+m-1, \tag{2.50}$$

where $c_j = \sum_{k=-N}^{j} r_k W_{j-k}$ and $d = V_{th}$ are some constants that can be derived from (2.45) and (2.47), given the value of $\tilde{t}_1$. The resulted problem becomes that can we find at least one feasible solution, which each a_j is restricted to either 0 or 1, satisfies a linear equality constraint.

To solve this problem, we can simply relax the equality constraint by adding an error term ϵ and convert the equality constraint into two inequality constraints, as shown in problem (P2): (**P2**)

$$(\mathbf{P2}) \quad \min \quad t \tag{2.51}$$

$$s.t. \quad \sum_{j=-N}^{N} c_j a_j \leq d + \epsilon t \tag{2.52}$$

$$\sum_{j=-N}^{N} (-c_j) a_j \leq -d + \epsilon t \tag{2.53}$$

$$a_j \in \{0,1\} \quad t \in \{0,1\} \tag{2.54}$$

$$-N \leq j \leq N+m-1. \tag{2.55}$$

Note that an extra variable t is added to convert the feasibility problem into an optimization problem.

(P2) is actually an instance of the pseudo-Boolean optimization (PBO) problem, which generally has the following structure:

$$\min \quad \sum_{j=1}^{n} c_j \cdot x_j \tag{2.56}$$

$$s.t. \quad \sum_{j=1}^{n} b_{ij} l_j \leq d_i \tag{2.57}$$

$$x_j \in \{0,1\}, \quad i \in \{1 \cdots m\}, \tag{2.58}$$

where x_j is a Boolean variable and a literal l_j denotes either a variable x_j or its complement $\bar{x}_j$. The PBO can be solved by using existing pseudo-Boolean solver. In this work, we use miniSAT+ [ES06], a SAT-based pseudo-Boolean solver, to solve (P2) in our experiment, given the error bound defined by ϵ. Note that if a given t_1 value is feasible with error bound ϵ, the SAT-based pseudo-Boolean solver returns a feasible input pattern $\tilde{a}_j$. Then we can use the pattern to calculate the timing jitter t_1 using (2.43) and use it as the new bound in the binary search algorithm.

Note that due to the complexity of our original problem, there is neither guarantee for optimal solution nor meaningful upper bound of runtime complexity for the combined relaxation-based binary search and pseudo-Boolean optimization. Experimental results, however, show that our algorithm gives more pessimistic result than brute-force enumeration method, yet achieve significant runtime reduction. In the following section, we further propose a heuristic algorithm replacing PBO solver to reduce runtime.

2.4.4 Proposed Heuristic

Instead of solving the feasibility problem through pseudo-Boolean solver, in this section, we propose a heuristic algorithm based on the specific structure of our problem so that we can solve it more efficiently.

To begin with, we first relax the integer constraint (2.48) to

$$0 \leq a_j \leq 1 \quad -N \leq j \leq N+m-1, \tag{2.59}$$

and solve the problem (**P3**)

$$(\mathbf{P3}) \quad \max \quad \sum_{j=-N}^{N+m-1} |a_j - 0.5| \tag{2.60}$$

$$s.t. \quad b_j = \sum_{i=0}^{m-1} W_i a_{i+j} \quad -N \leq j \leq N \tag{2.61}$$

$$\sum_{j=-N}^{N} b_j r(\tilde{t}_1 - jT) = V_{th} \tag{2.62}$$

$$0 \leq a_j \leq 1 \quad -N \leq j \leq N+m-1. \tag{2.63}$$

The objective function (2.60) tries to find the solution set a_j that is as close to the integer as possible. For the time being, let's assume that we know how to solve (**P3**). Then we denote the optimal solution as $\tilde{a}_j$ and round it to 0 or 1. After that, we can get $\tilde{b}_j$ from (2.61), and insert them in the equation

$$\sum_{j=-N}^{N} \tilde{b}_j r(t_1 - jT) = V_{th} \tag{2.64}$$

to solve for t_1, which is close to $\tilde{t}_1$ and yet is a feasible solution of the original problem. This procedure can now be used as the core for the binary search. The overall algorithm for jitter computation is shown in Algorithm 1, where ϵ_0 is used to control the termination condition: when the lower bound and upper bound have a difference smaller than ϵ_0, the search stops.

Now we discuss how problem (**P3**) can be solved efficiently. For the sake of efficiency, we propose an heuristic to obtain its solution directly from the structure of (**P3**). Let

$$x_j = a_j - 0.5 \quad -N \leq j \leq N+m-1, \tag{2.65}$$

and insert (2.61) into (2.62). Then(**P3**) can be transformed into an equivalent

form

$$\max \quad \sum_{j=-N}^{N+m-1} |x_j| \tag{2.66}$$

$$s.t. \quad \sum_{j=-N}^{N+m-1} c_j x_j = d \tag{2.67}$$

$$-0.5 \leq x_j \leq 0.5 \quad -N \leq j \leq N+m-1, \tag{2.68}$$

where c_j and d are some constants that can be derived easily.

The incentive of the heuristic to be proposed below is to let as many x_i take the maximum absolute value as possible. Due to the symmetry of the problem, without loss of generosity, we can assume

$$|c_{-N}| \leq |c_{-N+1}| ... \leq |c_k| \leq |c_{k+1}| \leq |c_{N+m-1}| \tag{2.69}$$

Then according to this ascending order of $|c_i|$, we assign -0.5 or 0.5 as the optimal value $\tilde{x}_i$ based on the following criteria

$$\tilde{x}_i = \begin{cases} -sgn(c_i) \times 0.5 & \text{if } d - \sum_{j=-N}^{i-1} c_j x_j > 0 \\ sgn(c_i) \times 0.5 & otherwise \end{cases} \tag{2.70}$$

This assignment is continued until

$$\sum_{j=i+1}^{N+m-1} 0.5|c_j| < d - \sum_{j=-N}^{i} c_j \tilde{x}_j. \tag{2.71}$$

And the solutions for the remaining x_i are

$$\tilde{x}_i = sgn(c_i) \times \frac{\sum_{j=-N}^{i} c_j \tilde{x}_j - d}{|c_i|}. \tag{2.72}$$

Again, due to the complexity of the original problem, we cannot guarantee that the solution obtained from our algorithm is optimal (or even locally-optimal). However, experimental results show that our algorithm gives a result that is very close or more pessimistic to the enumeration method, yet achieves further speedup compared to using PBO solver. The overall algorithm for solving problem (**P1**) is summarized in Algorithm 1.

Algorithm 1 Algorithm for solving problem (**P1**).

Initialize: $t_1^{lb} = t_0$; $t_1^{ub} = T$;
while $t_1^{lb} < t_1^{ub} - \epsilon_0$ **do**
  $\tilde{t}_1 = (t_1^{lb} + t_1^{ub})/2$;
  Solve problem (**P3**) for $\tilde{a}_i$ and round it to 0 or 1.
  Compute $\tilde{b}_i$ based on the rounded $\tilde{a}_i$ from (2.61);
  Solve (2.64) for t_1;
  if $t_1 > t_1^{lb}$ **then**
    $t_1^{lb} = t_1$;
  else
    $t_1^{ub} = \tilde{t}_1$;
Return t_1^{lb};

2.5 Worst-case Amplitude Noise

The amplitude noise is the difference between the maximum amplitude deviation and the minimum amplitude deviation, at the optimal sampling time. To find the worst-case noise, we could use the following formulation:

$$(\mathbf{P4}) \quad \max_{a_i} \text{ or } \min_{a_i} \quad \sum_{i=-N}^{N} b_i r(t_s - iT) \tag{2.73}$$

$$s.t. \quad b_i = \sum_{j=0}^{m-1} W_j a_{i+j} \tag{2.74}$$

$$a_i \in \{0, 1\}, \tag{2.75}$$

where

$$t_s = \arg\max_t \{p(t)\} \tag{2.76}$$

is the optimal sampling time. The difference between maximum and minimum deviation determines the peak-to-peak amplitude noise for the eye diagram. Given the t_s calculated from (2.76), we can rewrite (**P4**) as (use the maximum problem as an example)

$$\max_{a_i} \quad \sum_{i=-N}^{N+m-1} c_i a_i \tag{2.77}$$

$$s.t. \quad a_i \in \{0, 1\} \tag{2.78}$$

where $c_i = \sum_{k=-N}^{i} r_k W_{i-k}$ and $r_k = r(t_s - kT)$ can be derived from (2.73) and (2.74). As a result, it is a linear programming problem and, moreover, the solution can be obtained directly without calling the general linear programming solver. Obviously, to maximize the objective function, we just let a_i be 1 if c_i is positive and be 0 if c_i is negative [HWM05]. For the minimum case, it is vice versa. So the amplitude noise can be expressed as

$$\sum_{i=-N}^{N+m-1} |c_i|, \tag{2.79}$$

and the complexity is $O(N^2)$ given $m \ll N$.

2.6 Experimental Results

In this section, we report our experiments on a Pentium 4 computer with 2.66G CPU and 1G RAM.

2.6.1 Jitter and Amplitude Noise Model Validation

We first verify our transmission line channel model. Table 2.1 lists the detailed design information for our various testbench. And Fig. 2.9 shows the comparison of the transient simulation result between our analytical channel model and SPICE simulation for Design 1, given the same RLGC model for channel and the same pre-emphasis filter coefficients. The RLGC value for the channel can be calculated by giving the geometry parameters and using the methods as discussed in Section II.A. For example, the resulted RLGC values for Design 1 are $r = 17.24\Omega$, $l = 325nH$, $m = 119nH$, $c = 135pF$, and $c_p = 49pF$. From Fig. 2.9, we can find out that the transient behavior is pretty similar and both give the same amount of timing jitter and amplitude noise. Note that the origin point is different between SPICE and our model. This is due to different input setting and does not affect the noise and jitter measurement.

Table 2.1: Differential transmission line testbench design information: width(w), spacing(s), thickness(t), dielectric height(h), length(L) and characteristic impedance.

Design	w (um)	s (um)	t (um)	h (um)	L (cm)	Char. impedance
#1	100	193.86	10	300	15	49.03
#2	50	117.48	50	200	15	49.2
#3	50	117.48	50	200	25	49.2
#4	100	80	10	300	15	52.51
#5	50	500	10	300	30	58.55

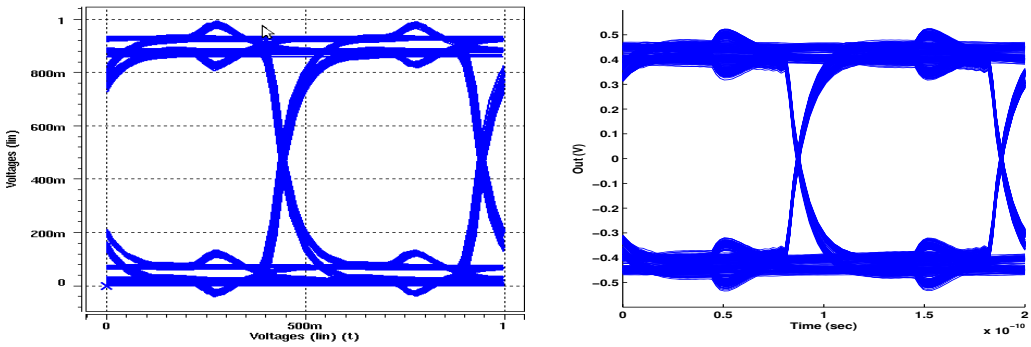

Figure 2.9: Transient simulation comparison between (left) SPICE and (right) our model. The origin point is different.

Next, we verify our jitter and noise model with SPICE given the same set of input patterns in Table 2.2. The test pattern contains 100 symbol with a data rate at 10Gb/s. From Table 2.2, we discover that, given the same input pattern, our model can accurately calculate jitter and noise with similar runtime, compared to SPICE results. The error is within 4.5% for timing jitter and 5% for amplitude noise. Although the runtime improvement is not much, our model is easier to be embedded into other tools or algorithms.

To emphasize the importance of considering a long period of time domain response, Fig. 2.10 shows the time domain response for Design 5, but with unmatched termination resistance. The impedance mismatch at the receiver end

Table 2.2: Jitter and amplitude noise model validation.

	SPICE			Our Model		
	Jitter (ps)	Noise (V)	Runtime (sec)	Jitter (ps)	Noise (V)	Runtime (sec)
#1	11.8	0.27	0.26	11.9	0.27	0.17
#2	5.0	0.20	0.26	5.0	0.21	0.17
#3	4.6	0.20	0.25	4.7	0.19	0.16
#4	11.0	0.29	0.26	10.5	0.30	0.17
#5	7.7	0.11	0.23	7.7	0.11	0.18
Avg. relative error				1.1%	2.8%	

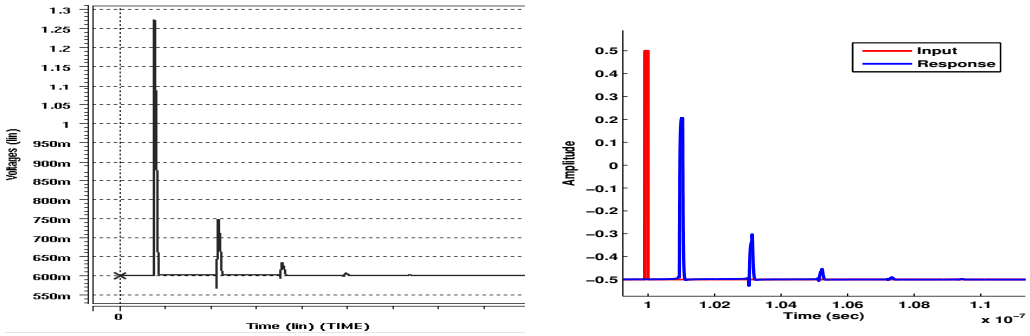

Figure 2.10: Time domain response: SPICE simulation(left) and MATLAB simulation with our model(right). The origin point is different.

will cause severe signal reflection. From Fig. 2.10, both SPICE simulation and our model clearly illustrate the signal reflection behavior. As a result, only a few taps of time domain response is not sufficient to determine the jitter and noise performance.

2.6.2 Worst-case Jitter and Amplitude Noise Calculation

The worst-case jitter, amplitude noise, and runtime comparison for various design cases are listed in Table 2.3. The pre-emphasis filter is optimized in advance for different channel characteristics. The jitter and amplitude performance is calculated through our formula-based model and we consider 40 taps of transmission line time domain response. We first test 10000 sets of pseudo-random bit se-

Table 2.3: Worst-case jitter, amplitude noise and runtime comparison: PRBS denotes pseudo-random bit sequence. For worst-case jitter, relaxation-based binary search (BS) is used along with pseudo-Boolean optimizer (PBO) or our heuristic algorithm (Heuristic). For worst-case noise, LP denotes directly solving linear programming.

	Worst-case Jitter (ps)			Worst-case Noise (volts)		Total runtime (sec)		
	PRBS	BS+PBO	BS+Heuristic	PRBS	LP	PRBS	BS+PBO+LP	BS+Heuristic+LP
#1	16	15	16	0.34	0.36	37.8	1.56	0.25
#2	8	8	8	0.25	0.27	38.9	1.55	0.25
#3	9	10	9	0.25	0.30	38.2	1.56	0.26
#4	20	25	24	0.37	0.41	37.8	1.50	0.26
#5	12	12	12	0.14	0.17	37.9	1.55	0.26

quences (PRBS) in order to find the worst-case scenario for both jitter and noise. We then use relaxation-based binary search to directly determine the required input pattern for the worst-case jitter and solve the linear programming (LP) problem for the worst-case amplitude noise. Moreover, within the relaxation-based binary search for worst-case jitter, two different methods are compared: pseudo-Boolean optimization (PBO) using miniSAT+ [ES06] and our proposed heuristic algorithm. Table 2.3 shows that, for all the cases, our algorithm obtains more reliable worst-case jitter and noise compared to PRBS simulations. For worst-case jitter, the results are similar no matter we use PBO solver or the proposed heuristic algorithm but the heuristic algorithm provides better runtime. From Table 2.3, it shows our algorithm obtains worst-case jitter and noise by up to 20% bigger than PRBS. At the same time, our algorithm is 25× faster than PRBS when we use binary search and PBO solver for worst-case jitter and solve LP for worst-case noise. And 150× faster than PRBS when we replace the PBO solver with our proposed heuristic algorithm.

2.7 Conclusion

In this chapter, we develop efficient algorithms to calculate the worst-case data-dependent jitter and noise directly for a differential microstrip line without lengthy simulation. We first propose formula-based jitter and noise models that consider the combined effect of ISI, crosstalk and the pre-emphasis filter. With the same set of input patterns, our models achieve within 5% difference compared to SPICE simulation. By utilizing these formula-based models, we then use binary search along with pseudo-Boolean optimization and mathematical programming to directly predict the input patterns that cause worst-case jitter and worst-case amplitude noise. Experiments show our algorithms obtain more reliable worst-case jitter and noise compared to PRBS simulation and, meanwhile, achieve a 25× runtime reduction when using binary search and PBO solver for worst-case jitter and solving LP for worst-case noise. In addition, by replacing the PBO solver with our proposed heuristic algorithm, a further 150× runtime reduction compared to PRBS can be achieved. Note that our modeling and algorithms are not restricted to differential signaling and can be applied to any multiconductor transmission lines.

CHAPTER 3

Joint Design-Time and Post-Silicon Optimization for Digitally Tuned Analog Circuits

3.1 Introduction

As process technologies scale down to 90nm and below, traditional circuit design methodologies are confronted by the prominent problem of process variation. To deal with process variation for analog circuits, which are highly sensitive to device matching, traditional corner-based design is adopted to guarantee performance in the worst-case scenarios at the cost of substantial circuit overhead. Such corner-based design methodology, however, is becoming insufficient and may eventually be inviable as variation increases with technology scaling.

Statistical design is proposed to analyze the performance distribution from process variation and defines *parametric yield* as the probability the design meets a specified performance or power constraint. Different techniques exist to maximize the parametric yield for analog circuits and generally fall into two complementary categories: *design-time optimization* and *post-silicon tuning*.

Design-time optimization techniques explore the design space at system-level and device-level to maximize the yield for analog circuits. At system-level, different circuit architectures are explored for a trade-off between power, area, and performance. Moreover, some architectures such as closed-loop negative feedback

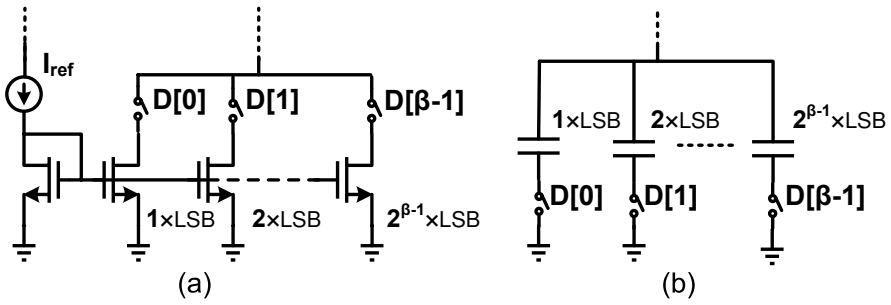

Figure 3.1: Examples of digitally tuned analog circuits: (a) CMOS current source and (b) capacitance array.

have good immunity from process variation. On the other hand, the impact of process variation can also be reduced by device-level optimization such as transistor sizing [PDW89] and layout optimization. Design-time optimization, however, has difficulty covering all process corners in a cost efficient fashion and may result in high area/power overhead.

Post-silicon tuning in analog design has been widely adopted to combat process variation. Tunable elements such as programmable capacitance array (PCA) [DKC01] and resistance array are proposed to adjust analog circuit performance after chip fabrication [HL01, MTH03]. Fig. 3.1 shows two examples of the tunable elements in analog design: tunable CMOS current source and capacitance array, where β is the resolution (number of control bits). By applying appropriate, potentially different, control signals $D[i]$ $(1 \leq i \leq \beta - 1)$ on individual chips, performance can be adjusted to maximize yield. While this will be discussed in more detail in Section II, we would like to point out that in both examples the tuning values are digitized. Such *digitally tuned analog circuits* have wide applications because of their noise-insensitivity and good technology scalability [MB04].

Post-silicon tuning has been shown to directly impact the design-time optimization for analog circuits [MB04]. On one hand, post-silicon tunability can significantly relax the analog design by providing a certain capacity to "correct"

performance deviation after fabrication. On the other hand, tuning circuitry consumes extra area and power which needs to be considered during design-time optimization in order to meet design specifications. The strong coupling between design-time optimization and post-silicon tuning has already led to joint optimization in the respective domains of both digital circuit design [MSO06] and high-level synthesis [WWX08]. It is natural to expect that by extending joint design-time and post-silicon optimization to analog design, a better parametric yield can be achieved. The complication of modeling and optimizing tunable analog circuits, however, leaves co-optimization an open problem in literature.

In this chapter, we study the joint design-time and post-silicon optimization with focus on digitally tuned analog circuits. This type of circuit has two special properties: first, variables such as the transistor sizes are continuous, while variables such as the tuning resolution are discrete in nature. Second, if the resolutions are the only changing variables and all the remaining variables are fixed, we can show that finding the performance upper bound among all permissible resolutions is easy. To make use of these two properties, we propose a general optimization framework combining the branch-and-bound algorithm on the resolutions and gradient-ascent method on the unpruned branches. We use the high-speed serial link as our application and provide two analog design examples to demonstrate the joint optimization framework: transmitter equalization filter design and phase-lock loop (PLL) design. In the transmitter design, we use the transistor sizes, number of taps, resolution, and the least significant bit (LSB) size of the pre-emphasis filter as the optimization variables and propose mathematical models of bit error rate (BER), power, and area with respect to those variables. Our experimental results show that compared with the design heuristic commonly used by analog designers, joint design-time and post-silicon optimization can improve the yield by up to 47% under the same area and power constraints. The same framework is applied to a tunable PLL as another example. We use the

charge pump currents as our design variables and formulate the problem to maximize the yield defined by output clock jitter. Result shows the jitter yield can be improved by up to 56% with power and area constraints when compared with the design heuristic. To the best of the authors' knowledge, this work is the first yield-driven analog circuit design technique that considers post-silicon tuning and design-time optimization at the same time.

3.2 Preliminaries on Digitally Tuned Analog Circuits

Analog circuits are very sensitive to process, voltage, and temperature (PVT) variations. Among all sources of variations, the random mismatches caused by doping fluctuations are expected to become dominant within the next few technology generations. In this work, we focus on the transistor threshold voltage(V_{th}) mismatch and use it as our main source of process variation. Fig. 3.2 shows an example of threshold voltage(V_{th}) variation and the resulting transistor drain current mismatch. The relation between the V_{th} variation and the resulting drain current I_D can be linearly approximated [Raz95] as

$$I_D = I_{D0} + \eta \Delta V_{th}, \tag{3.1}$$

where η and I_{D0} can be obtained through SPICE simulation, as shown in Fig. 3.2(a). Such drain current variation then causes significant power and performance variation in analog design.

To address this issue, various analog design techniques are proposed to reduce the impact of variations. In particular, post-silicon tuning is widely used to calibrate process variation after fabrication using tunable elements. Examples of tunable elements can be found in Fig. 3.1. In those tuning elements, digital binary control signal is adopted because digital signal is not sensitive to noise and as a result, makes itself immune to variation sources. Those digitally tuned analog circuits conceptually operate as a digital-to-analog conversion (DAC) circuit.

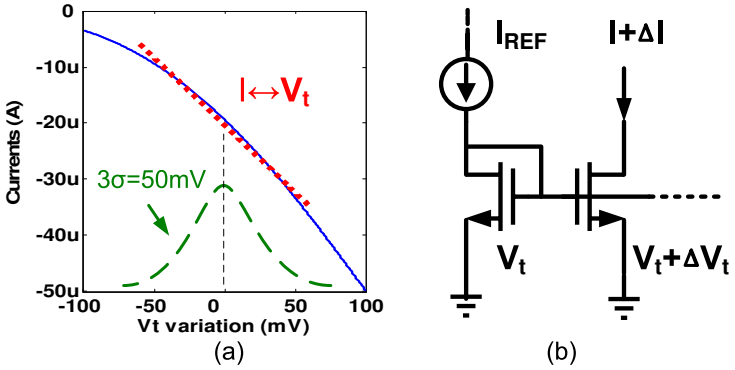

Figure 3.2: V_{th} variation model (a) and current mirror with V_{th} mismatch (b).

By given a control signal i.e. D, an analog output i.e. A, is produced proportionally. There are two major design aspects for digitally tuned analog circuits: least-significant-bit (LSB) size and resolution. The LSB size determines the minimum step in the digital-to-analog conversion. In the CMOS current source shown in Fig. 3.1(a), for example, it physically represents the drain current for the LSB transistor (I_{LSB}). In the capacitance array shown in Fig. 3.1(b), it represents the minimum size capacitance (C_{LSB}) in the array. Resolution, on the other hand, is the number of bits used as input control signal. Given the LSB size and resolution, the tuning range can be directly determined. In this chapter, we denote its resolution as β and the LSB size as γ.

An example of a digital-to-analog conversion curve is shown in Fig. 3.3. Assume that digital input D is designed to generate analog output A. With the V_{th} variation, however, the conversion curve becomes nonlinear, and input D generates output with a ΔA deviation with respect to A. To make the analog output closer to the desired value, one can change the input from D to D' and, therefore, a smaller deviation $\Delta A'$ can be obtained. In general, post-silicon tuning is performed by increasing or decreasing the input stepwise to find the minimum deviation.

By applying the tuning technique, the effect of process variation can be sig-

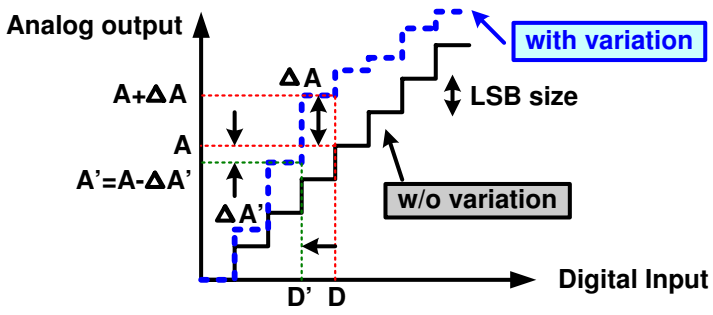

Figure 3.3: Post-silicon tuning through DAC

nificantly reduced. Extra circuits, however, are needed to provide tunability. We assume $D = [100]$ and generate $A = 4 \cdot I_{LSB}$ in Fig. 3.3. In addition to the required 4 LSB current sources, we need to implement a total of 7 LSB current sources to achieve 3-bit tunability, almost doubling the required area. Moreover, extra sources add capacitance and can potentially increase power consumption. Therefore, an optimal balance between the performance and area/power cost considering system design and post-silicon tuning must be found.

3.3 Problem Formulation

Without loss of generosity, analog design-time optimization can be described to determine the optimal design parameters to maximize the parametric yield, subject to the power and area constraints [1]. Mathematically,

$$(\mathbf{P0}) \quad \max \quad Prob(F(\boldsymbol{x}) \leq \bar{f}) \tag{3.2}$$

$$s.t. \quad Prob(P(\boldsymbol{x}) \geq \bar{p}) \leq \epsilon, \tag{3.3}$$

$$A(\boldsymbol{x}) \leq \bar{a} \tag{3.4}$$

$$\boldsymbol{x}_l \preceq \boldsymbol{x} \preceq \boldsymbol{x}_u, \quad \boldsymbol{x} \in R^k \tag{3.5}$$

[1] Note that we can also formulate the problem to minimize the power with given performance and area constraints. The joint optimization problem to be proposed can be re-formulated accordingly, and the same optimization framework still applies with little change.

where $F(\cdot)$, $P(\cdot)$, and $A(\cdot)$ represent the functions of performance metric, area, and power, respectively. $\bar{f}$, $\bar{p}$ and $\bar{a}$ are the upper bounds of the performance metric; power and area given by the design specifications; $\boldsymbol{x}$ is the vector of length k formed by the design variables with lower bound $\boldsymbol{x}_l$ and upper bound $\boldsymbol{x}_u$ given by the design specifications; k is the total number of design variables; ϵ is a small positive number indicating the tolerance for power variation over the upper bound $\bar{p}$. Note that in the above formulation we have assumed process variation has little impact on the area and the cut-off metric for the performance is an upper bound.

With post-silicon tuning, we first consider the special structure of the digitally tuned elements, as shown in Fig. 3.1. In this work, we adopt a simple but direct method based on the unit cell design technique for the tunable element. An example of a unit cell for the CMOS current source is shown in Fig. 3.4(a). Assume that we have characterized a total number of m unit cells with different transistor width/length and bias voltage under the condition that they all draw the same amount of current I_{unit}. Each unit cell α_i represents a set of transistor W/L and bias voltage V_b, where $0 \leq \alpha_i \leq m$. Any larger transistor, which draws larger current and provides larger swing at the output, can be obtained by connecting the unit cells of the same type in parallel. Such parallel connection ensures linear relationship for the parasitic capacitance and current driving capability, which is measured by output swing and delay as shown in Fig. 3.4(b). Moreover, by limiting the maximum number of connected cells, the transistor-level biasing constraints can be guaranteed to ensure all transistors work in the desired operation region. Note that similar unit cell design methodology can be extended to other digitally tuned elements, such as capacitance array.

As a result, the parametric yield can be rewritten as

$$Prob(\hat{F}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \leq \bar{f}) \tag{3.6}$$

where $\hat{F}(\cdot)$ is the performance metric after tuning, $\boldsymbol{\alpha}$ are the indices of the types of unit cell design, and $\boldsymbol{\beta}$ are vectors representing the resolution used for the digitally tuned elements. $\boldsymbol{\gamma}$ are the LSB sizes in terms of the number of unit cells used to implement the LSB of the digitally tuned element. In addition, post-silicon tuning also affects the power consumption, and (3.3) can be rewritten as

$$Prob(\hat{P}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \geq \bar{p}) \leq \epsilon, \tag{3.7}$$

where $\hat{P}(\cdot)$ is the power consumption after tuning.

Combining the above discussion, the joint design-time and post-silicon optimization can be extended from (**P0**) as

$$\begin{aligned}
(\mathbf{P1}) \qquad \max \quad & Prob(\hat{F}(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \leq \bar{f}) && (3.8)\\
s.t. \quad & Prob(\hat{P}(\boldsymbol{x}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \geq \bar{p}) \leq \epsilon && (3.9)\\
& A(\boldsymbol{x}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \leq \bar{a} && (3.10)\\
& \boldsymbol{x}_l \preceq \boldsymbol{x} \preceq \boldsymbol{x}_u, \quad \boldsymbol{x} \in R^k && (3.11)\\
& 0 \preceq \boldsymbol{\alpha} \preceq m\mathbf{1}, \quad \boldsymbol{\alpha} \in Z^n && (3.12)\\
& \mathbf{0} \preceq \boldsymbol{\beta}, \quad \boldsymbol{\beta} \in Z^n && (3.13)\\
& \mathbf{0} \preceq \boldsymbol{\gamma}, \quad \boldsymbol{\gamma} \in Z^n, && (3.14)
\end{aligned}$$

where m is the total number of unit cell designs and n is the total number of tuning elements in the circuit. Note that there is no explicit bound necessary for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ as they are implicitly bounded by the power and area constraints (3.9) and (3.10).

3.4 Optimization Framework

(**P1**) is hard to solve in general because it is a mixed integer non-convex programming problem, the complexity of which grows exponentially with the number of integer variables (the dimension of the vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$). Therefore, we pro-

pose to separate the integer variables and the continuous variables. We define a new function $Z(\boldsymbol{t})$ as the optimum value of (**P1**) when $\boldsymbol{x} = \boldsymbol{t}$. If (**P1**) is infeasible at $\boldsymbol{x} = \boldsymbol{t}$, then $Z(\boldsymbol{t}) = -\infty$. Accordingly, **P1** is equivalent to an unconstrained nonlinear optimization problem with a continuous feasible region:

$$\max Z(\boldsymbol{t}), \quad \boldsymbol{t} \in R^k, \tag{3.15}$$

which can be solved efficiently by the first order gradient method if we can evaluate $Z(\boldsymbol{t})$ and $\frac{\partial Z(\boldsymbol{t})}{\partial \boldsymbol{t}}$ at any point $\boldsymbol{t} = \hat{\boldsymbol{t}}$ to find local maximum. Below we will discuss how to evaluate the function value and first order derivative efficiently.

3.4.1 Algorithm Overview

To evaluate $Z(\boldsymbol{t})$ we need to solve problem (**P1**) for given $\boldsymbol{x} = \boldsymbol{t}$, i.e.,

$$\begin{aligned}
(\mathbf{P2}) \quad Z(\boldsymbol{t}) = \quad \max \quad & Prob(\hat{F}(\boldsymbol{t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \leq \bar{f}) && (3.16)\\
s.t. \quad & Prob(\hat{P}(\boldsymbol{t}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \geq \bar{p}) \leq \epsilon && (3.17)\\
& A(\boldsymbol{t}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) \leq \bar{a} && (3.18)\\
& 0 \preceq \boldsymbol{\alpha} \preceq m\mathbf{1}, \quad \boldsymbol{\alpha} \in Z^n && (3.19)\\
& \mathbf{0} \preceq \boldsymbol{\beta}, \quad \boldsymbol{\beta} \in Z^n && (3.20)\\
& \mathbf{0} \preceq \boldsymbol{\gamma}, \quad \boldsymbol{\gamma} \in Z^n, && (3.21)
\end{aligned}$$

with variables $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. (**P2**) is an integer programming problem, which is an NP-hard problem. Though software does exist in literature to solve general integer programming problems, in this work we propose an optimization framework to efficiently solve it using the special properties of digitally tuned analog circuits.

As delineated in Algorithm 1, the optimization framework combines the branch-and-bound (BnB) algorithm with the gradient ascent method (GDA). Assume that we know how to partition the feasibility space into different regions and how to efficiently obtain an upper bound of the objective function (3.16) for each region. Then, according to the principles of the BnB algorithm, we can prune regions that

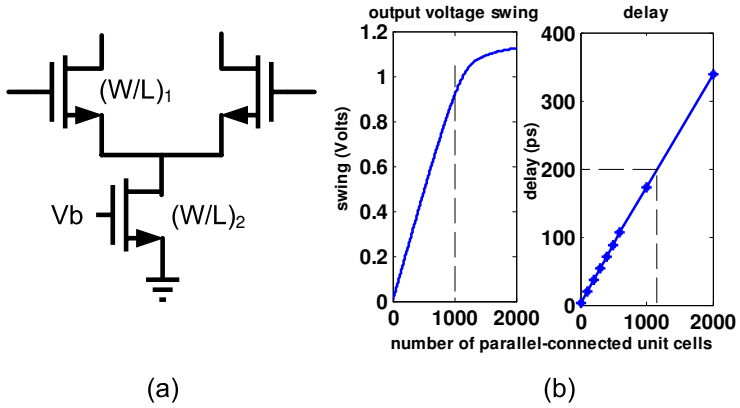

Figure 3.4: (a) Unit cell design. (b) Swing and delay vs. number of parallel-connected cells.

have an upper bound worse than the existing solutions, thereby maximizing the performance metric. If a region cannot be pruned, we employ GDA optimization to find a local maximum in it. The final solution $Z(\boldsymbol{t})$ is obtained by comparing the optimal solutions found in each unpruned region.

To evaluate the first order derivative $\frac{\partial Z(\boldsymbol{t})}{\partial \boldsymbol{t}}$, a direct method would be to use the finite different method: Compute $Z(\boldsymbol{t} + \delta \boldsymbol{e_i})$ $(1 \leq i \leq k)$ for some small positive number δ, where $\boldsymbol{e_i}$ is a unit vector with the i^{th} element equal to 1 and other elements equal to 0. Then the i^{th} element of $\frac{\partial Z(\boldsymbol{t})}{\partial \boldsymbol{t}}$ can be obtained by

$$\frac{\partial Z(\boldsymbol{t})}{\partial t_i} \approx \frac{1}{\delta}(Z(\boldsymbol{t} + \delta \boldsymbol{e_i}) - Z(\boldsymbol{t})). \tag{3.22}$$

As such, the cost for evaluating $\frac{\partial Z(\boldsymbol{t})}{\partial \boldsymbol{t}}$ would be quite expensive as we would have to solve k integer programming problems. Note that k is the total number of design variables, which can be quite large in practical problems. This urges us to turn to some alternative approach to approximate the computation with affordable cost.

As delineated in Algorithm 1, since we can obtain the upper bound of the objective function in each region efficiently, the upper bound of $Z(\boldsymbol{t})$ is just the maximum of all those upper bounds. Denoting the upper bound of $Z(\boldsymbol{t})$ as $\bar{Z}(\boldsymbol{t})$, the derivative of $Z(\boldsymbol{t})$ can be approximated by applying finite difference method

on $\bar{Z}(\boldsymbol{t})$, i.e.,

$$\frac{\partial Z(\boldsymbol{t})}{\partial t_i} \approx \frac{1}{\delta}(\bar{Z}(\boldsymbol{t} + \delta \boldsymbol{e_i}) - \bar{Z}(\boldsymbol{t})). \tag{3.23}$$

Note that the accuracy of the approximation depends on how the upper bound is calculated. If the upper bound is tight, then the approximation will converge to the exact derivatives.

Next we will discuss how to solve the two critical sub-problems: (**P3**) how to partition the feasible space and derive the upper bound of the objective function for each partitioned region and (**P4**) how to use the GDA method to find a local maximum in each region that cannot be pruned.

Algorithm 2 BnB+GDA algorithm framework for computing $Z(\boldsymbol{t})$ and $\frac{\partial Z(\boldsymbol{t})}{\partial \boldsymbol{t}}$.

Evaluate (3.16) to get $\hat{z}$ by initial guess.
(**P3**): Partition the feasible space Ω into regions ω_i $(1 \leq i \leq d)$ and derive the upper bound of the objective function $\bar{z}_i$ in each region.
$\bar{Z}(\boldsymbol{t}) = \max_i\{z_i\}$.
for $i = 1$; $i \leq d$; i++ **do**
  if $\bar{z}_i \leq \hat{z}$ **then**
    Continue;
  else
    (**P4**): Solve (**P2**) in ω_i for optimal value $\tilde{z}_i$ by the GDA method.
    if $\tilde{z}_i \geq \hat{z}$ **then**
      $\hat{z} = \tilde{z}_i$.
$Z(\boldsymbol{t}) = \hat{z}$.
Evaluate $\bar{Z}(\boldsymbol{t} + \delta \boldsymbol{e_i})$ for a small positive number δ.
$\frac{\partial Z(\boldsymbol{t})}{\partial t_i} \approx \frac{1}{\delta}(\bar{Z}(\boldsymbol{t} + \delta \boldsymbol{e_i}) - \bar{Z}(\boldsymbol{t}))$.

3.4.2 Partitioning and Bound Estimation

From Algorithm 1 we can see that the BnB+GDA framework offers a trade-off between runtime and quality: a finer partition of the solution space results in fewer local optimums in each region and accordingly, better GDA optimization quality, but at the cost of a increased runtime for the BnB algorithm as the number of total regions increases. In this work, we partition the solution space according to

the unit cell index and LSB size of each tap. In other words, each region has a unique set of unit cell indices and LSB sizes. Our experiments show that such partitioning provides a good balance between the runtime and the solution quality.

In general, the yield upper bound for a given region is hard to compute. Fortunately, in this particular type of problem, where digitally tuned analog circuits are involved, we are able to obtain the bound through a special relaxation. Suppose we can solve (**P2**) without power and area constraints, then such an optimal value can serve as the upper bound of the constrained problem (**P2**) since we have expanded the feasible space. Note that such an upper bound might not be a tight one since the corresponding solution may violate the area or power constraint.

To solve (**P2**) without constraints, we need to resort to its physical meaning: given the unit cell design and LSB sizes, find the optimal resolution that gives the maximum yield. The optimal resolution can be determined according to the target values for the tuning parameters in an iterative way, as delineated in Algorithm 2. The iterative procedure is required because in most cases the target values are also related to the resolution due to the area-dependent parasitics. In experiments, we find that the algorithm converges quickly within two or three iterations. The optimality of the solution is guaranteed because any increase in the resolution only increases the total area and the parasitics while the minimum distance to the target values remains the same, which will downgrade the performance.

3.4.3 Gradient Ascent Method

Given the partitioning method discussed in the previous section, if a particular region cannot be pruned by comparing its upper bound with the current solution, we need to solve (**P2**) for optimal $\boldsymbol{\beta}$ with given unit cell indices $\tilde{\boldsymbol{\alpha}}$ and LSB size set $\tilde{\boldsymbol{\gamma}}$.

In essence, the gradient ascent method sequentially takes steps in a direction

Algorithm 3 Yield upper bound computation for given unit cell design and LSB sizes (**P3**).

INPUT: Unit cell indices $\tilde{\boldsymbol{\alpha}}$ and LSB sizes $\tilde{\boldsymbol{\gamma}}$;
OUTPUT: Yield upper bound $\bar{z}$;
INIT: Set initial guess $\beta^{(0)}$; $k = 1$;
while $\max_k |\boldsymbol{\beta}^{(\boldsymbol{k})} - \boldsymbol{\beta}^{(\boldsymbol{k-1})}| > \epsilon ||\boldsymbol{\beta}^{(\boldsymbol{k-1})}||$ **do**
  Calculate the system parasitics according to $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}^{(\boldsymbol{k})}$;
  Update system response;
  Find the target optimal values for all tuning parameters ;
  Determine $\boldsymbol{\beta}^{(\boldsymbol{k+1})}$ according to $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\gamma}}$ and the target optimal values;
  $k = k + 1$;
$\bar{z} = Prob(\hat{F}(\boldsymbol{t}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}^{(\boldsymbol{k})}, \tilde{\boldsymbol{\gamma}}) \leq \bar{f})$;

proportional to the gradient, until a local maximum of the objective function is reached [BV04]. At each step we increase/decrease each variable by 1 and check the change of the objective function. Note that by doing so we are actually computing the gradient because all the variables are integers. We then move along the direction that causes the maximum increase. This is iteratively done until the relative change of the objective value is below a certain threshold. The termination of the algorithm indicates that one of the local maxima has been reached or that we have reached the boundary. The initial guess for the GDA can be arbitrarily chosen. In our experiments, we found that it did not influence runtime or quality significantly for both of the examples studied. In addition, we observed that the algorithm always converges to local optimum within two or three iterations.

Next we will use a high-speed serial link as our application and provide two analog design examples, the transmitter design and the phase-lock loops (PLLs) design, to demonstrate our joint optimization framework.

3.5 Transmitter Design in High-speed Serial Link

The system diagram of a high-speed serial link is shown in Fig. 3.5. At the transmitter end, the pre-driver drives the FIR pre-emphasis filter at the designated

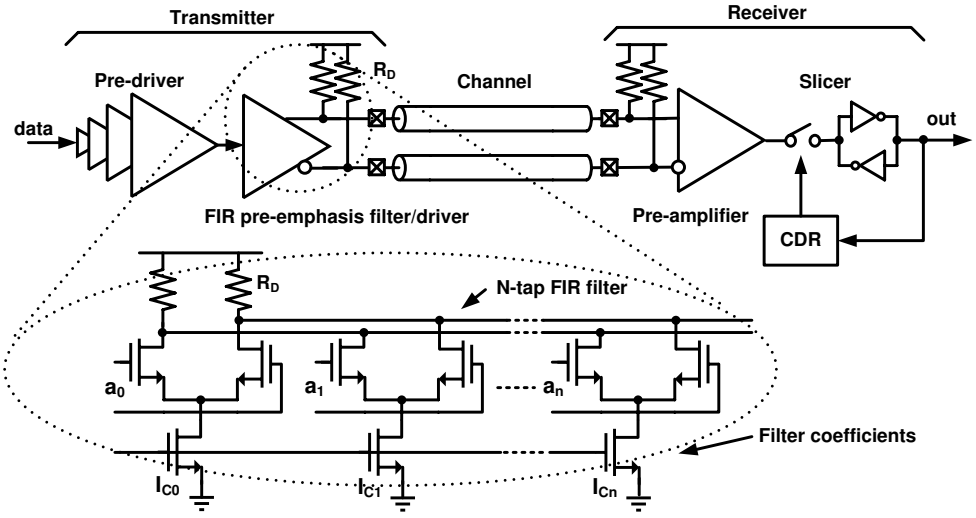

Figure 3.5: System diagram of a high-speed serial link.

data rate. The pre-emphasis filter is used to counteract the inter-symbol interference (ISI) [TBP05] caused by the bandwidth-limited channel, which behaves as a transmission line and can be characterized by the Telegrapher's equations with RLGC per-unit-length model. The pre-emphasis filter can be expressed as

$$b_i = \sum_{j=0}^{n-1} W_j a_{i-j}, \tag{3.24}$$

where n is the number of filter taps, W_i is the tap coefficient for tap i, and a_i is the transmitted non-return-to-zero (NRZ) symbol. At receiver end, the pre-amplifier, along with the slicer decision circuit, is responsible for detecting the data from the received signal. Moreover, the clock is embedded in the transmitted data and the clock data recovery (CDR) sub-system is used to extract the clock from the serial data stream.

In order to focus on the transmitter optimization, in our first example, we assume that the frequency domain response for the channel and the receiver is given. In addition, we assume that an ideal sampling clock is obtained through the receiver CDR circuits.

3.5.1 Design-time Optimization

The performance of the overall system is usually quantified in terms of BER, the rate at which errors occur during data transmission. To start with, we formulate the design-time optimization problem to minimize the BER of the system subject to power and area constraints. The design variables include the number of taps n of the filter, the transistors sizing W/L, and the bias voltage V_b in the CMOS current source. Assume that we have characterized a total number of m unit cells and each unit cell α_i represents a set of transistor W/L and bias voltage V_b, as shown in Fig. 3.4.

Since directly measuring the BER requires a long period of time, error vector magnitude (EVM) is used in this work to estimate the BER because of their monotonic relationship [SNS08].

$$EVM = \sqrt{\frac{1}{M}\frac{\sum_1^M |r_i - a_i|^2}{|r_{max}|^2}}, \tag{3.25}$$

where

$$r_i = \sum_{j=-\infty}^{\infty} b_j p_{i-j} + n_i, \tag{3.26}$$

is the received data with respect to filter output b_i from (3.24), time domain symbol response p_i, and circuit thermal noise n_i. Moreover, r_{max} is the outermost received data in the constellation and M (usually less than 10^4) is the total number of data used for computation. We can easily map the EVM to the BER from table look-up and accordingly, the objective function (3.2) takes the form

$$Prob(BER(n, \boldsymbol{\alpha}) \leq \bar{f}). \tag{3.27}$$

The area $A(n, \boldsymbol{\alpha})$ and power $P(n, \boldsymbol{\alpha})$ of the transmitter are mainly contributed by the pre-emphasis filter and the pre-driver, i.e.

$$A(n, \boldsymbol{\alpha}) = A_{pre-driver}(n, \boldsymbol{\alpha}) + A_{filter}(n, \boldsymbol{\alpha}), \tag{3.28}$$

$$P(n, \boldsymbol{\alpha}) = P_{pre-driver}(n, \boldsymbol{\alpha}) + P_{filter}(n, \boldsymbol{\alpha}). \tag{3.29}$$

For tap i ($1 \leq i \leq n$), we use unit cells of type α_i ($1 \leq \alpha_i \leq m$) with the parasitic capacitance $C_{unit}^{\alpha_i}$ and the occupied area $A_{unit}^{\alpha_i}$. The required number of cells q_i for that tap is determined by its coefficient W_i and the unit current I_{unit}:

$$q_i = \lceil \frac{W_i}{I_{unit}} \rceil. \tag{3.30}$$

Accordingly, the total area used in the pre-emphasis filter can be calculated as

$$A_{filter}(n, \boldsymbol{\alpha}) = \sum_{i=1}^{n} q_i A_{unit}^{\alpha_i}. \tag{3.31}$$

The total parasitic capacitance C_{para} can be calculated as

$$C_{para}(n, \boldsymbol{\alpha}) = \sum_{i=1}^{n} q_i C_{unit}^{\alpha_i}. \tag{3.32}$$

The power consumed by the filter (P_{filter}) contains both static power and dynamic switching power and can be expressed as

$$P_{filter}(n, \boldsymbol{\alpha}) = \rho \sum_{i=1}^{n} q_i \cdot I_{unit} \cdot V_{dd} + (1-\rho) f \cdot V_{dd}^2 \cdot C_{para}, \tag{3.33}$$

where f is the data rate. ρ is the ratio between static power and total power, which depends on detailed delay and switching probability and can be obtained from simulation.

The pre-driver is designed according to the total gate capacitance at the filter input $C_{gate} = \sum_{i=1}^{n} q_i C_g^{\alpha_i}$, where $C_g^{\alpha_i}$ is the input transistor gate capacitance of unit cell α_i. We assume the pre-driver is designed through logic effort using a simple inverter chain. Note that other configurations like CML pre-drivers with swing control can also be applied. As a result, the occupied area can be determined by

$$A_{pre-driver} = A_{inv} \cdot (1 + f_p + \cdots {f_p}^{N_p}), \tag{3.34}$$

where $N_p = ln\lfloor \frac{C_{gate}}{C_{inv}} \rfloor$ and $f_p = (\frac{C_{gate}}{C_{inv}})^{\frac{1}{N_p}}$. A_{inv} and C_{inv} are the area and input capacitance for a unit inverter and N_p is the number of pre-driver stages. The pre-driver consumes only dynamic power:

$$P_{pre-driver} = \frac{1}{2} f \cdot vdd^2 \cdot C_{inv} \cdot (1 + f_p + \cdots {f_p}^{N_p}). \tag{3.35}$$

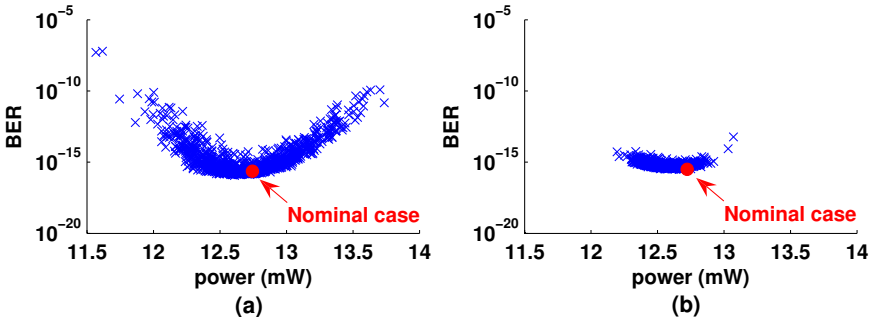

Figure 3.6: Power and performance variation for 1000 die samples by Monte Carlo simulation: (a) without tuning and (b) with tuning.

Combining (3.27)-(3.35), the optimization problem can then be mathematically formulated as shown in (**P0**).

3.5.2 Post-silicon Tuning and Joint Optimization

In the presence of process variation, assuming transistor threshold voltage V_{th} has a normal distribution with 10% variation [YLN08], the power consumed by the transmitter varies by 30% variation and the BER varies in the magnitude of $10^8\times$ for the same design, as demonstrated in Fig. 3.6(a). By applying the tuning technique, simulation results show that the span of power and BER variation becomes much smaller as shown in Fig. 3.6(b). Extra circuits, however, are needed to provide this tunability and an optimal balance between the performance and area/power cost has to be found.

To cast the problem into the format of (**P1**), we need to find $\hat{F}(\cdot)$, $\hat{A}(\cdot)$ and $\hat{P}(\cdot)$. The $\hat{F}(\cdot)$ is straightforward to obtain:

$$\hat{F} = BER(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}), \tag{3.36}$$

where $\boldsymbol{\alpha}$ is the vector indicating the LSB design for each tap. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors in R^n containing resolution and LSB size for each tap, and $\bar{e}$ is the allowed BER upper bound. Note that the number of taps n is no longer a variable: by allowing

$\beta_i = 0$, tap i is removed. Accordingly, we only need to specify n_{max}, a maximum number of taps to be considered ($n_{max} = 10$ in this work).

The power P_{filter} (3.33) and area A_{filter} (3.31) of the pre-emphasis filter also need to be modified with the introduction of the DAC:

$$P_{filter} = \rho \sum_{i=1}^{n_{max}} \boldsymbol{D_i}^T [2^{\beta_i - 1}, \cdots, 2^0] \cdot \gamma_i I_{unit} \cdot V_{dd} + (1-\rho) f \cdot V_{dd}^2 \cdot C_{para}, \tag{3.37}$$

$$A_{filter}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} 2^{\beta_i} \gamma_i A_{unit}^{\alpha_i}, \tag{3.38}$$

$$C_{para} = \sum_{i=1}^{n_{max}} 2^{\beta_i} \gamma_i C_{unit}^{\alpha_i}. \tag{3.39}$$

Note that vector $\boldsymbol{D_i}$ represents the digital control bits and P_{filter} becomes a distribution instead of a deterministic value because of the I_{unit} variation from V_{th} mismatch. The other calculations are kept the same and the total area and power can be obtained by (3.28) and (3.29), accordingly.

3.6 PLL Design in High-speed Serial Link

Phase locked-loops (PLLs) are widely used to generate well-timed on-chip clocks in high-speed transceivers [Raz96]. Any timing jitter or phase noise significantly degrades the performance of the system, especially as operating frequency increases.

Timing jitter can be expressed as $\sigma_{\Delta T} = (T/2\pi) \cdot \sigma_{\Delta\phi}$, where ω_0 is the clock frequency, $T = 2\pi/\omega_0$ is the clock period, and $\sigma_{\Delta\phi}$ is the phase jitter of the clock. Phase jitter is defined as the standard deviation of the phase difference between the first cycle and mth cycle of the clock [MK02].

An example of second-order PLL as shown in Fig. 3.7 comprises of several components: (1) the phase frequency detector, (2) the charge pump, (3) the loop filter, and (4) the voltage-controlled oscillator. Phase and frequency detector is used to detect phase and frequency difference and provides the UP/DN signal to

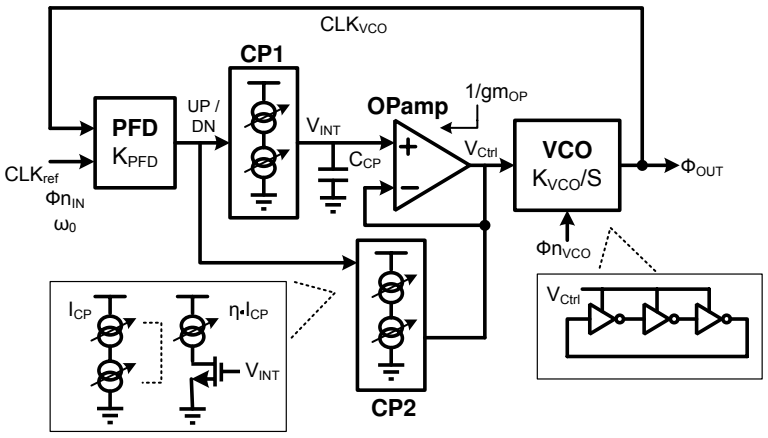

Figure 3.7: Tunable and adaptive bandwidth PLL. [SLK00]

the charge pump. The charge-pump circuit comprises of two switches driven by the UP and DN signal and injects the charge into or out of the loop filter capacitor (C_{CP}). The combination of charge-pump and C_{CP} is an integrator that generates the average voltage of UP (or DN) signal, V_{Ctrl}, and adjusts the frequency of the subsequent oscillator circuit. In Fig. 3.7, a power-supply regulated ring oscillator is shown with the voltage-to-frequency gain K_{VCO}. The VCO output frequency is controlled by its supply voltage V_{Ctrl}.

3.6.1 Design-time Optimization

The performance of PLL is measured by its output clock jitter. The jitter mainly comes from the reference clock (N_{in}) and VCO (N_{VCO}), which can be expressed as [MK02]:

$$Jitter = \sigma^2_{\Delta T} = \frac{8}{w_0^2}\int_0^\infty S_\phi(f) sin^2(\pi f \Delta T) df, \tag{3.40}$$

and

$$S_\phi(f) = \frac{N_{in}}{f^2} \cdot |Hn_{in}(j2\pi f)|^2 + \frac{N_{VCO}}{f^2} \cdot |Hn_{VCO}(j2\pi f)|^2. \tag{3.41}$$

Note that Hn_{in} and Hn_{VCO} are the noise transfer functions of the reference clock noise (N_{in}) and VCO noise (N_{VCO}) accordingly. Here we assume white noise sources and ignore the noise from the clock buffers.

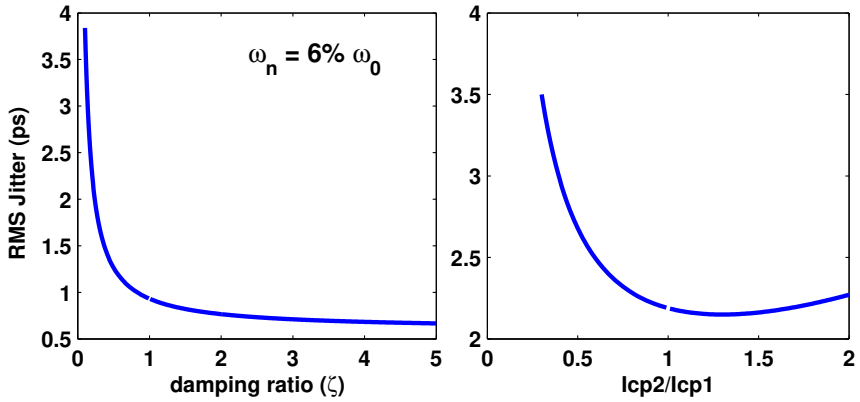

Figure 3.8: Output jitter sensitivity to the (a) loop damping factor ζ and (b) charge pump current ratio I_{CP2}/I_{CP1}.

Considering the PLL shown in Fig. 3.7, the noise transfer functions Hn_{in} and Hn_{VCO} can be expressed using PLL design parameters:

$$\begin{aligned} Hn_{in}(s) = \frac{\phi_{out}}{\phi n_{in}} &= \frac{K_{loop}RC_{CP}s + K_{loop}}{s^2 + K_{loop}RC_{CP}s + K_{loop}} \\ &= \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \end{aligned} \tag{3.42}$$

$$\begin{aligned} Hn_{VCO}(s) = \frac{\phi_{out}}{\phi n_{VCO}} &= \frac{s^2}{s^2 + K_{loop}RC_{CP}s + K_{loop}} \\ &= \frac{s^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \end{aligned} \tag{3.43}$$

where $K_{loop} = I_{CP1}/(2\pi C_{CP})K_{PD}K_{VCO}$, $\omega_n = \sqrt{K_{loop}}$, $\zeta = \sqrt{K_{loop}}RC/2$, $R = (I_{CP1}/I_{CP2})(1/gm_{OP})$ [MK02]. We can see that the noise from the input reference clock and VCO are filtered through low-pass and high-pass filters, respectively.

As a result, the jitter performance is a function of the PLL design parameters ω_n and ζ. Fig. 3.8(a) shows an example of output root-mean-square (RMS) jitter with respect to the damping ratio (ζ) for a fixed $\omega_n = 0.06\omega_0$. Moreover, in the case of tunable PLL shown in Fig. 3.7, the natural frequency varies proportionally to $\sqrt{I_{CP1}}$ and the damping factor is proportional to $I_{CP2}/\sqrt{I_{CP1}}$ [SLK00]. By finding an optimum value of the absolute value and relative ratio of I_{CP2} and

I_{CP1}, we can minimize the PLL output jitter. We write the objective function of the design-time optimization as (3.2):

$$Prob(Jitter(\boldsymbol{\alpha}) \leq \bar{f}), \tag{3.44}$$

where $\boldsymbol{\alpha}$ is a vector which represents the number of unit cells used in the charge pumps. In other words, it represents the value of I_{CP1} and I_{CP2}. An example of the relation between output RMS jitter and the current ratio for the charge pumps (I_{CP2}/I_{CP1}) is shown in Fig. 3.8(b), with a fixed I_{CP1}.

For the design-time optimization, we want to minimize the output clock jitter, subject to power and area constraints. The design parameters are the charge pump currents I_{CP1} and I_{CP2}. The power consumption of the charge pump can be calculated by an approach similar to the one used in our first transmitter design example. Assume we use unit cells of type α_i ($1 \leq \alpha_i \leq m$) with unit current I_{α_i}, then the required number of cells q_i for the charge pump i can be determined by

$$q_i = \lceil \frac{I_{CPi}}{I_{\alpha_i}} \rceil. \tag{3.45}$$

As a result, the power consumed by the charge pump is:

$$P_{CP}(\boldsymbol{\alpha}) = \sum_i \frac{1}{2}(2\pi\omega_0) \cdot (1 + \frac{1}{\eta_i}) \cdot q_i I_{\alpha_i} \cdot V_{dd}, \tag{3.46}$$

where η_i represents the current mirror ratio for the biasing circuit of the charge pump i. The area can also be approximated using the similar method and details can be found in our first example. As a result, the optimization problem is mathematically formulated as shown in (**P0**). Note that the power dissipated by PLLs is often a small fraction of total active power. However, it can be quiet significant during sleep modes where the PLL must remain locked.

3.6.2 Post-silicon Tuning and Joint Optimization

In the presence of process variation, the output RMS jitter varies for the same design because of the variations on I_{CP1} and I_{CP2}, as demonstrated in Fig. 3.9(a).

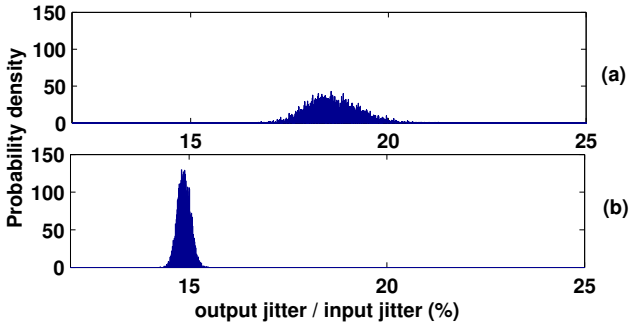

Figure 3.9: Probability density for output jitter(%). (a) without tuning (b) with tuning circuit and optimized digital control.

To reduce the impact of process variation and improve the parametric yield, post-silicon tuning techniques can be applied. Fig. 3.10 shows a schematic of the charge pump circuit with digitally tuned elements placed in the biasing circuit. By applying a proper digital control signal $\boldsymbol{D}$, the charge pump current ratio can be optimized to reduce the output jitter under the impact of process variation. The resulting histogram can be found in Fig. 3.9(b).

As discussed in Section III, we can change the objective function to the *Jitter parametric yield* as

$$Prob(Jitter(\boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \leq \bar{f}), \tag{3.47}$$

where $\boldsymbol{\alpha}$ is the vector indicating the LSB design for each tap in the tunable element. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ contain resolution and LSB size for each charge pump and $\boldsymbol{\eta}$ represents the biasing current ratio; $\bar{f}$ is the allowed jitter upper bound. The power consumed by the charge pump can be re-written as

$$P_{CP} = \sum_i \frac{1}{2}(2\pi\omega_0) \cdot (1 + \frac{1}{\eta_i}) \cdot \boldsymbol{D}_i{}^T[2^{\beta_i - 1}, \cdots, 2^0] \cdot \gamma_i I_{\alpha_i} \cdot V_{dd}.$$

Note that in this example, the tunable element is inserted in the biasing part with bias ratio η, which is considered as part of the design parameters $\boldsymbol{x}$ in (**P1**). When $\eta \ll 1$, only a small amount of current in the biasing circuit is required

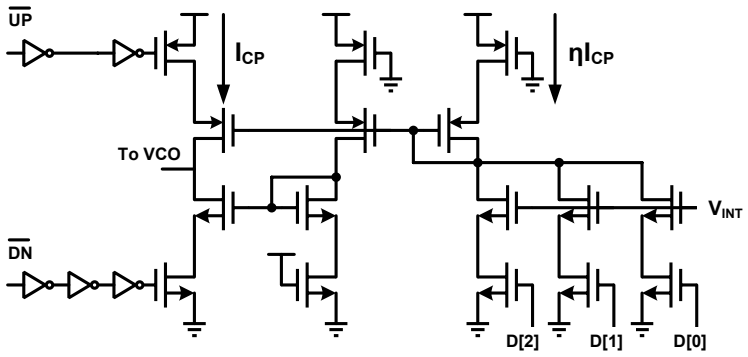

Figure 3.10: Charge pump schematic [SLK00].

to generate I_{CP}. As a result, the power consumed and the area occupied by the digitally tuned element can be ignored. In this case, however, the LSB size in the charge pump current becomes $\frac{1}{\eta}\gamma I_{\alpha}$, which is increased when η is decreased. The effect of tuning is reduced and may not provide the desired yield. On the other hand, when $\eta \sim 1$, the tunability is maximized but the power and area consumed by the tunable element is also increased. Obviously, a good balance needs to be found through our proposed framework.

3.7 Experimental Results

We extract the model parameters by SPICE simulation in IBM 90nm technology and implement the proposed algorithm in MATLAB. All the experiments are run on a Windows server with Pentium IV 3.2GHz CPU and 2G RAM.

3.7.1 Transmitter Design

We compare our algorithm with three different methods: no-tunability design, maximum tunability design, and design heuristic from designer's perspective. The design heuristic is guided by the designers' experience [VMA03, LL08]: (1) total number of filter taps is iteratively determined by the channel response and the LMS algorithm. (2) assume that each tap of the filter has the same LSB size;

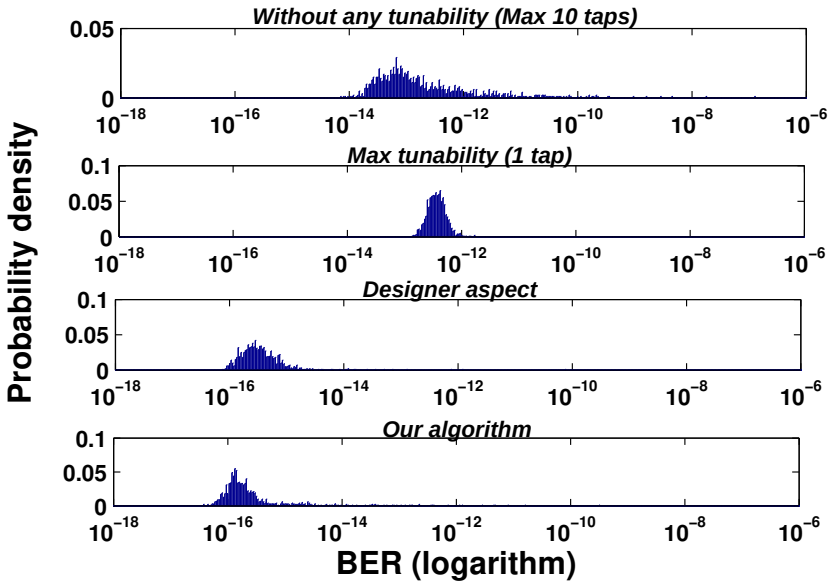

Figure 3.11: BER distribution for four different designs.

(3) the LSB size is determined by considering the maximum and minimum filter coefficient. This design methodology serves as a heuristic for this joint optimization problem and essentially solves the problem in a reduced solution space. The no-tunability design sets the resolution to be 1 ($\beta_i = 1$) for all taps and maximizes the precision of a pre-set pre-emphasis filter. The maximum tunability design uses only one-tap filter ($n_{max} = 1$) to allow maximum adjustability. The no-tunability design and maximum tunability design also serve as the representative of maximum design-time effort and maximum post-silicon effort, respectively.

For fair comparison, the data rate for all the designs is set to be $5GHz$ and the threshold BER for yield $\bar{e} = 1.0 \times 10^{-15}$. In our experiments, we assume that the channel is a 30cm differential microstrip line on FR-4 substrates and that the receiver has ideal timing recovery. We also assume that V_{th} variation follows normal distribution.

We first present the BER distribution with 20% V_{th} variation based on $10K$

Monte Carlo runs in Fig. 3.11. The area is constrained to $1000um^2$ and the power is constrained to $10mW$. First, for all the four methods, the distributions show strong non-symmetry and non-Gaussianity. This should be attributed to the non-linear relationship between the V_{th} and BER. Second, we can see that the ranges of BER vary for the four methods: the maximum and minimum tunability design gives the smallest and largest variations respectively, with the other two methods in between. This is in accordance with the intuition that more tunability corresponds to less variation. Third, we can see that our design gives the smallest mean BER while the minimum tunability design gives the largest mean BER. Moreover, compared with the design heuristic, our design optimizes the BER distribution with better mean and smaller variance. This verifies that our joint design-time and post-silicon optimization can significantly improve performance when compared with design-time or post-silicon only optimization, or the heuristic method.

Next, we quantitatively study how the yield from our design and design heuristic vary with respect to different area constraints for fixed power ($P = 10mW$) and 20% V_{th} variation. [2] The yield is defined as the percentage of the chips meeting the BER as in (3.6). The results are presented in Fig. 3.12 (a). From the figure we can see that for different area specs, our design always gives a larger yield than the design heuristic. Moreover, with the tightening of the area spec, the yield degradation of our method is slower than the design heuristic. When the area is limited to $700um^2$, we have a 47% yield improvement over the design heuristic. Finally, it is interesting to note the area saturation effect: When the area constraint is larger than $1200um^2$, the yield does not improve because the design is dominated by the power constraint. We observe that for the $10mW$ power limit, the optimized design area cannot exceed $1200um^2$, regardless of the

[2] The BER distributions from our design and design heuristic are better than those from the no/maximum tunability designs in orders of magnitude, thus rendering the yield of the latter two designs close to zero for the same threshold. Accordingly we exclude them for the quantitative comparison.

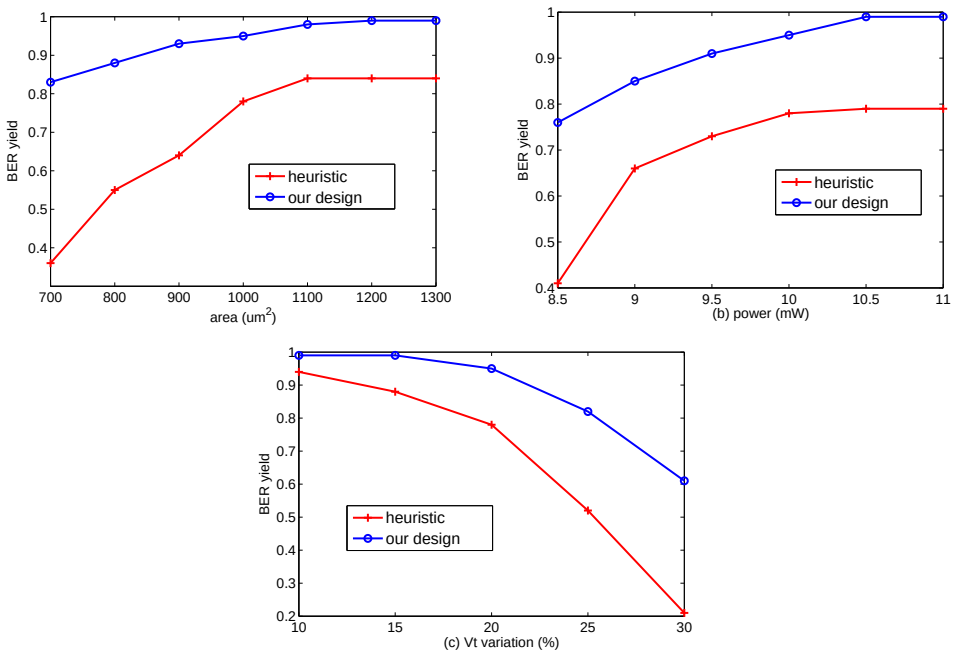

Figure 3.12: Yield curves for our designs and design heuristic with respect to area (a), power (b) and V_{th} (c).

maximum area allocated. This verifies our discussion that the power and area constraints are strongly coupled.

A similar study is conducted with respect to different power constraints for fixed area ($A = 1000um^2$) and 20% V_{th} variation as shown in Fig. 3.12 (b). From the figure we can see that for different power specifications, our design also gives better yield and better scalability than the design heuristic. When the power is limited to $8.5mW$, we have a 35% yield improvement over the design heuristic. The power saturation effect is also observed here when the area constraint becomes dominant. In addition, we study how the amount of V_{th} variation affects the yield for the four methods for fixed power ($P = 10mW$) and area ($A = 1000um^2$) constraints. Although V_{th} variation is not explicitly listed as a constraint and only appears in the power and area constraints, it affects the yield significantly. Our design improves the yield by 40% when compared with the design heuristic

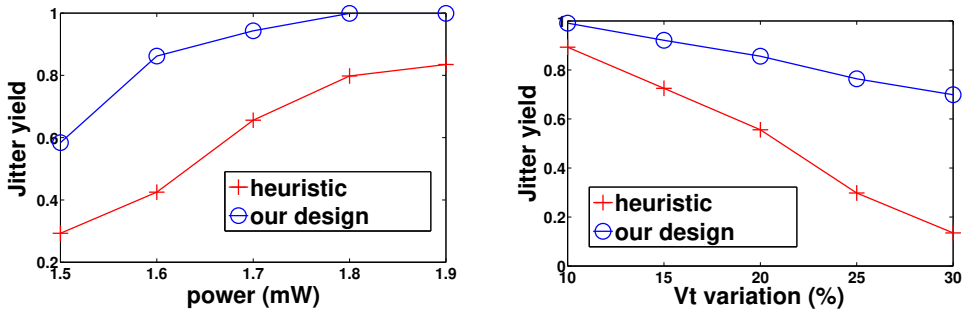

Figure 3.13: Yield for our algorithms and design heuristic w.r.t power (a) and V_{th} (b) in the PLL.

with 30% variation, as shown in Fig. 3.12 (c). In terms of runtime, the developed framework is very efficient. For different power and area constraints, the runtime varied between 30 minutes and 1 hour.

3.7.2 PLL Design

The same optimization framework is applied to a PLL design example and the result is provided in Fig. 3.13. We compare our algorithm with the design heuristic that has optimal I_{CP1} and I_{CP2} values through design time optimization and tunable elements in the biasing circuit consumes negligible power [MK02]. The reference clocks of the PLL for both designs are set to $700MHz$. We assume that the V_{th} variation follows normal distribution. The yield is defined as the percentage of the chips meeting the jitter requirement, as in (3.47). The experiment was conducted with respect to different power constraints for fixed area and 30% V_{th} variation, as shown in Fig. 3.13(a). From the figure we can see that for different power specs, our design provides better yield than the design heuristic and obtains up to 29% yield improvement. In Fig. 3.13(b), when the power is limited to $17mW$, we have a 56% yield improvement over the design heuristic.

3.8 Conclusions

Joint design time and post-silicon optimization for analog circuits has been an open problem in literature, given the complex nature of analog circuit modeling and optimization. In this chapter we formulate a co-optimization problem for digitally tuned analog circuits to optimize the parametric yield, subject to power and area constraints. A general optimization framework combing the branch-and-bound algorithm and gradient ascent method is proposed. We demonstrate our framework with two examples in high-speed serial link, the transmitter design and the phase-locked-loop (PLL) design. Experimental results show that compared with the design heuristic from analog designers' perspective, joint design-time and post-silicon optimization can improve the yield by up to 47% for transmitter design and up to 56% for PLL design under the same area and power constraints.

CHAPTER 4

Modeling and Application of Multi-Port TSV Networks in 3D IC

4.1 Introduction

As traditional CMOS scaling pace gradually slows down, three-dimensional (3D) integration offers another dimension of scaling by means of stacking functional blocks vertically and providing high integration density, fast signal transmission, low power consumption, and heterogeneous integration opportunities in the "More-than-Moore" era [BSK01, BAC07]. Through-silicon-via (TSV) has been well regarded as a key component in 3D integration, connecting chips vertically with shortened electrical delay and providing extremely dense I/O connections. While TSV fabrication technologies have progressed [PBW08], it is vitally important to understand TSV electrical properties accurately and efficiently for 3D system-performance analysis and subsequent design optimization.

In order to evaluate electrical behavior including delay, power consumption, signal integrity (SI), and power integrity (PI) for 3D ICs, it is desirable to have a SPICE-compatible equivalent circuit model for TSV networks. One approach is to employ an accurate, full-wave numerical simulator. However, this method is slow and memory intensive. Consequently, it is not suitable for large-scale analysis and design optimization. The partial element equivalent circuit (PEEC) model [Rue72,RB73] has been widely used in inductance and capacitance extraction tools for planar on-chip traces. In a typical PEEC model, the reference has to be defined

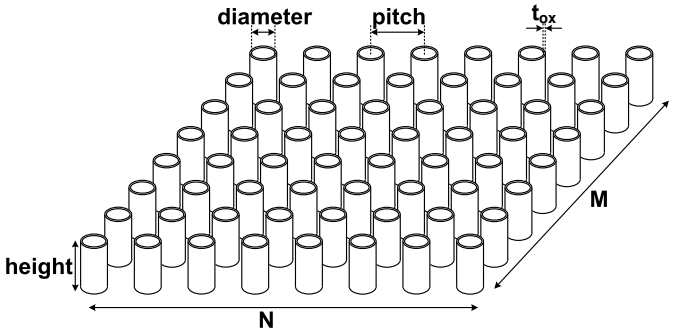

Figure 4.1: Multi-port TSV network example: A $M \times N$ TSV array

at infinity for partial inductance and capacitance [RB73, HCL99]. However, when applying the PEEC model to TSV networks, the complex 3D metal-insulator-semiconductor (MIS) structure of TSV and the lossy silicon substrate make it difficult to find the partial inductance and partial capacitance efficiently.

Instead, an extensive amount of work has been done to model a single signal-ground pair of TSVs analytically [PRK07, PCK10, CFB09, XL11, XLS10, XKS11]. As in [XLS10, XKS11], compact RLGC models for a single pair of TSVs are proposed for a wide frequency range, with consideration of the MOS depletion region effect, the alternating-current (AC) conduction and eddy currents in silicon, and the skin effect in TSV metal. Though these two-port TSV pair models are already verified against electrostatic measurements as well as electromagnetic (EM) simulations, they are no longer valid for any pairs of the TSV in the array structure, as the other surrounding TSVs can affect the distributions of electromagnetic fields. To model multi-port TSV networks, [SRI11] and [PKC11] both proposed empirical parasitic models for various TSV array structures by dimensional analysis and curve fitting through EM simulations. These multi-TSV models, however, are only available for a limited number of multi-TSV arrangements and, again, cannot be applied to general multi-port TSV networks.

In this chapter, we introduce a comprehensive yet accurate modeling method-

ology to expand TSV pair models for general multi-port TSV networks based on the proposed pair-based equivalent circuit model. An example of a multi-port TSV network is shown in Fig. 4.1. In our proposed method, we extend the PEEC method with impedance and admittance between TSV pairs and provide a frequency-dependent SPICE-compatible RLGC equivalent circuit for a multi-port TSV network. Design studies to evaluate crosstalk and power integrity of TSV arrays are also discussed based on our proposed multi-TSV modeling techniques.

4.2 Preliminary on TSV Modeling

The characteristics of TSV are dependent on its geometrical parameters such as TSV radius (r_{via}), height (H), oxide layer thickness (t_{ox}), center-to-center distance or pitch (d), and electrical parameters such as metal conductivity (σ_{metal}), oxide permittivity (ϵ_{ox}), and the silicon substrate permittivity (ϵ_{Si}) and conductivity (σ_{Si}).

The conventional PEEC method, firstly proposed in [Rue72, RB73], could be applied to extract the inductance and the capacitance between each conductor cylinder in the TSV Networks. The partial inverse capacitance matrices (P_s) and the inductance matrices (L) are defined as

$$P_{s_{ij}} = \frac{1}{A_i A_j} \int_{S_i} \int_{S_j} G_\phi(r, r') ds ds', \tag{4.1}$$

$$L_{ij} = \frac{1}{a_i a_j} \int_{\Omega_i} dv \int_{\Omega_j} dv' \overline{G_A}(r, r') f_i(r) f_j(r'), \tag{4.2}$$

where A_i, A_j are the total surface areas and S_i, S_j are the surfaces for conductor i and j. Ω_i, Ω_j, a_i, and a_j are the volumes and cross section areas of conductor i and j, respectively. f_i and f_j are the current distribution functions over a_i and a_j. $G_\phi(r, r')$ and $\overline{G_A}(r, r')$ are dyadic Greens functions for electric and magnetic potentials in MIS environments. However, the derivations of dyadic Greens function in inhomogeneous medias are generally difficult problems and are often

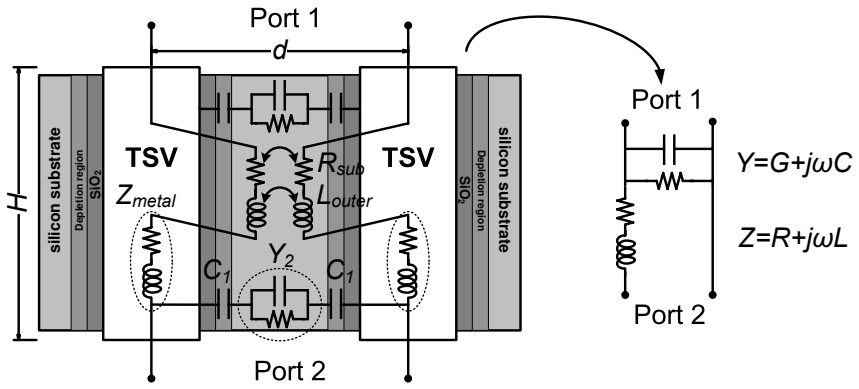

Figure 4.2: Structures and dimensional variables of a single TSV pair and its RLGC equivalent circuit model [XLS10].

solved in spectrum domain [Che99]. The calculations of those Greens functions are time-consuming as the numerical integration of Sommerfeld integrals has to be performed [CYF91].

Instead, an extensive amount of work has been done to model a single signal-ground pair of TSV analytically. In [PRK07, XL11, CFB09], a parametric and frequency-dependent equivalent circuit model is developed by employing EM simulations. In [PCK10], a MIS structure signal-ground (SG) TSV equivalent circuit model is proposed based on TSV physics and closed-form equations. In [XLS10], an equivalent circuit model that considers the width of MIS depletion region for a single TSV pair is proposed. With various models available for a TSV pair, one-dimensional frequency-dependent RLGC parameters can always be extracted from a two-port network, considering one TSV as the signal and the other as a reference as shown in Fig. 4.2. These parameters are actually loop impedance and admittance. For example, in [XLS10], the parallel admittance per unit TSV height (Y) can be expressed as a series of capacitances of the dielectric and silicon depletion region (C_1) and the coupling admittance due to bulk silicon (Y_2),

$$Y = [2(j\omega C_1)^{-1} + Y_2^{-1}]^{-1} = G + j\omega C, \tag{4.3}$$

where ω is the radial frequency and

$$C_1 = [\frac{1}{2\pi\epsilon_{ox}} \cdot ln(1 + \frac{t_{ox}}{r_{via}}) + \frac{1}{2\pi\epsilon_{Si}} \cdot ln(1 + \frac{w_{dep}}{r_{via} + t_{ox}})]^{-1}, \tag{4.4}$$

$$Y_2 = \pi(\sigma_{Si} + j\omega\epsilon_{Si})/arccosh(\frac{d}{2}/(r_{via} + t_{ox} + w_{dep})), \tag{4.5}$$

where w_{dep} is the silicon depletion width. The depletion width is dependent on the geometrical parameters, the Si/SiO_2 interface charge due to plasma damage during via-hole etching or dielectric deposition, the semiconductor type, and the bias voltage. Specifically, take n-type silicon substrate in depletion/weak inversion condition as an example, w_{dep} can be determined by solving the relation between the total charge and the electrical potential at the Si/SiO_2 interface [XLS10],

$$\begin{aligned} &\pi[(r_{via} + t_{ox} + w_{dep})^2 - (r_{via} + t_{ox})^2] \cdot qN_d + 2\pi(r_{via} + t_{ox})Q_i = \\ &-[V - \phi_{ms} - \psi(r_{via} + t_{ox})] \cdot 2\pi\epsilon_{ox}/ln(1 + \frac{t_{ox}}{r_{via}}), \end{aligned} \tag{4.6}$$

$$\begin{aligned} &\psi(r_{via} + t_{ox}) = \\ &-\frac{qN_d}{2\epsilon_{Si}} \cdot [(r_{via} + t_{ox} + w_{dep})^2 ln(1 + \frac{w_{dep}}{r_{via} + t_{ox}}) - \frac{1}{2}w_{dep}^2 - w_{dep}(r_{via} + t_{ox})], \end{aligned} \tag{4.7}$$

where N_d is n-type bulk doping concentration, q is the elementary charge, Q_i is the interface charge density, V is the bias voltage, ϕ_{ms} is the work function difference between the TSV metal and the doped silicon, and $\psi(r_{via} + t_{ox})$ is the potential drop in the depletion region.

The serial impedance per unit length (Z) can be treated as the sum of inner impedance of the TSV (Z_{metal}), the outer inductance (L_{outer}) and the resistance due to eddy current in the substrate (R_{sub}) [XLS10], i.e.

$$Z = 2Z_{metal} + j\omega L_{outer} + R_{sub} = R + j\omega L. \tag{4.8}$$

Among them,

$$Z_{metal} = \frac{(1 - j) \cdot J_0((1 - j)r_{via}/\delta_{metal})}{\sigma_{metal} \cdot 2\pi r_{via}\delta_{metal} \cdot J_1((1 - j)r_{via}/\delta_{metal})}, \tag{4.9}$$

where $\delta_{metal} = \sqrt{2/\omega\mu\sigma_{metal}}$ is the damping parameter for metal and J_0 and J_1 are the 0th order and 1st order Bessel functions of the first type.

$$R_{sub} = \frac{\omega\mu}{2} \cdot Re[H_0^{(2)}(\frac{1-j}{\delta_{Si}}(r_{via} + t_{ox} + w_{dep})) - H_0^{(2)}(\frac{(1-j)d}{\delta_{Si}})], \tag{4.10}$$

where $\delta_{Si} = \sqrt{2/\omega\mu(\sigma_{Si} + j\omega\epsilon_{Si})}$ is the damping parameter for silicon substrate. The purely real R_{sub} is physically due to eddy current. The outer inductance L_{outer} can be approximated by the reciprocal of the capacitance between two TSVs as if the medium permittivity is $1/\mu$:

$$L_{outer} \approx \frac{\mu}{\pi} arccosh(\frac{d}{2r_{via}}). \tag{4.11}$$

In [XKS11], the outer series impedance R_{sub} and L_{outer} are further improved from a 2D per-unit-length model to a first-order approximated 3D model when the TSV height is comparable to the TSV pitch, where

$$R_{sub} \approx \frac{H^4\omega^2\mu^2\sigma_{Si}}{12\pi\sqrt{r_{via}^2 + \frac{169}{400}H^2}} - \frac{H^4\omega^2\mu^2\sigma_{Si}}{12\pi\sqrt{d^2 + \frac{169}{400}H^2}}, \tag{4.12}$$

$$\begin{aligned} L_{outer} \approx & \frac{\mu}{\pi} \cdot [r_{via} + H \cdot arcsinh(\frac{H}{r_{via}}) - \sqrt{r_{via}^2 + H^2}] \\ & -\frac{\mu}{\pi} \cdot [d + H \cdot arcsinh(\frac{H}{d}) - \sqrt{d^2 + H^2}]. \end{aligned} \tag{4.13}$$

The TSV pair model proposed in [XLS10,XKS11] is used in the rest of the chapter to model each TSV pair in the multi-port TSV network. However, as long as these impedance and admittance between a single TSV pair can be extracted through other modeling techniques, simulation or even measurement, they can be directly applied in our techniques for modeling multi-port TSV network.

4.3 Multi-port TSV Network Modeling

4.3.1 Framework Overview

In this section, we propose a detailed procedure to reconstruct $S \times S$ RLGC matrices for N TSVs with S signals and $(N-S)$ ground TSVs. First, the potential matrix can be expressed as the inverse of the capacitance matrix [RB73] and, for a TSV pair with two individual conductors,

$$P_s = \begin{bmatrix} P_{s_{11}} & P_{s_{12}} \\ P_{s_{12}} & P_{s_{11}} \end{bmatrix} = \begin{bmatrix} C_{11}+C_{12} & -C_{12} \\ -C_{12} & C_{11}+C_{12} \end{bmatrix}^{-1}, \tag{4.14}$$

where C_{11} and C_{12} are the partial self and mutual capacitances defined when the reference is set to be at infinity. For simplicity, we assume all the TSVs in the network are identical. The pair-based capacitance C_p is defined as the capacitance between two TSVs, with one TSV as the signal and the other one as the reference, and can be expressed as

$$C_p = C_{12} + (C_{11}^{-1} + C_{11}^{-1})^{-1} = \frac{1}{2(P_{s_{11}} - P_{s_{12}})}. \tag{4.15}$$

In a multi-TSV network, as long as the charges uniformly distribute along the surface of the TSV conductor [1] , the self terms $P_{s_{ii}}, P_{s_{jj}}$ and mutual terms $P_{s_{ij}}$ of two TSVs are *solely* determined by the media and the geometry of TSV i and j themselves [RB73]. Thus, the partial potential matrix for a general N-TSV network can be expressed as

$$P_s = \begin{bmatrix} P_{s_{11}} & P_{s_{11}} - \frac{1}{2C_{p_{12}}} & \cdots & P_{s_{11}} - \frac{1}{2C_{p_{1N}}} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s_{11}} - \frac{1}{2C_{p_{N1}}} & P_{s_{11}} - \frac{1}{2C_{p_{N2}}} & \cdots & P_{s_{11}} \end{bmatrix}, \tag{4.16}$$

where $C_{p_{ij}}$ is the pair-based capacitance of a TSV pair between TSV i and j, which can be obtained using (4.3), and $P_{s_{11}}$ is the inverse of partial self capacitance for

[1] The charges in TSV are uniformly distributed as long as the center-to-center distance between TSVs is more than 6× TSV radius such that the proximity effect can be neglected. A 3-TSV network example, includes 1 signal TSV and 2 ground TSV, is showed in Fig. 4.3. When TSV pitch is larger than 6X TSV radius, the inductance error is saturated and the proximity effect can be neglected. When TSV pitch is less than 6X TSV radius, the inductance error rises up dramatically due to the non-uniform current distribution on each TSV.

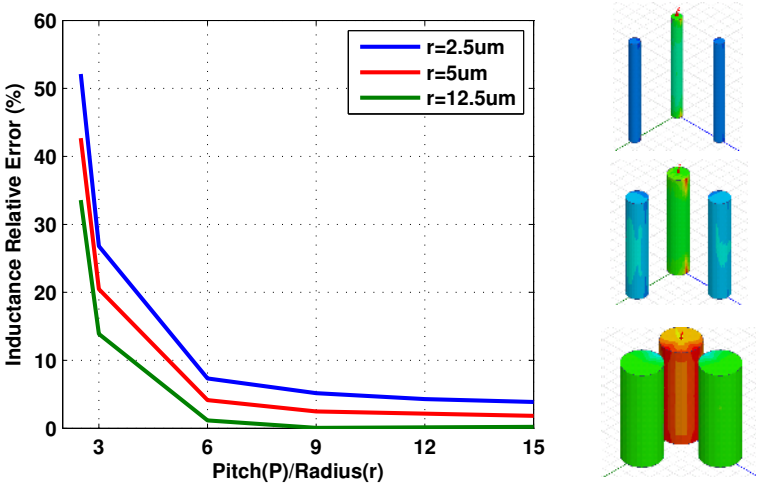

Figure 4.3: The charge distribution and the inductance comparison of a 3-TSV network show the proximity effect can be neglected when TSV pitch is larger than 6× TSV radius. The equivalent inductance is calculated at 20GHz based on our methodology and the TSV pair model in [XKS11]. The results are compared to Ansoft Q3D [ANS] with various TSV radius and pitch/radius ratios.

a single TSV. We again assume all the TSVs in the network are identical. To obtain the inverse of the conductance matrix, G^{-1}, for a general N-TSV network, the procedure is the same as the P_s matrix.

A similar procedure can also be applied to construct the inductance matrix L and resistance matrix R of a N-TSV network. Taking inductance as an example, the inductance matrix L for two TSV conductors can be expressed as

$$L = \begin{bmatrix} L_{11} & L_{12} \\ L_{12} & L_{11} \end{bmatrix} = \begin{bmatrix} L_{11} & L_{11} - \frac{1}{2}L_p \\ L_{11} - \frac{1}{2}L_p & L_{11} \end{bmatrix}, \tag{4.17}$$

where L_{11} is the partial self inductance of a single TSV and L_p is the loop inductance of a TSV pair. Fundamentally, in a multi-conductor system, the self inductance of one conductor is *solely* decided by the conductor itself, while the mutual inductance of two conductors is *solely* decided by the two conductors [Rue72, HCL99]. The full partial inductance matrix for a general N-TSV

network can be expressed as

$$L = \begin{bmatrix} L_{11} & L_{11} - \frac{1}{2}L_{p_{12}} & \cdots & L_{11} - \frac{1}{2}L_{p_{1N}} \\ \vdots & \vdots & \ddots & \vdots \\ L_{11} - \frac{1}{2}L_{p_{N1}} & L_{11} - \frac{1}{2}L_{p_{N2}} & \cdots & L_{11} \end{bmatrix}, \tag{4.18}$$

where $L_{p_{ij}}$ is the loop inductance between TSV i and j. Again, for simplicity, we assume all the self terms are the same and the loop inductance $L_{p_{ij}}$ can be obtained for each TSV pair using (4.8).

For typical TSV geometry, $C_{p_{ij}}$ and $L_{p_{ij}}$ can be calculated easily, for example by using equation (4.3) and (4.8). However, it is difficult to analytically calculate $P_{s_{11}}$ (or $1/C_{11}$) and L_{11} for a long cylinder with inhomogeneous media. Note the reference for the full potential matrix P_s and the return path for the full inductance matrix L are set to be at infinity. In reality, however, ground TSVs are designed to be the return path and reference of the high-speed signals. In Section 4.3.2, it can be proven that the reduced potential matrix P_{s_r} and the reduced inductance matrix L_r are independent of the values of $P_{s_{11}}$ and L_{11} if at least one of the TSVs is set as the reference of the network. Thus, we could simply choose both $P_{s_{11}}$ and L_{11} as 0 to simplify our procedure. Physically, it means the *relative* potential difference between any two TSVs is independent to the selection of the reference point.

To get the reduced inductance matrix L_r and the reduced potential matrix P_{s_r} for a N-TSV network with $(N - S)$ reference TSVs, we need to connect those reference TSVs in parallel and set them as the current return path for the S signal TSVs. Take a 3-TSV network with full inductance matrix L as an example. Since $V = j\omega L \cdot I$, then $\frac{1}{j\omega}L^{-1} \cdot V = I$, or equivalently,

$$\frac{1}{j\omega} \begin{bmatrix} (L^{-1})_{11} & (L^{-1})_{12} & (L^{-1})_{13} \\ (L^{-1})_{21} & (L^{-1})_{22} & (L^{-1})_{23} \\ (L^{-1})_{31} & (L^{-1})_{32} & (L^{-1})_{33} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}. \tag{4.19}$$

Assume TSV_1 is the only signal TSV and both TSV_2 and TSV_3 are reference TSVs. After connecting them in parallel and setting $V_2 = V_3$, the reduced inductance matrix L_1 should satisfy

$$\frac{1}{j\omega} L_1^{-1} \begin{bmatrix} V_1 \\ V_2 (= V_3) \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 + I_3 \end{bmatrix}, \tag{4.20}$$

with L_1^{-1} expressed as

$$\begin{bmatrix} (L^{-1})_{11} & (L^{-1})_{12} + (L^{-1})_{13} \\ (L^{-1})_{21} + (L^{-1})_{31} & (L^{-1})_{22} + (L^{-1})_{23} + (L^{-1})_{32} + (L^{-1})_{33} \end{bmatrix}, \tag{4.21}$$

or equivalently,

$$L_1 = (A(L^{-1})A')^{-1}, \quad where \quad A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \tag{4.22}$$

We now set these parallel-connected reference TSVs as the current return path, which means

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = j\omega L_1 \begin{bmatrix} I_1 \\ -I_1 \end{bmatrix}, \tag{4.23}$$

and the resulting reduced impedance matrix L_r satisfies

$$(V1 - V2) = j\omega L_r \cdot I_1, \tag{4.24}$$

where $I_1 = -I_2 - I_3$. As a result, L_r can be expressed as

$$\begin{aligned} L_r \;=\; & (L^{-1})_{11} \\ & -((L^{-1})_{12} + (L^{-1})_{13}) - ((L^{-1})_{21} + (L^{-1})_{31}) \\ & +((L^{-1})_{22} + (L^{-1})_{23} + (L^{-1})_{32} + (L^{-1})_{33}), \end{aligned} \tag{4.25}$$

or equivalently,

$$L_r = B(L_1)B' = B(AL^{-1}A')^{-1}B', \tag{4.26}$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \end{bmatrix}. \tag{4.27}$$

Algorithm 4 Pair-based equivalent circuit method for N-TSV network modeling.

for $i = 1 : N$ **do**
  for $j = 1 : N$ **do**
    if $i \neq j$ **then**
      $Z_{ij} = -\frac{1}{2} Z_p$;
      $(Y^{-1})_{ij} = \frac{-1}{2Y_p}$;
    else
      $Z_{ii} = 0$; $(Y^{-1})_{ii} = 0$;
Rearrange Z and Y matrix, if necessary, for reference TSVs;
Prepare matrix A and B according to (4.30);
$Z_r = B(AZ^{-1}A')^{-1}B'$;
$Y_r = (B(AYA')^{-1}B')^{-1}$

This procedure can be generalized to find the reduced impedance matrix Z_r and the reduced admittance matrix Y_r for any N-TSV networks, where S of them are signal TSVs and $(N - S)$ of them are reference TSVs, with

$$Z_r = B(AZ^{-1}A')^{-1}B', \tag{4.28}$$

$$Y_r = (B(AYA')^{-1}B')^{-1}, \tag{4.29}$$

and

$$A = \left[\begin{array}{c|c} I_{(S+1)} & \begin{array}{c} 0 \\ \\ 1_{1\times(N-S-1)} \end{array} \end{array}\right], \quad B = \left[\begin{array}{cc} I_S & -1_{S\times 1} \end{array}\right]. \tag{4.30}$$

Note that $S < N$ so that at least one of the TSVs is a reference. A is a $(S+1)\times N$ matrix and B is a $S \times (S+1)$ matrix. Here we assume the first S TSVs in the full Z and Y matrix are signal TSVs and the remaining $(N - S)$ TSVs are reference TSVs. If not, it can be re-arranged to this form by multiplying Z and Y with permutation matrices. A summary of our proposed modeling methodology can be found in Algorithm 1. The TSV loop impedance and pair-based admittance Z_p and Y_p can be obtained through (4.8) and (4.3) or by other TSV pair models as well.

4.3.2 Pair-based Equivalent Circuit Model

In a typical PEEC model, partial inductance and capacitance of a open loop are defined as the integration of the fields to infinity. Computational methods with filament approximation were widely developed [Rue72,HCL99,RB73] based on the concepts of those partial elements. In some situations, however, when integral-equation based PEEC methods cannot obtain partial elements easily while the loop or pair-based elements among filaments are available through other methods, a numerical scheme as Algorithm 1 can be used to obtain circuit models for these complex structures. The reference needs to be set on at least one or more filaments in the network when applying Algorithm 1. The following derivations demonstrate that the resulting equivalent circuit model for the network is only related to the impedance and admittance between each pair of filaments. The information for the partial self inductance and self capacitance with reference at infinity has no effect on the resulting equivalent circuit model. Note in the modeling of a multi-port TSV network, we can simply take each TSV as one filament as long as the proximity effect can be neglected as shown in Fig. 4.3.

Without loss of generality, we use impedance matrix Z as an example. For a general N-TSV system, the impedance matrix Z can be expressed as $Z = G + H$ where

$$G = \begin{bmatrix} 0 & -\frac{1}{2}Z_{p_{12}} & \cdots & -\frac{1}{2}Z_{p_{1N}} \\ -\frac{1}{2}Z_{p_{12}} & 0 & \cdots & -\frac{1}{2}Z_{p_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{2}Z_{p_{1N}} & -\frac{1}{2}Z_{p_{2N}} & \cdots & 0 \end{bmatrix}, \tag{4.31}$$

and

$$H = \begin{bmatrix} Z_{11} & \cdots & Z_{11} \\ \vdots & \ddots & \vdots \\ Z_{NN} & \cdots & Z_{NN} \end{bmatrix}, \tag{4.32}$$

where $Z_{p_{ij}}$ is the pair-based impedance between each TSV pair (i, j) and Z_{ii} is

the partial self impedance for TSV i. All the TSVs in the network are identical. Note that G is symmetric and H is rank-one. We would like to show that the choice of Z_{ii} actually has no impact on the reduced impedance matrix Z_r as long as we set at least one of the TSVs as the reference.

To start with, from (4.28), we have

$$Z_r = B(AZ^{-1}A')^{-1}B' = B(A(G+H)^{-1}A')^{-1}B', \tag{4.33}$$

where A and B can be found in (4.30).

Since G is symmetric and H is rank-one, by applying a Lemma in [Mil81],

$$\begin{aligned} Z_r &= B(A(G+H)^{-1}A')^{-1}B' \\ &= B(AG^{-1}A' + \frac{-1}{1+g}AG^{-1}HG^{-1}A')^{-1}B' \\ &= B(G_1+H_1)^{-1}B', \end{aligned} \tag{4.34}$$

where $g = tr(HG^{-1})$. H_1 is still rank-one since $rank(XY) \leq min(rank(X), rank(Y))$ for arbitrary matrix X and Y. The lemma in [Mil81] can be applied again and we get

$$\begin{aligned} Z_r &= B(G_1^{-1} + \frac{-1}{1+g_1}G_1^{-1}H_1G_1^{-1})B' \\ &= B(AG^{-1}A')^{-1}B' + \frac{1}{(1+g)(1+g_1)} \cdot KHK', \end{aligned} \tag{4.35}$$

where $g_1 = tr(H_1G_1^{-1})$ and $K = B(AG^{-1}A')^{-1}AG^{-1}$.

It can be shown that

$$KHK' = 0, \tag{4.36}$$

such that

$$Z_r = B(AG^{-1}A')^{-1}B', \tag{4.37}$$

is not related to H, and so Z_r is not related to the choice of Z_{ii}. The detailed proof can be found in the Appendix.

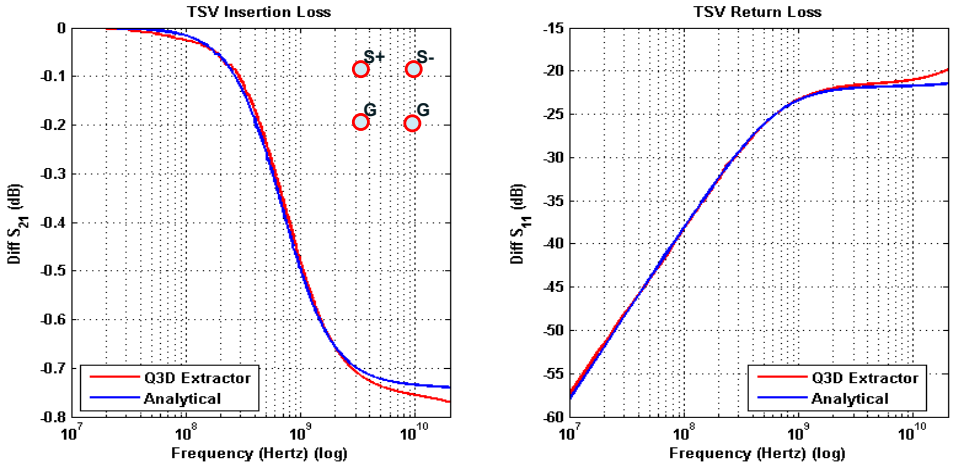

Figure 4.4: 2x2 TSV differential pair differential S-parameter comparison. The TSV diameter is 25μm and pitch is 150μm

4.3.3 Model Validation via Simulation

From the previous section, it has been shown that by applying our proposed methodology for general multi-port TSV networks, the resulting RLGC models in the reduced impedance (Z_r) and admittance (Y_r) matrices are only related to the equivalent circuit model of each TSV pair within the network. As long as the pair model for each TSV pair is accurate, the resulting multi-port TSV network model is also accurate without the knowledge of partial self elements. To even further verify our method for a multi-port TSV network, a 2×2 TSV array with 2 signal TSVs and 2 ground TSVs is first simulated for its differential insertion loss (S21 magnitude) and differential return loss (S11 magnitude) and compared to a commercial EM simulator (Ansoft Q3D [ANS]), as shown in Fig. 4.4. The TSV diameter is 25μm and pitch is 150μm. The height of the TSV is 150μm and the silicon dioxide thickness is 0.5μm. The resistivity of silicon substrate is 10Ω-cm. From Fig. 4.4, it can be shown that our model correlates well with the commercial EM simulator up to $20GHz$ for this 2×2 TSV differential network.

Another 4×4 TSV array case with 12 signal TSVs in peripheral and 4 ground

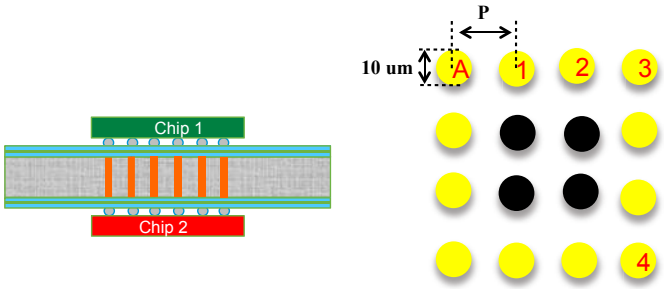

Figure 4.5: TSV Networks in silicon interposer for 2.5D chip-to-chip communication

TSVs in center, as shown in Fig. 4.5, is also modeled using our proposed methodology. The reduced impedance and admittance matrices are compared to a commercial EM simulator (Ansoft Q3D [ANS]) with its extracted RLGC values, as shown in Fig. 4.6. The TSV diameter is 10μm and silicon dioxide thickness is 0.5μm. The height of the TSV is 150μm and the pitch for this TSV array is 40μm. The resistivity of silicon substrate is 10Ω-cm. In Fig. 4.6, the self terms (TSV_A) and the mutual terms (between TSV_A and TSV_1) of all the RLGC elements are compared between our method and the extracted results. It can be shown that, again, our modeling method correlates well with the commercial EM simulator for both self and mutual RLGC values up to 20GHz.

Note that since in the experiment our methodology is based on an analytical model between each TSV pair without any meshing or other computational electromagnetic procedures, our method could achieve orders of magnitude improvement in terms of efficiency while providing similar accuracy compared to commercial EM simulation tools.

4.4 Multi-port TSV Network Characteristics in Signal and Power Integrity

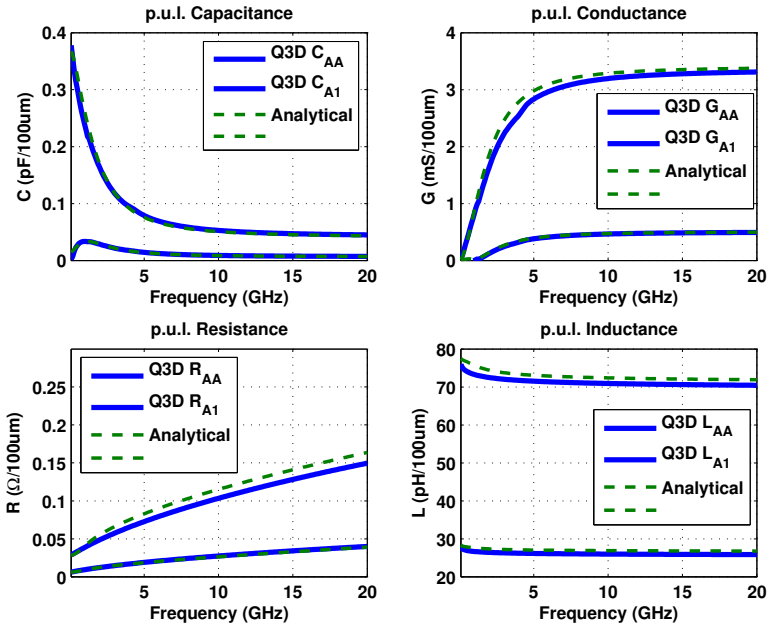

Figure 4.6: 4x4 TSV array RLGC comparison for (1) self terms of TSV_A and (2) mutual terms between TSV_A and TSV_1.

4.4.1 Crosstalk Analysis for Chip-to-chip TSV Networks in Silicon Interposer

Crosstalk in TSV networks is a critical issue in 3D integration and may have significant impacts on timing margins and signal integrity, especially for high density TSV arrays on a lossy silicon substrate. Although the TSV arrays can be modeled as a multi-conductor transmission line, the crosstalk among the TSV arrays behaves differently than the crosstalk of transmission lines with homogeneous media.

We first review the near-end crosstalk (NEXT) and far-end crosstalk (FEXT) for electrically short transmission lines by the equations below

$$V_{NE} = \frac{R_{NE}}{R_{NE}+R_{FE}} j\omega L_m I_{ag} + \frac{R_{NE}R_{FE}}{R_{NE}+R_{FE}} j\omega C_m V_{ag} \tag{4.38}$$

$$V_{FE} = \frac{-R_{NE}}{R_{NE}+R_{FE}} j\omega L_m I_{ag} + \frac{R_{NE}R_{FE}}{R_{NE}+R_{FE}} j\omega C_m V_{ag} \tag{4.39}$$

where V_{NE} and V_{FE} are the near-end and far-end phasor crosstalk voltages while

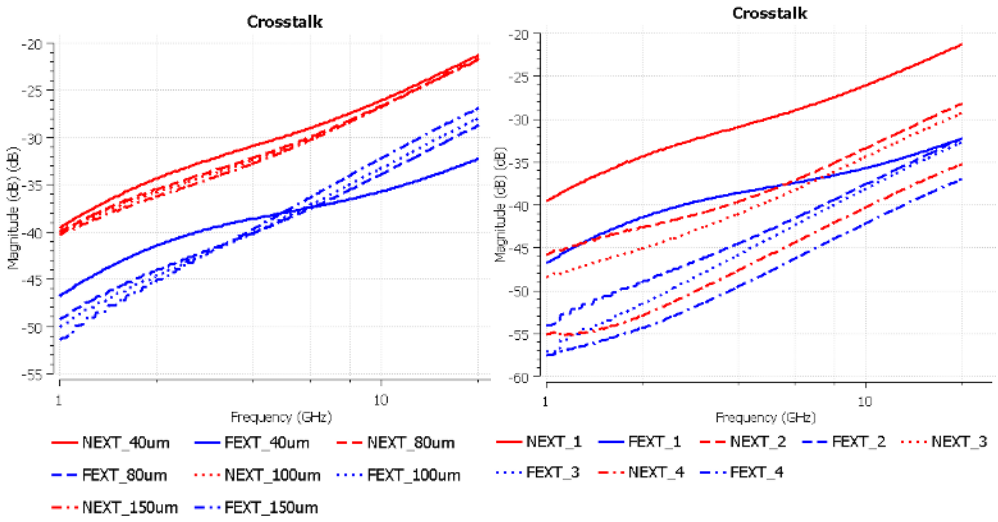

Figure 4.7: S Parameters of NEXT and FEXT between TSV_A and TSV_1 with different pitch sizes and between TSV_A and other TSVs with pitch $40\mu m$

R_{NE} and R_{FE} are terminations at the near end and far end, respectively. L_m and C_m denote the mutual inductance and capacitance between the aggressor and the victim. V_{ag} and I_{ag} are the aggressor voltage and current, respectively. (4.38) and (4.39) assume the transmission lines are weakly coupled and lossless and the crosstalk can be separated to inductive and capacitive couplings.

TSV structures can usually be considered as electrically small and weakly coupled transmission lines below 20 GHz. Thus, the near-end and far-end crosstalk in terms of the S-parameters between TSV i and j can be written as

$$S_{near} = (Z_m/Z_{ref} + Ym \cdot Z_{ref})/2 \quad (4.40)$$

$$S_{far} = (-Z_m/Z_{ref} + Ym \cdot Z_{ref})/2 \quad (4.41)$$

where Z_m and Y_m are the mutual impedance and admittance between TSV i and j in the reduced impedance and admittance matrices.

Fig. 4.5 shows a 4×4 TSV array structure, including 12 peripheral high-speed signal IOs and 4 center ground TSVs. Crosstalk between the aggressor (TSV labeled as 'A' in Fig. 4.5) and other TSVs is evaluated using our proposed model.

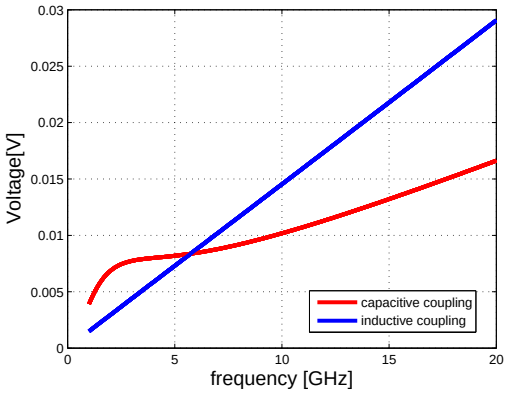

Figure 4.8: Normalized crosstalk voltages between TSV_A and TSV_1. The noise voltages are separated to capacitive and inductive couplings using (4.40) and (4.41). The pitch size is 40 μm.

The aggressor has 50Ω terminations at both source and load and all the other TSVs also have 50Ω terminations. Because the TSV array needs to be aligned with package microbumps in this 2.5D silicon interposer integration, as shown in Fig. 4.5, the pitch size for each adjacent TSV pair in this array structure has to be identical. Both near-end crosstalk (NEXT) and far-end crosstalk (FEXT) between aggressor TSV_A and TSV_1 with different pitch sizes are illustrated in Fig. 4.7(a) , while the crosstalk between TSV_A and TSV_1, TSV_2, TSV_3, and TSV_4 for a fixed pitch size of $40\mu m$ are shown in Fig. 4.7(b). First, compared to the crosstalk in ordinary transmission lines with homogeneous media, the frequency responses of total crosstalk transfer functions do not maintain 20 dB/decade. This is mainly due to the transitions of slow mode to TEM mode of semi-conductive media [Has71], which results in frequency-dependent mutual capacitance. At low frequencies of the slow-mode region, as shown in Fig. 4.8, capacitive coupling dominates while inductive coupling dominates at the TEM-mode region.

Interestingly, from Fig. 4.7(a), it can be observed that the crosstalk cannot be significantly reduced by increasing the pitch size, especially at the high frequency range. This is because although capacitive coupling decreases as pitch

Table 4.1: Mutual inductance and capacitance between TSV_A and TSV_1 at $10GHz$ with different pitch sizes.

Pitch	40μm	80μm	100μm	150μm
Mutual Inductance (nH/m)	320	350	360	380
Mutual Capacitance (pF/m)	75	50	47	40

size increases, the mutual inductance between TSV_A and TSV_1 actually increases, as shown in Table 4.1. When increasing the pitch size, not only signal TSVs are further away from each other, ground TSVs are also further away from signal TSVs, which leads to more magnetic couplings between TSV_A and TSV_1. Decreasing TSV pitch hurts NEXT with more capacitive coupling but, on the other hand, benefits FEXT at high frequency where inductive coupling dominates, as shown in Fig. 4.7(a). From Fig. 4.7(b), for a fixed pitch size, crosstalk from both inductive and capacitive coupling affects the nearest victim most. This is similar to the crosstalk in ordinary transmission lines.

Fig. 4.9 shows the time-domain simulations of the crosstalk between TSV_A and TSV_1 with different pitch sizes. The signal has a 1V voltage swing and the rise time is $30ps$, which corresponds to the knee frequency of $16.7GHz$. The TSVs are terminated at both ends by a 50Ω resistance. From Fig. 4.9, for NEXT, the peak-to-peak values of the crosstalk only change slightly with different pitch sizes, as when pitch size increases, inductive coupling increases but capacitive coupling decreases. For FEXT, a deep (inductive coupling) occurs first followed by a pump (capacitive coupling). The deep increases with the pitch size while the pump decreases with the pitch size. Increasing TSV pitch size cannot reduce the crosstalk since inductive coupling is dominant at $16.7GHz$ and the peak-to-peak crosstalk voltage actually increases when pitch size is increased. On the other hand, from Fig. 4.9, it can be shown that adding extra ground TSVs between two signal TSVs is a very effective way to reduce the crosstalk.

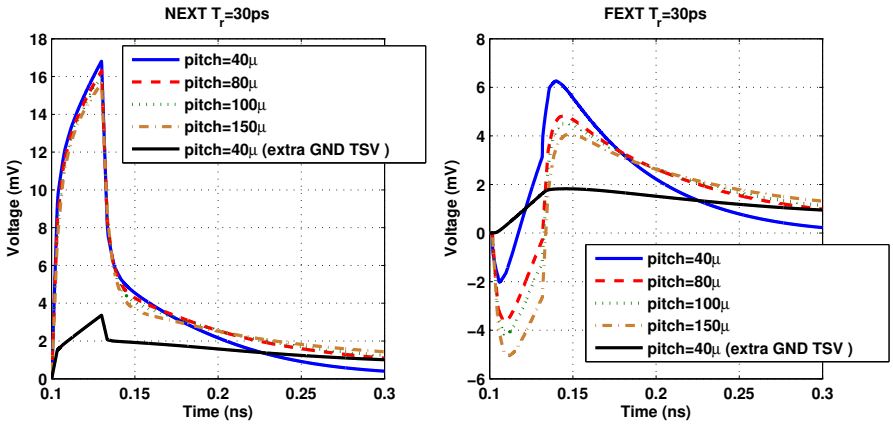

Figure 4.9: Time-domain crosstalk voltage at near and far ends between TSV_A and TSV_1, with 1V signal strength and 30ps rise time.

4.4.2 Impedance Analysis for Power/Ground TSV Array in 3D Power Distributed Network (PDN)

In order to reduce simultaneous switching noise (SSN) in a 3D integration PDN design, the impedance properties of the power/ground (P/G) TSV array must be estimated and analyzed [PKC11]. Fig. 4.10 shows a system-level PDN with three different P/G TSV array arrangements connecting one of the active die and package. TSV diameter is 10μm and the pitch is 100μm. The height of the TSV is 150μm. The current required by the fast switching chip is delivered from the voltage regulation module (VRM), through the PCB, package, and TSV P/G networks.

Fig. 4.11(a) shows the impedance comparison between different array arrangements for a P/G TSV array. In the distributed P/G TSV array, for any particular TSV, all other adjacent TSVs have currents in a reverse direction, which results in reduced mutual inductance and thus smaller impedance at high frequency. The impedance at high frequency for the grouped P/G TSV array is higher than the other two cases because its longer distance between power and ground TSVs results in higher loop inductance. Similar behavior can be found if we increase the

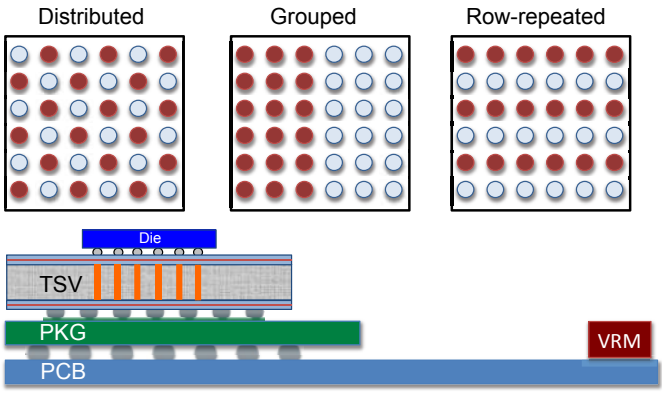

Figure 4.10: Power/ground TSV array arrangements and system-level power distribution network (PDN)

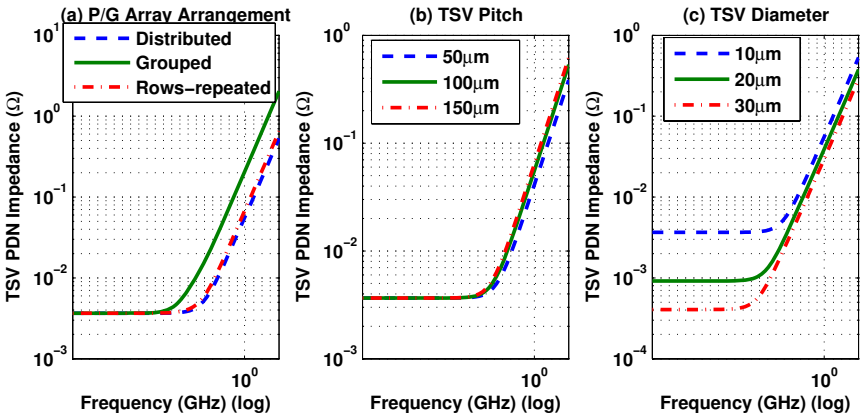

Figure 4.11: P/G TSV array impedance v.s. (a) different P/G TSV array arrangement (b) different pitch size and (c) different diameter.

pitch in a distributed P/G TSV array, as shown in Fig. 4.11(b). As pitch increases, the loop inductance increases and so does the high-frequency impedance. For frequencies less than $10MHz$, the impedance is dominated by the resistance of the P/G TSV and is not affected by the array arrangement or the pitch. By increasing the TSV diameter, the resistance of the P/G TSV is decreased and so is the low frequency TSV array impedance, as shown in Fig. 4.11(c).

By connecting our P/G TSV array model with the extracted package and PCB PDN models, a system-level impedance analysis for 3D PDN can be performed

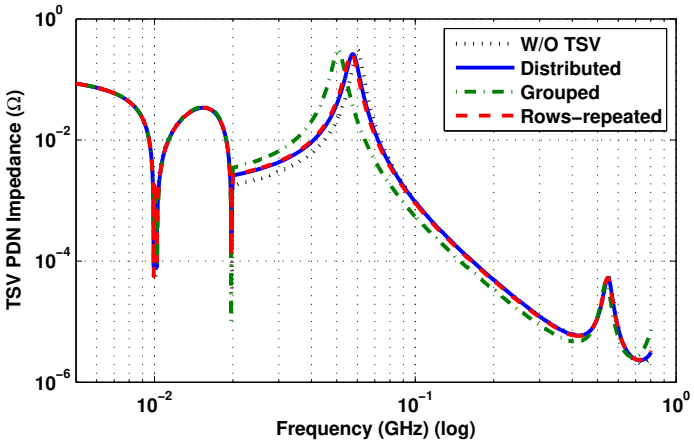

Figure 4.12: 3D PDN impedance comparison between different P/G TSV array arrangements.

between different P/G TSV array arrangements and the result is shown in Fig. 4.12. From the figure, the PDN impedance is still generally dominated by the large off-chip inductance and on-chip capacitance at frequencies below $10MHz$. However, with the grouped P/G TSV array, the PDN peak impedance gets shifted and the impedance between 20MHZ to 100MHz increases significantly due to larger inductance introduced by the grouped P/G TSV array.

4.5 TSV Array Bandwidth Analysis for 3D IC

To evaluate the bandwidth benefit of future 3D IC, stacked 4×4 TSV arrays are used to find the maximum achievable data rate per TSV channel considering different layers of stacked silicon and different CMOS technology nodes, as shown in Fig. 4.13. A simple push-pull driver is assumed for each technology node and the transistors are sized separately to have similar driving strength as a linear 40Ω resistance. The transistor models are based on the predicted technology model (PTM) for low-power applications (PTM LP), incorporating high-k/metal gate and stress effect from 45nm to 16nm [Pre]. A simple RC model (R=$150m\Omega$ and

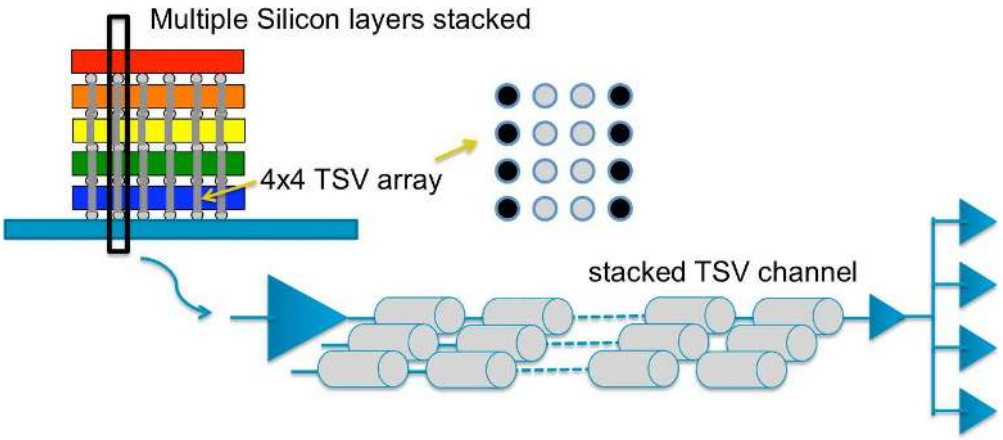

Figure 4.13: 3D IC TSV array bandwidth analysis considering different layers of stacking and different CMOS technology nodes.

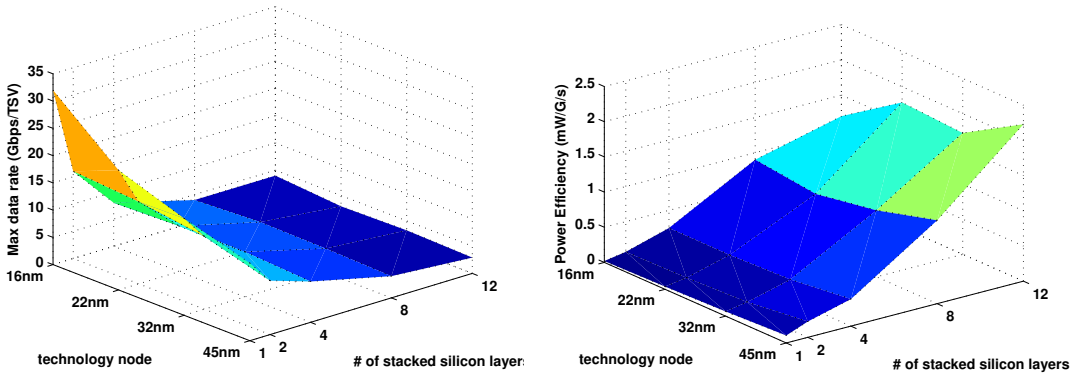

Figure 4.14: TSV array date rate per TSV and power efficiency vs number of stacked silicon layers and technology nodes.

C=$0.1fF$) [RFC12] is used to model the micro-bumps between each silicon layer. The input voltage swing is assumed to be $1.0V$ for all the cases and the maximum data rate for each stacked TSV channel is determined by its time-domain eye diagram constraints: data-dependent jitter has to be less than 10% of the unit interval (UI) and the voltage noise has to be less than $150mV$. Note that the TSV diameter is fixed to $10\mu m$ with $60\mu m$ pitch and $50\mu m$ height. The geometry of TSV is not scaled along with different technology nodes in this study.

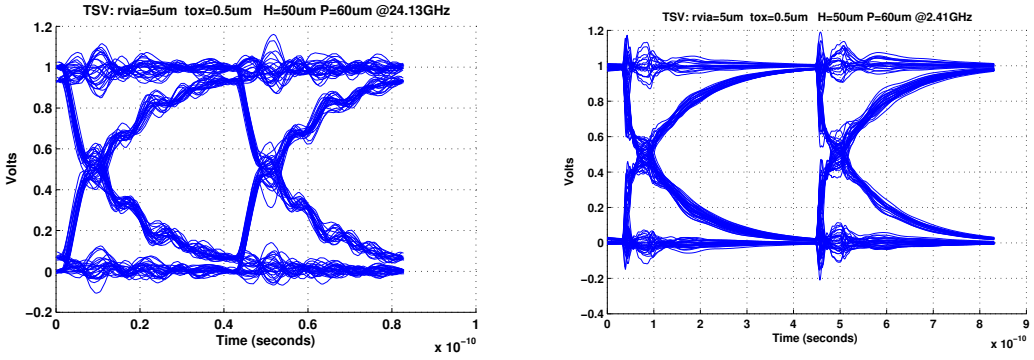

Figure 4.15: Eye diagram of (a) 1 layer of 4×4 TSV array at 24.13Gbps and (b) 12 layers of 4×4 TSV arrays at 2.41Gbps. Both cases are assumed using 16nm technology and with the same driver driving strength.

As shown in Fig. 4.14, when the number of stacked silicon layers increases, the maximum achievable date rate for each TSV channel drops significantly. At 16nm technology node, the maximum achievable data rate for only one layer of TSV array is as high as 24.1Gbps. By increasing the number of stacking to 12 layers, the data rate per TSV channel drops 10× lower to 2.4Gbps. From Fig. 4.15, it can be shown that with 12 layers of stacking, the crosstalk between signal TSVs in the array becomes significant and thus limit the data rate. On the other hand, with only one layers of TSV array, the data rate is limited by the reflection noise due to the purely capacitive loading provided by the gate capacitance in the receiver end.

With only one layer of TSV array, the link power efficiency is easily lower than $0.1mW/Gb/s$ even with the simple push-pull driver due to the short distance of the TSV. With the number of stacking increased to 12 layers, the power efficiency increases and is over $1mW/Gb/s$, which has no advantage over the current high-speed serial links. Different technology node results in different load capacitance as well as different driver output capacitance. As shown in Fig. 4.14, with only one layer of TSV array, the data rate per TSV channel increases from 14.23Gbps to 24.13Gbps when the technology improves from 45nm to 16nm. However, the data

rate differences is not obvious once the number of stacked silicon layer increases. This is because the data rate is limited by the crosstalk performance between signal TSVs in the TSV array instead of the load and source capacitance.

4.6 Conclusions

In this chapter, we introduced a general pair-based model for multi-port TSV networks. Frequency-dependent RLGC equivalent circuits for multi-port TSV networks can be extracted and composed based on pair-based impedance and admittance from the TSV pair models. The results of our proposed multi-port TSV network modeling are validated against commercial electromagnetic simulations up to 20GHz. Design guidelines for TSV arrays based on crosstalk analysis and power integrity analysis are also investigated using the proposed modeling technique.

CHAPTER 5

Modeling of RDL Coplanar Waveguide on Silicon Interposer

5.1 Introduction

Recently, 3D integration explores a new solution to keep the pace with Moores Law for the semiconductor industry. Nowadays the performance of the integrated circuits is highly depends on the interconnect designs as the interconnect delay has exceeded the gate switching delay. On the other hand, the total number of Input/Output (IO) and Power/Ground (PG) terminals keep increasing and drive the requirements of rapid advance in integration density and performance. As the required wiring and bump pitches on the substrate continue to shrink, passive interposer using silicon substrate has emerged as a reliable solution to connect chips using shorter and denser interconnects [ZCL09].

Coplanar waveguide (CPW) transmission lines are commonly realized in the conventional CMOS processes for global power and signal routing. In order to evaluate the electrical behaviors of the broad-band transmission line for CPW structure on lossy silicon substrate, a variety of numerical techniques and modeling methodologies have been proposed. In [Has71], three different modes of wave propagations on metal-isolator-semiconductor (MIS) structure, including "skin-effect", "slow-wave" and "quasi-TEM" modes, are proposed according to the comparison between operation frequency and silicon relaxation frequency. In [AMH99], a quasi-static approach in spectral domain for layered isolator-semiconductor sub-

strate is proposed and is extended to include non-perfect metallic conductors in [PMM06]. However, a simple distributed frequency-dependent resistance, inductance, conductance and capacitance (RLGC) transmission line model is always desired for fast spice-compatible time domain simulations. [ZHT00,MOD98,KNR08] have proposed RLGC equivalent circuit models for CPW with silicon substrate but the derivations of the circuit topology are not rigorous and fitted empirical formulas were used for the circuit elements.

In this chapter, we proposed a circuit topology for MIS CPW structures based on partial element equivalent circuit (PEEC) [Rue72, WKC92]. The per-unit-length shunt admittance and series impedance are calculated separately due to the slow-wave transmission line property, as demonstrated in [TPH07,SRG09]. While the effective series impedance is mostly dominate by the skin loss of the conductor rather than the loss from eddy current in silicon, the calculations of inductance and resistance could simply use the similar methods mentioned in [WKC92]. On the other hand, the shunt capacitance and conductance in inhomogeneous regions are rigorously derived through PEEC method combined with complex image theory. The proposed method assumes the ratio between trace space and width should not be too small, which is valid in most conventional CMOS processes due to design rules restrictions, and the error induced by this method is also analyzed for different geometrical constrains.

5.2 CPW Equivalent Circuit Model: Shunt Admittance

Fig. 5.1 shows a typical geometry of coplanar waveguide on the silicon interposer. Through quasi-static analysis, the electrical properties of the CPW structure can be accurately determined by per-unit-length impedance and admittance parameters, as demonstrated in [ZTW01].

Although the total shunt admittance for on-chip CPW is highly frequency-

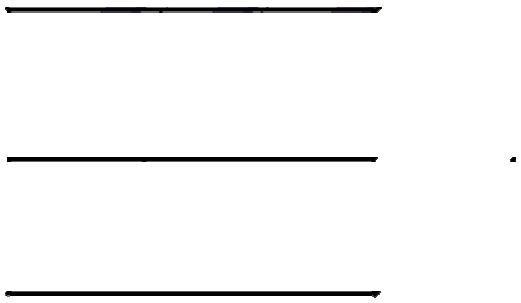

Figure 5.1: Coplanar waveguide on silicon interposer.

dependent, it still can be modeled for a wide frequency range with frequency-independent capacitance and conductance if an appropriate circuit topology is utilized.

From the integral solution of Maxwells equations, the relationship between the charge density and the potential is given as [RB73]

$$\Phi(\boldsymbol{r}) = \int G_c(\boldsymbol{r}, \boldsymbol{r'}) q(\boldsymbol{r'}) ds', \tag{5.1}$$

where q is the charge density on the conductor surfaces, and G_c is the Greens function. $\boldsymbol{r}$ and $\boldsymbol{r'}$ are the position vectors of the source and observation points. If the charge uniformly distributes on the surface of the conductors, (5.1) becomes

$$\Phi(\boldsymbol{r}) = \sum_{i=1,2} \frac{Q_i}{S_i} \int G_c(\boldsymbol{r}, \boldsymbol{r'}) ds', \tag{5.2}$$

where S_1 and S_2 represent the total surface areas for signal and ground conductors. After applying Galerkins approach to match the potential for signal and ground by averaging (5.2) over their surfaces, we can obtain

$$\Phi_{j=1,2} = \sum_{i=1,2} Q_i p_{s_{ij}}, \tag{5.3}$$

where

$$p_{s_{ij}} = \frac{1}{S_i S_j} \int_{S_i} \int_{S_j} G_c(\boldsymbol{r}, \boldsymbol{r'}) ds' ds. \tag{5.4}$$

For on-silicon CPW structure, inhomogeneous SiO2-Silicon media are usually involved. The Greens function for this layered media can be derived from the free space Greens function using the complex image theory. Note both signal and ground conductors are in the SiO2 layer, source and observation points are in the same media. Thus, the Greens function for CPW structure in SiO2-Silicon media can be written as

$$G_c(\boldsymbol{r}, \boldsymbol{r'}) = \frac{1}{4\pi} \frac{1}{|\boldsymbol{r} - \boldsymbol{r'}|} \frac{1}{\epsilon_{SiO_2}} + \frac{1}{4\pi} \frac{1}{|\boldsymbol{r} - \boldsymbol{r'_{im}}|} \frac{1}{\epsilon_{SiO_2}} \frac{\epsilon_{SiO_2} - \epsilon^c_{Si}}{(\epsilon_{SiO_2} + \epsilon^c_{Si})}, \tag{5.5}$$

where r'_{im} is the position vectors of the image of observation points. ϵ_{SiO_2} is the permittivity of SiO_2 and

$$\epsilon^c_{Si} = \epsilon_{Si} + \frac{\sigma_{Si}}{j2\pi f}, \tag{5.6}$$

where ϵ_{Si} and σ_{Si} are the permittivity and conductivity of the silicon, respectively.

In order to develop an intuitive physics-based circuit model, we assume charge distributes uniformly on the 2-D surface of signal and ground conductors. By applying (5.5), the $\boldsymbol{p_s}$ matrix defined in (5.4) becomes

$$\boldsymbol{p_s} = \begin{bmatrix} a_{ss} \frac{1}{\epsilon_{SiO_2}} + a_{ss'} \frac{1}{\epsilon_{SiO_2}} \frac{\epsilon_{SiO_2} - \epsilon^c_{Si}}{(\epsilon_{SiO_2} + \epsilon^c_{Si})} & a_{sg} \frac{1}{\epsilon_{SiO_2}} + a_{sg'} \frac{1}{\epsilon_{SiO_2}} \frac{\epsilon_{SiO_2} - \epsilon^c_{Si}}{(\epsilon_{SiO_2} + \epsilon^c_{Si})} \\ a_{sg} \frac{1}{\epsilon_{SiO_2}} + a_{sg'} \frac{1}{\epsilon_{SiO_2}} \frac{\epsilon_{SiO_2} - \epsilon^c_{Si}}{(\epsilon_{SiO_2} + \epsilon^c_{Si})} & a_{gg} \frac{1}{\epsilon_{SiO_2}} + a_{gg'} \frac{1}{\epsilon_{SiO_2}} \frac{\epsilon_{SiO_2} - \epsilon^c_{Si}}{(\epsilon_{SiO_2} + \epsilon^c_{Si})} \end{bmatrix}, \tag{5.7}$$

where a_{ss}, a_{sg}, and a_{gg} are purely related to the geometry of signal and ground conductors in the homogeneous air, as shown in Fig. 5.2(a). $a_{ss'}$, $a_{sg'}$, and $a_{gg'}$ can be calculated using (5.4) with free space Green function in air between the original conductor and its image, as shown in Fig. 5.2(b), (c) and (d), respectively. For 2-D rectangular conductors in homogeneous media, all the coefficients in (5.7) can be evaluated analytically [RB73, ZK02].

The shunt admittance between the signal and ground conductors in SiO_2-

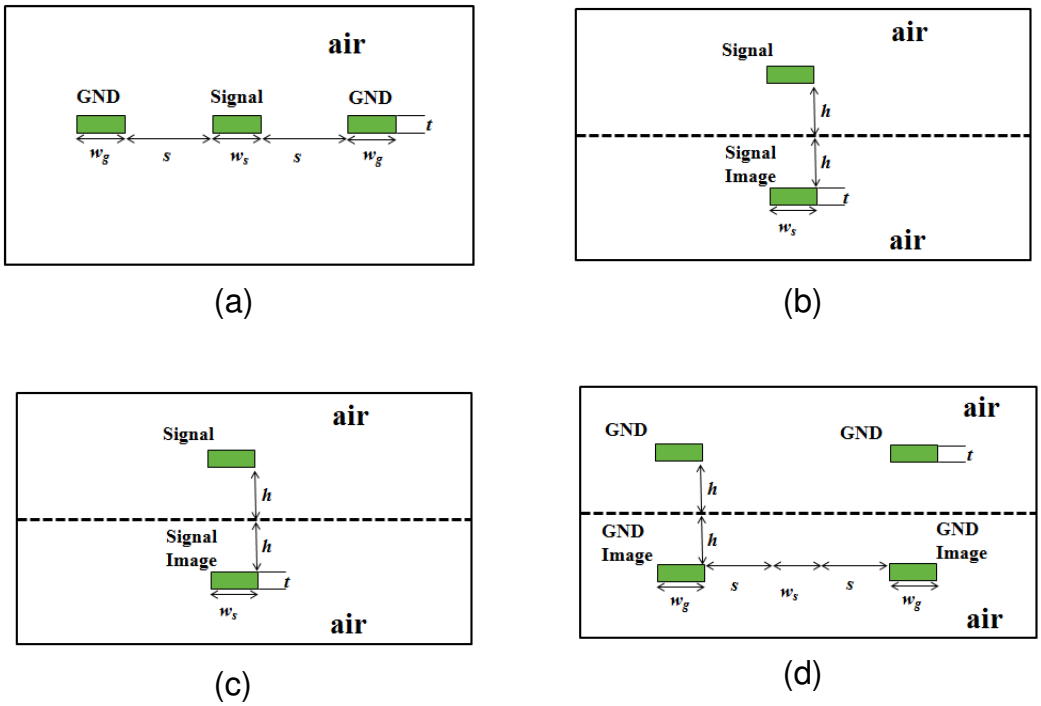

Figure 5.2: Geometries to calculate the coefficients in the $\boldsymbol{p_s}$ matrix.

Silicon media can be further obtained through (5.7), as

$$Y_{cpw} = \frac{1}{(a_{ss} + a_{gg} - 2a_{sg})\frac{1}{\epsilon_{SiO_2}} + (a_{ss'} + a_{gg'} - 2a_{sg'})\frac{1}{\epsilon_{SiO_2}}\frac{\epsilon_{SiO_2} - \epsilon^c_{Si}}{(\epsilon_{SiO_2} + \epsilon^c_{Si})}}. \tag{5.8}$$

By exploiting the pair capacitance between the two conductors from the four configurations in Fig. 5.2, we can write the equivalent admittance in (5.8) as

$$Y_{cpw} = \frac{1}{b_1\frac{1}{\epsilon_{SiO_2}} + b_2\frac{1}{\epsilon_{SiO_2} + \epsilon^c_{Si}}}, \tag{5.9}$$

where

$$b_1 = b_{(a)} + \frac{b_{(b)}}{2} + \frac{b_{(d)}}{2} - b_{(c)}, \tag{5.10}$$

$$b_2 = b_{(c)} - b_{(b)} - b_{(d)}, \tag{5.11}$$

$$b_{(i)} = 1/C_{(i)} \quad i = a, b, c, d \tag{5.12}$$

and $C_{(i)}$ is the pair capacitance between two groups of conductors in homogeneous media in Fig. 5.2(i).

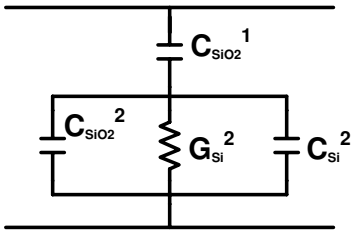

Figure 5.3: Equivalent circuit model for the shunt admittance between signal and ground conductors of CPW.

Finally, the equivalent circuit model for the shunt admittance between signal and ground CPW on silicon can be derived based on (5.9) and is shown in Fig. 5.3 with

$$C^1_{SiO_2} = \epsilon_{SiO_2}/b_1, \tag{5.13}$$

$$C^2_{SiO_2} = \epsilon_{SiO_2}/b_2, \tag{5.14}$$

$$C^2_{Si} = \epsilon_{Si}/b_2, \tag{5.15}$$

$$G^2_{Si} = \sigma_{Si}/b_2. \tag{5.16}$$

The elements in the equivalent circuit model are directly related to the physical geometry. For example, in the design of on-chip interconnects, smaller conductance is usually desired to reduce loss for signal transition. In order to reduce G^2_{Si}, the capacitance from the configuration in Fig. 5.2(c) needs to be increased while the capacitance of the geometry in Fig. 5.2(b) and (d) needs to be decreased. Thus, to reduce the spacing between signal and ground conductors and to increase the vertical distance between conductors and silicon substrate helps to reduce loss caused by shunt conductance.

5.3 CPW Equivalent Circuit Model: Series Impedance

The existence of lossy silicon can affect the current distribution of the on-chip CPW structure due to the eddy currents in the substrate induced by the time

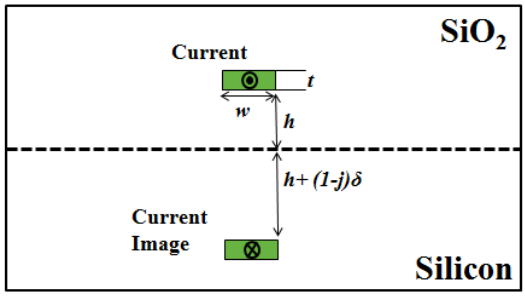

Figure 5.4: Complex image of line current source in SiO_2-Silicon media. The current flows in the opposite direction in the image lines.

varying magnetic field. For a line current above a semi-infinite silicon substrate, the solution of the vector magnetic potential A_z can be expressed in term of the 2-D Green's function as

$$A_z(\boldsymbol{r}) = \int G_l(\boldsymbol{r}, \boldsymbol{r'}) J_z(\boldsymbol{r'}) ds', \tag{5.17}$$

where J_z is the z-direction current density vertical to the substrate and G_l is the Green's function. $\boldsymbol{r}$ and $\boldsymbol{r'}$ are the position vectors of the source and observation points.

By Applying the complex image theory to the Greens function in integral form, the Greens function in SiO_2 region can be expressed as the vector magnetic potential due to original line current sources and its complex image, as

$$G_l(\boldsymbol{r}, \boldsymbol{r'}) = G_{l_{air}}(\boldsymbol{r}, \boldsymbol{r'}) + G_{l_{Si}}(\boldsymbol{r}, \boldsymbol{r'}), \tag{5.18}$$

where

$$G_{l_{air}}(\boldsymbol{r}, \boldsymbol{r'}) = \frac{\mu_0}{4\pi} \frac{1}{|\boldsymbol{r} - \boldsymbol{r'}|} \tag{5.19}$$

$$G_{l_{Si}}(\boldsymbol{r}, \boldsymbol{r'}) = \frac{\mu_0}{4\pi} \frac{1}{|\boldsymbol{r} - \boldsymbol{r'_{im}}|}, \tag{5.20}$$

and $\boldsymbol{r'_{im}}$ is the position vectors of the complex image of observation points, as shown in Fig. 5.4. The mirror distance between the original and complex image

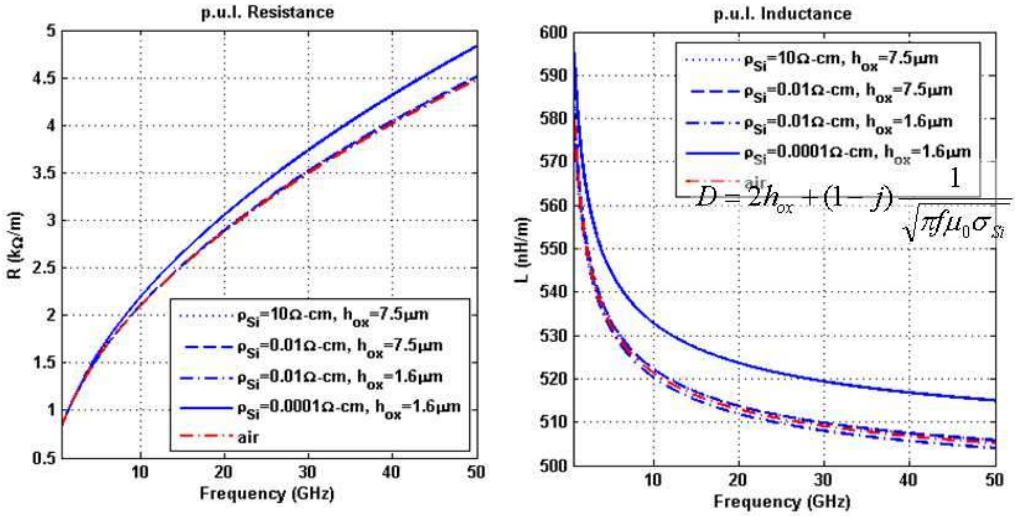

Figure 5.5: Comparisons of the series impedance for on-chip CPW structures between different silicon conductivity.

points is defined as

$$d = 2h + (1 - j)\delta, \tag{5.21}$$

where

$$\delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma_{Si}}}, \tag{5.22}$$

The per-unit-length partial inductance of the CPW structures on silicon substrate then can be obtained as

$$L_{mn} = \frac{1}{A_m A_n} \int_{A_m} \int_{A_n} G_l(\boldsymbol{r}, \boldsymbol{r'}) dA_m dA_n, \tag{5.23}$$

where A_m and A_n are the cross-sectional areas of two filaments, in which current distribution varies very little. If $\delta \gg h$, the series impedance for CPW structure in SiO_2-Silicon media can be simply calculated as the impedance of the same conductors in homogeneous media, as shown in Fig. 5.5. In todays conventional CMOS process, the conductivity of silicon is usually less than $100 S/m$. At the frequency of interests up to 50 GHz, δ is still much larger than the dimensions of the interconnects. Thus, the eddy current in the silicon substrate generally has little effects on the series impedance in the equivalent circuit model.

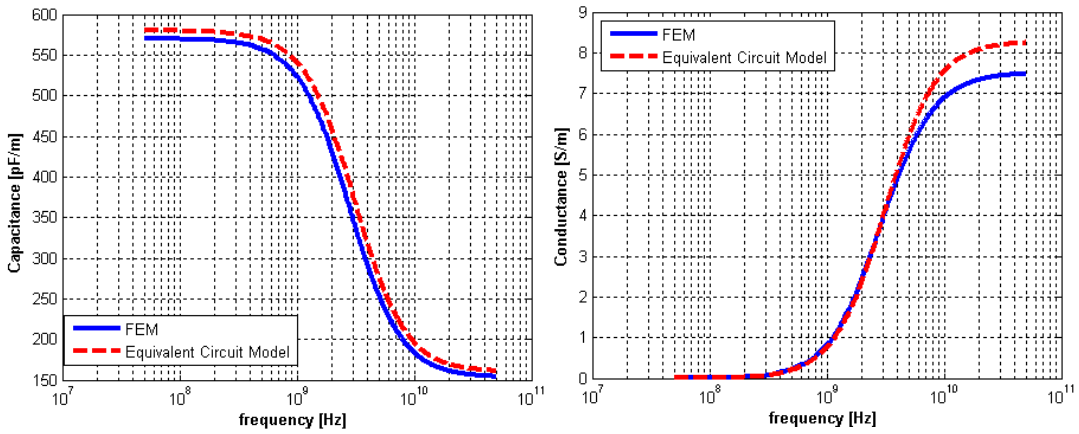

Figure 5.6: Shunt admittance obtained from our model is compared with simulated results using FEM: (left) capacitance (right) conductance.

5.4 Model Validation via Simulation

To validate the proposed circuit model for the on-chip CPW in Fig. 5.1, the admittance from the circuit model are compared with that obtained from full-wave simulator based on FEM, as shown in Fig. 5.6. The resistivity of silicon is $10\Omega - cm$. In the derivation of the equivalent circuit model for shunt admittance of on-chip CPW structure, we assumed the charge uniformly distribute on the surfaces of the conductors and ignored the proximity effect. Theoretically, this assumption will generate errors when calculate the elements in $\boldsymbol{p_s}$ matrix. As generally the width of on-chip conductors is greater than its thickness, the accuracy of the proposed model can be is directly related to the ratio of trace spacing and their widths, as

$$r = \frac{s}{max(w_s, w_g)}. \tag{5.24}$$

Fig. 5.7 shows the relative deviations between the admittance from the proposed circuit model and that from FEM. The deviation is defined as the percentage of the average difference of the curves from two different methods in the frequency range from 50MHz to 50GHz. The deviation is plotted versus the ratio

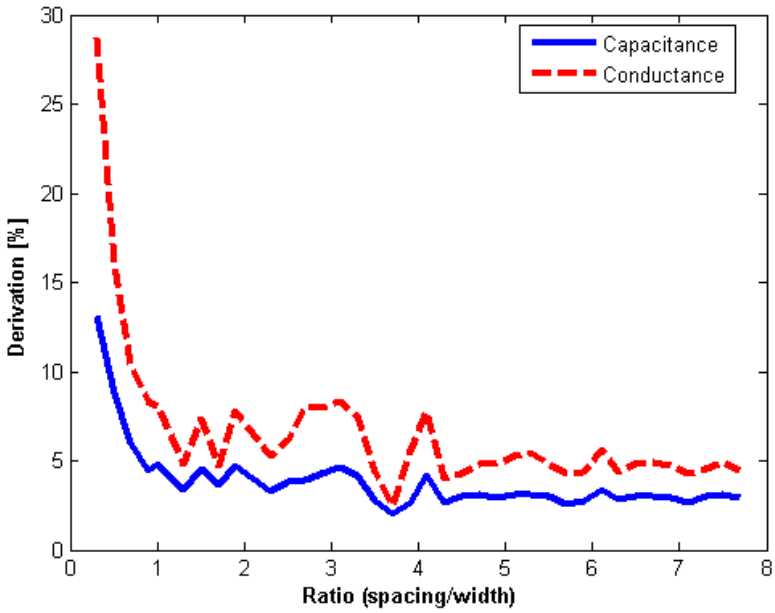

Figure 5.7: Derivations of capacitance and conductance for on-chip CPW using FEM and the model proposed. The x-axis is the ratio of trace spacing to trace width. The y-axis is the average deviation in percentage from 50 MHz to 50 GHz.

r defined in (5.24). When the ratio of trace spacing and width is smaller than 1, the distributions charges cannot be considered as uniform on the surface of the each conductor, thus, the proposed circuit model introduces large errors. However, when the ratio is larger than 1, the deviations of both the capacitance and conductance from circuit model and FEM are only around 5%.

5.5 Conclusion

A new equivalent circuit is proposed in this chapter to model the 2D lossy redistribution layer (RDL) coplanar waveguide (CPW) on the silicon interposer. The frequency-dependent shunt admittance of the CPW is derived rigorously based on the partial equivalent element circuit (PEEC) method considering the metal-isolator-semiconductor (MIS) layered media. The silicon effects on series

impedance are also considered by employing the modified green functions with complex image theory. The presented equivalent circuit model shows good correlations with full-wave EM simulations up to 25GHz.

CHAPTER 6

Power-Bandwidth Trade-off on TSV Array and TSV-RDL Junction Design Challenges on Silicon Interposer

6.1 Introduction

As traditional CMOS scaling pace gradually slows down, three-dimensional (3D) integration offers another dimension of scaling by means of stacking functional blocks vertically and providing high integration density, fast signal transmission, low power consumption, and heterogeneous integration opportunities in the "More-than-Moore" era [BSK01], as shown in Fig. 6.1. Through-silicon-via (TSV) has been well regarded as a key component connecting chips vertically with shortened electrical delay and providing extremely dense I/O connections. Owing to its short distance, if designed properly, TSV could provide superior performances in terms of signal attenuation, rail collapse, and crosstalk [XBZ10], as compared to traditional inter-chip channels. On one hand, the improved signal integrity can potentially provide thousands of multi-Gb/s I/O and support for tens of Tb/s data bandwidth between local chips without power-hungry equalization schemes [LLF12, SCA10]. On the other hand, reduced swing approaches offer a path to further reduce the power for TSV channels [LLF12]. However, detailed analysis considering various TSV array patterns and geometries with different signaling and termination techniques is desperately required in order to find a balance between power and bandwidth under different design constraints.

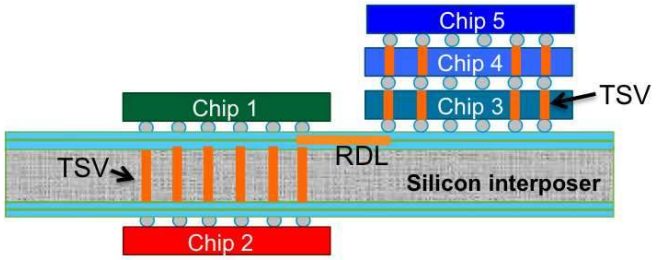

Figure 6.1: Concept of 3D IC using TSV and silicon interposer

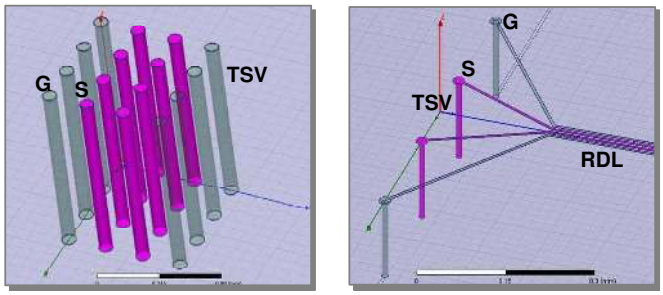

Figure 6.2: (a) 4×4 TSV array and (b) an example of junction structure between TSVs and interposer RDL traces

Beyond the TSV itself, horizontal interconnects such as redistribution layer (RDL) traces, are also essential to 3D multi-chip integration, as shown in Fig. 6.1. Though various works have been proposed regarding signal integrity on RDL traces [YKL09,KPC11] and its application on high-speed serial links [DLR12], the impact of TSV-RDL junction structure design has not been sufficiently investigated, especially when a large pitch difference exits between the TSV and RDL trace.

In this chapter, both frquency-domain and time-domain signal integrity on TSV arrays are first evaluated with high-speed signaling. 3D electromagnetic field solver [ANS] is employed to model the TSV array and TSV-RDL junction accurately, as shown in Fig. 6.2. By combining the TSV array model with selected I/O circuits, the overall transient behavior is further analyzed through SPICE simulation. In Section 6.3, reduced-swing signaling is then applied to the TSV array

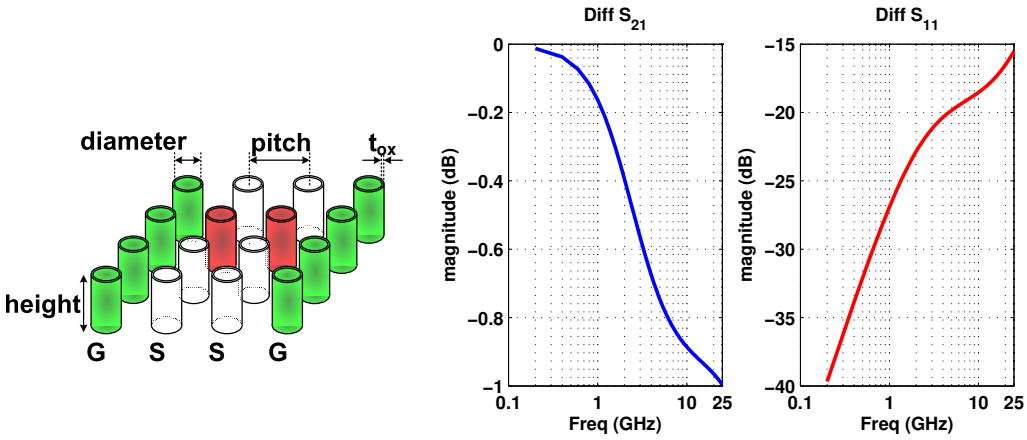

Figure 6.3: An array of cylindrical Cu-filled TSVs and the differential S_{21} and S_{11} on one of the differential TSV pair (marked in red).

for power and bandwidth trade-off analysis with different I/O and termination techniques. Simulation shows unterminated single-ended reduced-swing signaling can achieve the best power efficiency, while terminated single-ended signaling can provide the maximum bandwidth. In Section 6.4, critical design challenges for the junction structure between TSVs and RDL traces are also analyzed. Simulation shows that at $20GHz$, the fanout-like junction structure could cause about $10dB$ S_{11} degradation when changing TSV pitch from $50\mu m$ to $200\mu m$ and even contribute more insertion loss (S_{21}) than the TSV itself.

6.2 Signal Integrity Analysis on TSV Array

An example of a 4×4 TSV array is shown in Fig. 6.3. The characteristics of TSV are dependent on its geometrical parameters such as TSV diameter, height, pitch, oxide layer thickness, and electrical parameters such as metal conductivity and the silicon substrate resistivity. The TSV array is arranged by differential G-S-S-G pattern with $30\mu m$ pitch and each TSV in the array is identical with $10\mu m$ diameter, $150\mu m$ height, and $0.5\mu m$ oxide layer thickness. A $1dB$ differential insertion loss is observed at $25GHz$ when the silicon substrate has 10Ω-cm

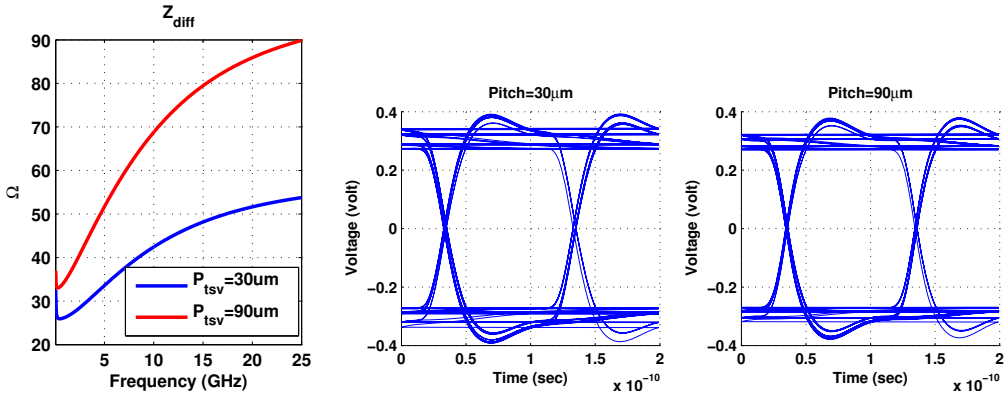

Figure 6.4: The differential impedance (Z_{diff}) and eye diagram comparison between $30\mu m$ and $90\mu m$ TSV array pitch. The TSV diameter is $10\mu m$ and height is $150\mu m$.

resistivity, as shown in Fig. 6.3.

Fig. 6.4 shows the differential characteristic impedance (Z_{diff}) and eye diagram comparison when changing the TSV array pitch from $30\mu m$ to $90\mu m$. As shown in Fig. 6.4, the Z_{diff} of TSV is so frequency dependent that it is difficult to control the impedance over a wide frequency range. By increasing the TSV pitch from $30\mu m$ to $90\mu m$, Z_{diff} increases from 40Ω to 70Ω at $10GHz$ due to the increase of loop inductance and the decrease of admittance between the signal and ground TSV. Fig. 6.5 shows another impedance comparison when changing the center-to-center distance between differential signal TSVs from $30\mu m$ to $90\mu m$ while keeping the signal to ground distance fixed at $30\mu m$. Similarly, Z_{diff} increases from 40Ω to 50Ω at $10GHz$. Though the impedance difference is obvious, especially for the first comparison between $30\mu m$ and $90\mu m$ TSV array pitch, the transient behavior of signal integrity, however, is not much different at $10GHz$. An industrial transistor-level driver model is adopted with 10Gbps PRBS differential signals (V_{diff} is $600mV$ with V_h is $1.0V$ and V_l is $0.7V$) to simulate the eye diagram using SPICE. Both transmitter and receiver have 100Ω equivalent differential impedance. Note only one pair of signals are switching in the array so

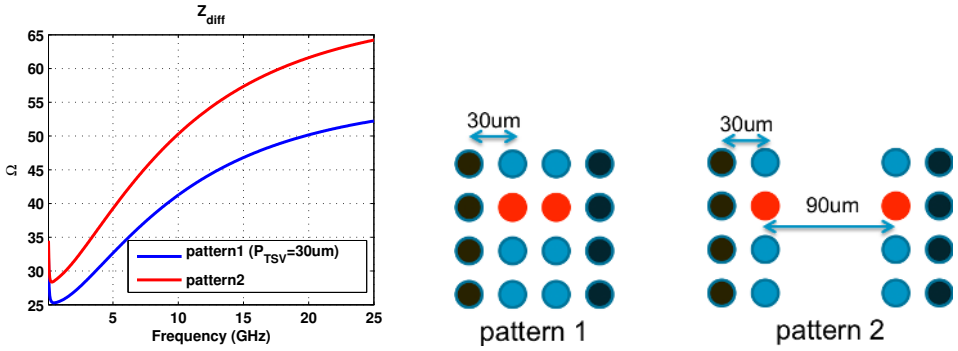

Figure 6.5: The differential impedance (Z_{diff}) comparison when changing the center-to-center distance between differential signal TSVs from $30\mu m$ to $90\mu m$ while keeping the signal to ground distance fixed at $30\mu m$.

that there is no crosstalk noise.

From Fig. 6.4, it can be shown that both cases can achieve around $550mV$ eye opening, while the $30\mu m$ pitch case has slightly larger timing jitter ($3.11ps$) than the $90\mu m$ pitch case ($2.25ps$) because it causes more impedance discontinuity. The TSV array pitch impact on signal reflection caused by impedance mismatch is not significant due to the electrically short length of TSV.

Other than reflection, crosstalk also plays an important role in both traditional parallel buses and advanced multiple serial links. Fig. 6.6 shows the signal-to-crosstalk-ratio (SCR) comparison on the $30\mu m$ pitch 4×4 TSV array. As shown in Fig. 6.6, when applying single-ended signaling, the victim in the TSV array has around $12dB$ SCR at $20GHz$ with multiple aggressor switching at the same time. When applying differential signaling, SCR can be further improved to $21dB$ at $20GHz$. Note that the impact of crosstalk increases when frequency increases, but the SCR curve has a slope transition at around $5GHz$ due to the transition from the slow-wave mode to the quasi-TEM mode.

Fig. 6.7 shows the maximum achievable data rate per TSV in the 4×4 TSV array versus various TSV pitches and diameters. The maximum data rate is determined by the eye diagram constraints including at least 300mV eye open and

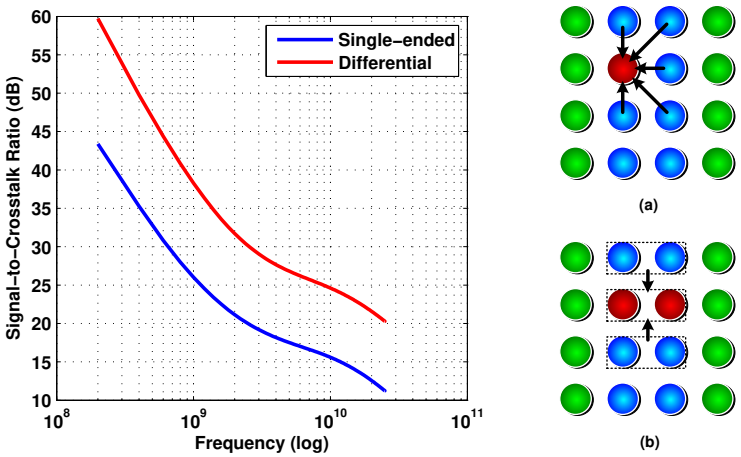

Figure 6.6: Signal-to-Crosstalk-Ratio (SCR) comparison on the 4×4 TSV array when considering (a) single-ended signaling and (b) differential signaling.

10%UI jitter tolerance. As shown in Fig. 6.7, the maximum data rate is limited by the 2pF loading capacitance and the coupling noise from adjacent TSV pairs since increasing TSV diameter does not improve the data rate much. Surprisingly, increasing TSV pitch actually decreases the achievable data rate due to the increased crosstalk. This is because although capacitive coupling decreases as pitch size increases, the mutual inductance between signal TSVs actually increases. When increasing the pitch size, not only signal TSVs are further away from each other, ground TSVs are also further away from signal TSVs, which leads to more magnetic couplings between signal TSVs.

6.3 TSV Power and Bandwidth Trade-off Analysis

Since TSV provides superior performances in terms of signal attenuation, ISI and crosstalk due to its short electrical length as discussed in precious section, reduced swing approaches can further be applied to TSV channels for power reduction purpose [LLF12]. A balance between power and bandwidth, however, has to be found under various design objectives and different signaling and termination

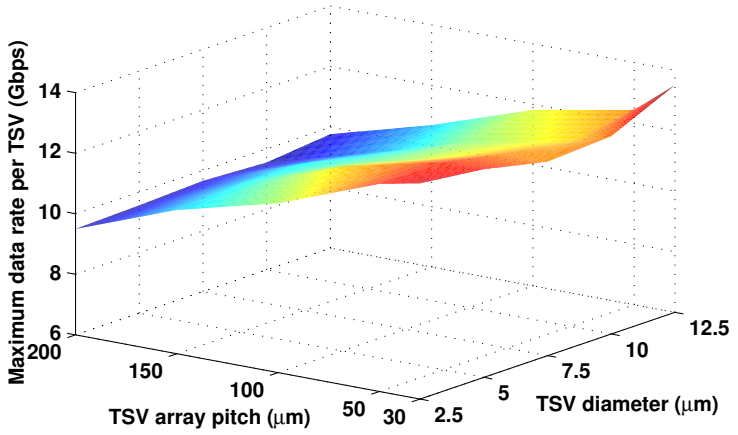

Figure 6.7: Maximum achievable data rate per TSV in a 4×4 TSV array versus various TSV pitches and diameters.

techniques.

Fig. 6.8 shows a couple of output driver schemes considered in our trade-off analysis, including traditional voltage-mode push-pull single-ended signaling and differential CML-type current-mode signaling. Furthermore, two termination techniques are considered for the push-pull driver: one has matched resistive termination while the other has purely capacitive loading. Increased signal reflection is expected for capacitive loading when data rate increases, however, power consumption can be significantly improved compared to matched resistive termination. For push-pull drivers, the output swing is determined by the supply voltage, thus an extra voltage regulator is required to reduce the signal swing. For CML-like drivers, the output swing can be easily controlled by adjusting the current source but it requires more driving current to develop the same voltage swing since the current source sees a parallel connection of the two termination resistors.

A 4×4 TSV array with $10\mu m$ diameter, $150\mu m$ height and $30\mu m$ pitch, as shown in Fig. 6.6, is first considered in the trade-off analysis. All three types of

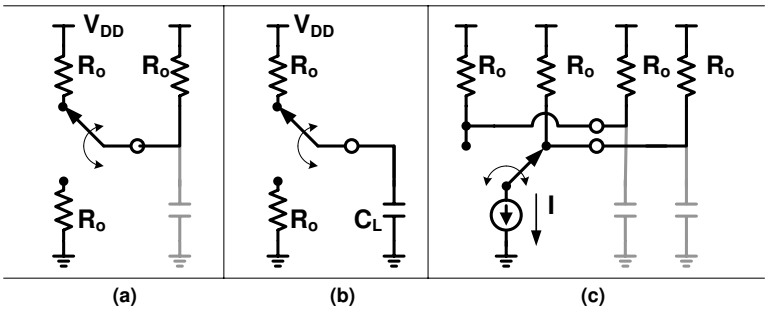

Figure 6.8: Different types of drivers and terminations: (a) push-pull single-ended with resistive termination, (b) push-pull single-ended with pure capacitive loading, and (c) differential CML driver. $R_0 = 40\Omega$ and $C = 2pF$.

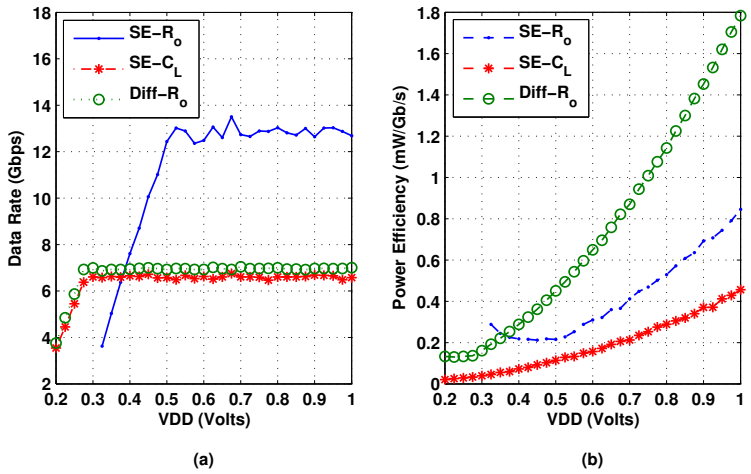

Figure 6.9: (a) Maximum achievable data rate per TSV versus input swing voltages and (b) corresponding power efficiency versus input swing voltages for various output drivers and terminations shown in Fig. 6.8.

drivers are combined with the TSV array electrical model and simulated in SPICE to find out the maximum data rate per TSV for each case under different reduced swing. The maximum data rate is determined by the eye diagram constraints including at least a 200mV eye opening [LLF12] and 10%UI jitter tolerance. As shown in Fig. 6.9, for all the driver schemes, the swing can be lowered down by almost half without hurting the maximum achievable bandwidth and significantly

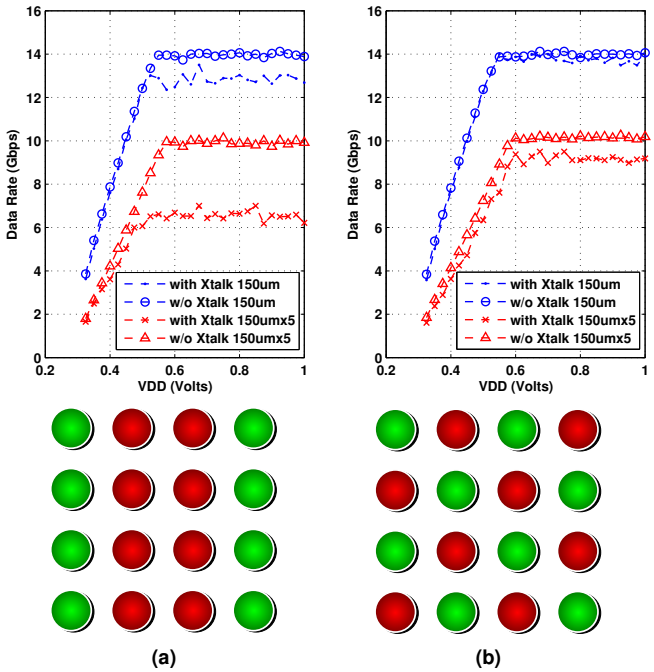

Figure 6.10: Maximum achievable data rate per TSV using single-ended signaling versus driver input swing with different TSV array patterns: (a) G-S-S-G column-by-column and (b) distributed signal and ground.

improves the power efficiency. This is due to the extra margin provided by the short TSV channels. Moreover, to achieve the best power efficiency, the unterminated single-ended case should be applied, while the terminated single-ended case can provide the maximum bandwidth. If extremely low swing is targeted, the terminated single-ended case is no longer suitable and only the unterminated single-ended or differential signaling can be used. Note that differential signaling is not as dominant as in traditional chip-to-chip channels because many of its advantages, such as robustness against crosstalk and better discontinuity tolerance in the return path, are not taken due to the short length of TSV. However, a pair of signal TSV is required for each differential channel, which significantly limits the total bandwidth.

If we intentionally increase the length of the TSV array by 5 times, similar

to the stacked-dies case as shown in Fig. 6.1 but assuming an extreme situation that each die has the same substrate thickness as silicon interposer, the signal integrity becomes much worse, as shown in Fig. 6.10(a) with terminated single-ended signaling. Without considering crosstalk, the original maximum achievable data rate per TSV is around $14Gbps$. By increasing the length by 5 times, the achievable data rate drops down to $10Gbps$. This is mainly due to channel loss and reflection caused by impedance mismatch. If all the aggressors start switching at the same time, crosstalk noise is induced and the achievable data rate drops down further from $10Gbps$ to only $6Gbps$. This shows the signal integrity is still playing an important role on TSV channels but is limited by the short length of the TSV, because if the TSV length is as short as $150\mu m$, crosstalk only affects the data rate slightly. A similar simulation is performed but this time with the TSV array having distributed signal and ground TSV arrangements, as shown in Fig. 6.10(b). It can be shown that the crosstalk impact can be reduced significantly for the single-ended case by interleaving signal and ground TSV in the array.

6.4 Design Challenges on Junction Structures between TSV and RDL Trace

Beyond the TSV itself, horizontal interconnects such as redistribution layer (RDL) traces, are also essential to 3D multi-chip integration, as shown in Fig. 6.1 between chip2 and chip3. Since the RDL trace length is usually much longer than the length of TSV, severe signal attenuation, ISI, and coupling noise are expected and have been studied extensively [YKL09, KPC11, DLR12]. However, the importance of the junction structure between TSV and RDL interconnect has not been sufficiently investigated, especially if we consider the fact that TSV pitch is usually much larger than RDL traces due to process limitations and mechanical issues. Specifically, a fanout-like junction structure is required for a successful in-

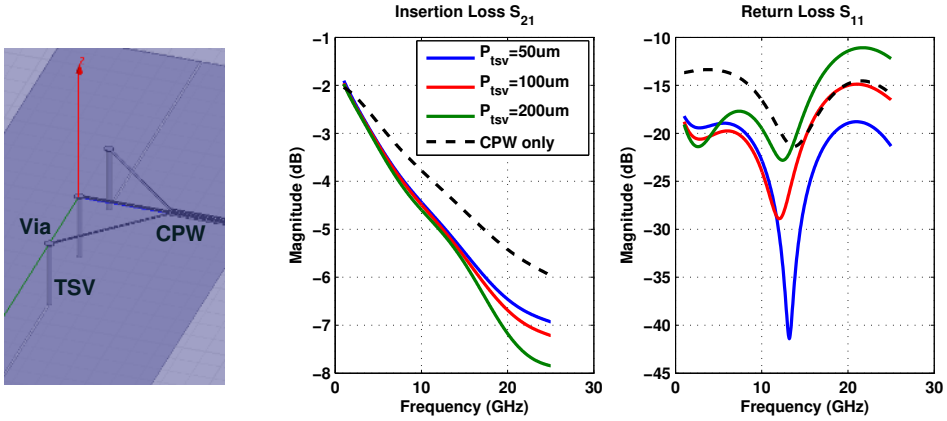

Figure 6.11: Fanout-like junction structure between TSV and RDL and the full link S_{21} and S_{11} comparison for different TSV pitches.

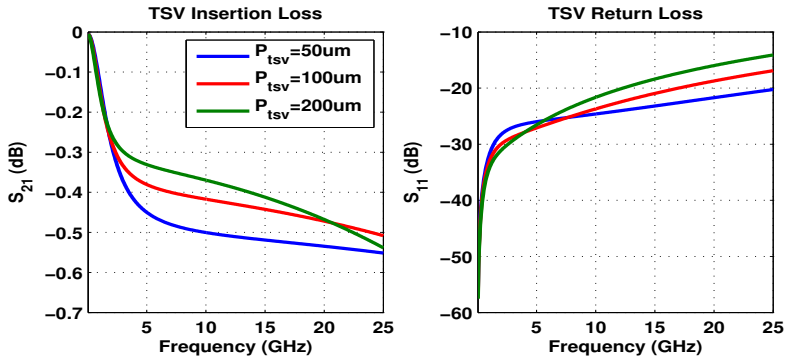

Figure 6.12: G-S-G TSV S_{21} and S_{11} comparison for different TSV pitches.

terconnection between TSV and RDL. Fig. 6.11 shows an example of a fanout-like junction structure between TSV and RDL and its full link (TSV+RDL) S_{21} and S_{11} comparison for different TSV pitches. The RDL coplanar waveguide (CPW) has a $5\mu m$ metal width, $5\mu m$ metal spacing, and $5mm$ length. The TSV diameter is $10\mu m$ and the height is $150\mu m$. Both TSV and CPW are arranged in the ground-signal-ground (G-S-G) pattern. From Fig. 6.11, it can be shown that the insertion of a $50\mu m$ pitch G-S-G TSV introduces an extra insertion loss into the system (around $1dB$ at $20GHz$). However, it actually improves the return loss by $5dB$ at $20GHz$. This is because the Z_0 of the $50\mu m$ pitch TSV happens to be between source impedance (50Ω) and the Z_0 of $5\mu m/5\mu m$ CPW.

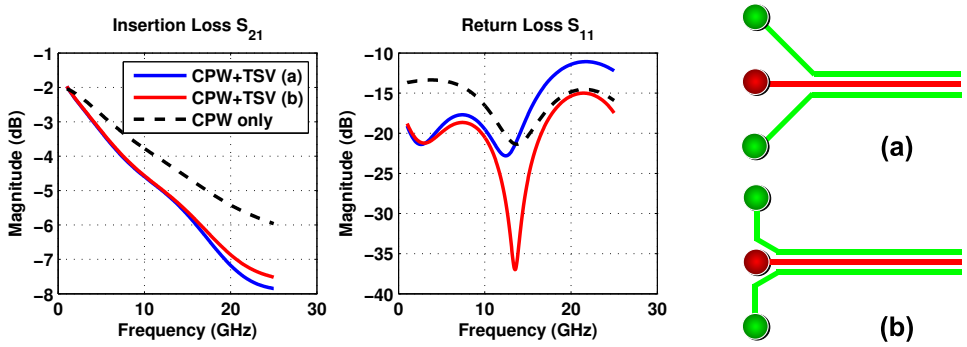

Figure 6.13: Full link S_{21} and S_{11} comparison between two different junction structure between TSV and RDL.

If we keep increasing the pitch of TSV, as shown in Fig. 6.11, not only does the insertion loss keep increasing, the return loss also rises up and has around $10dB$ S_{11} degradation at $20GHz$ when changing TSV pitch from $50\mu m$ to $200\mu m$. Part of the reason is because of the impedance change on the G-S-G TSV itself with different pitch, as shown in Fig. 6.12. The discontinuity introduced by the junction structure between TSV and CPW also causes comparably extra return loss. Moreover, from Fig. 6.12, it can be shown that the TSV-only S_{21} decreases only by $0.1dB$ at $20GHz$ when increasing TSV pitch from $50\mu m$ to $200\mu m$. Compared to Fig. 6.11, it is evident that more than $1dB$ extra S_{21} loss is introduced by the fanout-like junction structure when increasing TSV pitch from $50\mu m$ to $200\mu m$. The reason is that since CPW is used instead of popular microstrip or stripline due to the lack of ground plane in silicon interposer, signal and ground traces should be designed close to each other in order to provide a well-defined return path. However, it becomes a design challenge when connecting between TSV and CPW with a large pitch difference. A possible improvement for the TSV-RDL junction can be found in Fig. 6.13. By minimizing the length of the fanout-like structure at the TSV-RDL junction, the S_{11} can be improved by as much as $5dB$ at $20GHz$. However, the S_{21} improvement is still limited since extra loss is still introduced by the isolated extra ground segments, which is comparable with the

loss introduced by TSV.

6.5 Conclusions

In this chapter, we first evalute the signal integrity issues on TSV array based on 3D EM field solver and SPICE simulation. Power and bandwidth trade-off analysis is further applied to consider reduced-swing signaling with different I/O and termination techniques. In order to achieve the best power efficiency, unterminated single-ended reduced-swing signaling should be applied, while terminated single-ended signaling can provide the maximum bandwidth. Critical design challenges for the junction structure between TSVs and RDL traces are also revealed and simulation shows that at $20GHz$, the fanout-like junction structure contributes an even larger performance impact than TSV itself.

CHAPTER 7

Conclusion

To increase the performance of today's integrated digital systems along with the trend of computation and communication efficiency, the I/O interconnection bandwidth must be maximized by either increasing the data rate per channel or increasing the bus width or the number of available channels. Serial links have successfully evolved and achieved the bit-rate of several tens of Gb/s per channel by applying new generations of IC process and advanced circuit techniques. However, as process technologies further scale down, severe process variations significantly impact the performance of high bit-rate serial links and makes todays circuit designs have to be optimized not only for nominal performance but also for a reasonable yield. On the other hand, 3D IC provides a much dense I/O interconnection with higher performance and lower power consumption than conventional 2D integration by stacking multiple dies vertically.

In Chapter 2, a formula-based jitter and noise models considering the combined effect of ISI, crosstalk, and pre-emphasis filter is proposed. By utilizing these formula-based models, we then develop algorithms to directly find out the input patterns for worst-case jitter and worst-case amplitude noise through pseudo-Boolean optimization and mathematical programming. In addition, heuristic algorithm is proposed to further reduce runtime. In Chapter 3, we formulate a co-optimization problem for digitally tuned analog circuits to optimize the parametric yield, subject to power and area constraints. A general optimization framework combing the branch-and-bound algorithm and gradient ascent method is

proposed. We demonstrate our framework with two examples in high-speed serial link, the transmitter design and the PLL design. Results show that the proposed optimization framework improves the yield significantly compared to the design heuristic from analog designers perspective.

In Chapter 4, a pair-based model for multi-port TSV network is proposed by decomposing the network into a number of TSV pairs and then applying circuit models for each of them. This modeling methodology is first verified against a commercial electromagnetic solver and subsequently employed for a variety of examples for signal and power integrity analysis. Specifically, in Chapter 6, a power and bandwidth trade-off analysis is performed on TSV array between different signaling and termination techniques, such as single-ended, differential and reduced-swing signaling. From our study, to achieve the best power efficiency, unterminated single-ended reduced-swing signaling should be applied, while terminated single-ended signaling can provide the maximum bandwidth. For 2D lossy redistribution layer CPW on the silicon interposer, a new equivalent circuit model is also proposed in Chapter 5. The shunt admittance of the CPW is derived rigorously based on the partial equivalent element circuit method considering the metal-isolator-semiconductor layered media. The silicon effects on series impedance are also considered by employing the modified green functions with complex image theory. Beyond TSV and RDL trace, critical design challenges for the junction structure between TSVs and RDL traces are also revealed and analyzed in Chapter 6.

APPENDIX A

Mathematical Proof for Pair-based TSV Array Model

For any nonsingular $N \times N$ matrix X and rank-one $N \times N$ matrix H, it can be shown that

$$KHK' = 0, \tag{A.1}$$

where $K = B(AXA')^{-1}AX$. A is a $(S+1) \times N$ matrix and B is a $S \times (S+1)$ matrix, which can be found in (4.30).

To apply to our case, we can simply define H and G by using (4.32) and let $X = G^{-1}$ since G is symmetric.

Proof. (A.1) can be proven true by showing that for any nonsingular $N \times N$ matrix X and $K = B(AXA')^{-1}AX$, the sum of each row of K equals to zero such that KHK' equals to zero for any $N \times N$ rank-one matrix H.

Equivalently, we want to show $\sum_{j=1}^{N} k_{ij} = 0$ for row $i = 1, 2, \cdots, S$ in K. We first let $Y = AX$ and $M = (YA')^{-1}Y$ such that K can be expressed as

$$K = B(AXA')^{-1}AX = B \cdot (YA')^{-1}Y = B \cdot M. \tag{A.2}$$

Without loss of generality, we simply assume

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ y_{(S+1)1} & y_{(S+1)2} & \cdots & y_{(S+1)N} \end{bmatrix}, \tag{A.3}$$

with the following equality must be satisfied:

$$Y = (YA')M, \tag{A.4}$$

or equivalently

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ y_{(S+1)1} & y_{(S+1)2} & \cdots & y_{(S+1)N} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1S} & \sum_{i=S+1}^{N} y_{1i} \\ y_{21} & y_{22} & \cdots & y_{2S} & \sum_{i=S+1}^{N} y_{2i} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{(S+1)1} & y_{(S+1)2} & \cdots & y_{(S+1)S} & \sum_{i=S+1}^{N} y_{(S+1)i} \end{bmatrix} \cdot M, \tag{A.5}$$

where

$$\begin{bmatrix} y_{1(S+1)} & \cdots & y_{1N} \\ y_{2(S+1)} & \cdots & y_{2N} \\ \vdots & \vdots & \vdots \\ y_{(S+1)(S+1)} & \cdots & y_{(S+1)N} \end{bmatrix} = \tag{A.6}$$

$$\begin{bmatrix} y_{11} & \cdots & y_{1S} & \sum_{i=S+1}^{N} y_{1i} \\ y_{21} & \cdots & y_{2S} & \sum_{i=S+1}^{N} y_{2i} \\ \vdots & \vdots & \vdots & \vdots \\ y_{(S+1)1} & \cdots & y_{(S+1)S} & \sum_{i=S+1}^{N} y_{(S+1)i} \end{bmatrix} \cdot \begin{bmatrix} m_{1(S+1)} & \cdots & m_{1N} \\ m_{2(S+1)} & \cdots & m_{2N} \\ \vdots & \vdots & \vdots \\ m_{(S+1)(S+1)} & \cdots & m_{(S+1)N} \end{bmatrix}.$$

As a result, M can be expressed as

$$M = \begin{bmatrix} 1 & 0 & \cdots & 0 & m_{1(S+1)} & \cdots & m_{1N} \\ 0 & 1 & \cdots & 0 & m_{2(S+1)} & \cdots & m_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & m_{S(S+1)} & \cdots & m_{SN} \\ 0 & 0 & \cdots & 0 & m_{(S+1)(S+1)} & \cdots & m_{(S+1)N} \end{bmatrix},$$

where m_{ij} is defined in (A.6) for $i = 1, 2, \cdots, (S+1)$, $j = (S+1), \cdots, N$.

In other words,

$$y_{i(S+j)} = \sum_{k=1}^{S} y_{ik} m_{k(S+j)} + m_{(S+1)(S+j)} \sum_{l=S+1}^{N} y_{il}, \tag{A.7}$$

for $i = 1, 2, \cdots, (S+1)$ and $j = 1, 2, \cdots, (N-S)$, and we can find that the summation for each row i in (A.6):

$$\sum_{j=1}^{N-S} y_{i(S+j)} \tag{A.8}$$

$$= \sum_{j=1}^{N-S}\sum_{k=1}^{S} y_{ik} m_{k(S+j)} + \sum_{j=1}^{N-S}\sum_{k=1}^{N-S} y_{i(S+k)} m_{(S+1)(S+j)}$$

$$= \sum_{k=1}^{S} y_{ik}\Big(\sum_{j=1}^{N-S} m_{k(S+j)}\Big) + \sum_{k=1}^{N-S} y_{i(S+k)}\Big(\sum_{j=1}^{N-S} m_{(S+1)(S+j)}\Big)$$

$$= \sum_{k=1}^{S} y_{ik}\Sigma_k + \Sigma_{(S+1)} \sum_{k=1}^{N-S} y_{i(S+k)}. \tag{A.9}$$

From (A.9), it is clear that

$$\Sigma_k = \begin{cases} 0 & \text{if } k \le S \\ 1 & \text{if } k = S+1 \end{cases} \tag{A.10}$$

and the sum for any row k in M all equals to 1 since

$$\sum_{j=1}^{N} m_{kj} = \begin{cases} 1 + \sum_{j=N-S}^{N} m_{k(S+j)} = 1 + \Sigma_k = 1 & \text{if } k \le S \\ \sum_{j=N-S}^{N} m_{(S+1)(S+j)} = \Sigma_{(S+1)} = 1 & \text{if } k = S+1 \end{cases} \tag{A.11}$$

Equivalently,

$$\sum_{j=1}^{N} m_{ij} = \sum_{j=1}^{N} m_{kj} \quad for \quad i, k = 1, 2, \cdots, S+1, \tag{A.12}$$

for any $(S+1) \times N$ matrix Y and $M = (YA')^{-1}Y$.

Since $B = \begin{bmatrix} I_S & -1_{S\times 1} \end{bmatrix}$ and $K = BM$, it can be easily shown that the sum of each row in K equals to zero,

$$\sum_{j=1}^{N} k_{ij} = 0 \quad for \quad i = 1, 2, \cdots, S, \tag{A.13}$$

and KHK' equals to zero for any $N \times N$ rank-one matrix H, which proves (A.1).

□

APPENDIX B

3D IC Technology and TSV Dimension in ITRS Roadmap

The International Technology Roadmap for Semiconductors (ITRS) 2007 identified the interconnection problem as one of the near- term (through 2015) grand challenges since additional device and interconnect scaling alone could not deliver the required increase in IC performance. A 3D technology with TSVs aligned on a tight pitch was one of the new technologies identified to meet that challenge.

The 3D IC technology potential is not leveraged due to insufficient scaling of the through-silicon via (TSV) pitch. Since technological limits restrict the TSV ratio at about 10:1, minimum chip thickness enables us to take full advantage of the 3D IC concept. ITRS 2008 projects decreasing chip thickness to support three-dimensional integrated circuit (3D IC) solutions. Wafer thinning was the only technique considered for achieving thin chips. It was projected that at thicknesses below $10\mu m$ a sequential combination of mechanical grinding, chemicalmechanical polishing (CMP), wet etching and plasma treatment, and dry chemical etching would be required to allow for control of such small chip thickness and to produce a die free of stress. Table B.1 shows a summary of the ITRS projection on high-density TSV [Ass08].

In this dissertation, most of the TSV signal integrity studies focus on the TSV array in the silicon interposer. For the interposer, typical silicon substrate thickness is usually larger than $50\mu m$. With the 10:1 aspect ratio, the minimum TSV diameter is around $5\mu m$. On the other hand, the TSV pitch is limited by

Table B.1: High-density TSV projections in 2008 ITRS update [Ass08].

	2008	2009	2010	2011	2012	2013	2014	2015
Min Wafer Thickness (μm)	15	15	10	10	10	10	10	8
TSV diameter (μm)	1.6	1.5	1.4	1.3	1.3	1.2	1.2	1
TSV pitch (μm)	5.6	5.5	4.4	3.8	3.8	2.7	2.6	2.5
PAD spacing (μm)	1	1	1	0.5	0.5	0.5	0.5	0.5
PAD diameter (μm)	4.6	4.5	3.4	3.3	3.3	2.2	2.1	2

the micro-bump pitch, the landing pad size and the keep-out zone (KOZ), and is usually larger than $30\mu m$ for $10\mu m$ diameter TSVs. However, the proposed TSV array modeling method is not only valid for the interposer TSV but also valid for the high-density TSV in future 3D IC as long as the proximity effect assumption holds. As shown in Table B.1, the $6\times$ ratio between TSV pitch and radius still holds for future high-density TSV.

References

[AMH99] J. Aguilera, R. Marques, and M. Horno. "Improved quasi-static spectral domain analysis of microstrip lines on high-conductivity insulator-semiconductor substrates." *Microwave and Guided Wave Letters, IEEE*, **9**(2):57 –59, feb 1999.

[AN01] R. Achar and M.S. Nakhla. "Simulation of high-speed interconnects." *Proceedings of the IEEE*, **89**(5):693–728, May 2001.

[ANS] ANSYS, Inc. "http://www.ansys.com/.".

[Ass05] Semiconductor Industry Associate. *Int. Technology Roadmap for Semiconductors*. 2005.

[Ass08] Semiconductor Industry Associate. *Int. Technology Roadmap for Semiconductors*. 2008.

[BA04] J. Buckwalter and B. Analui. "Predicting Data-Dependent Jitter." *IEEE Transactions on Circuits and Systems*, **51**(9), Sep 2004.

[BAC07] K Bernstein, P Andry, J Cann, P Emma, D Greenberg, W Haensch, M Ignatowski, S Koester, J Magerlein, R Puri, and A Young. "Interconnects in the Third Dimension: Design Challenges for 3D ICs." *Design Automation Conference, 2007. 44th ACM/IEEE*, pp. 562 – 567, 2007.

[BAH04] J. Buckwalter, B. Analui, and A. Hajimiri. "Data-dependent jitter and crosstalk-induced bounded uncorrelated jitter in copper interconnects." *Microwave Symposium Digest, 2004 IEEE MTT-S International*, **3**:1627–1630 Vol.3, June 2004.

[BH05] J. Buckwalter and A. Hajimiri. "Crosstalk-induced jitter equalization." *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005*, pp. 409–412, Sept. 2005.

[BKN03] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De. "Parameter variations and impact on circuits and microarchitecture." In *Proc. Design Automation Conf. (DAC)*, 2003.

[Bre05] G. Breed. "Analyzing Signals Using the Eye Diagram." *High Frequency Electronics*, pp. 50–53, Nov. 2005.

[BS98] W.T. Beyene and J.E. Schutt-Aine. "Efficient transient simulation of high-speed interconnects characterized by sampled data." *Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, IEEE Transactions on*, **21**(1):105–114, Feb 1998.

[BSK01] K Banerjee, S Souri, P Kapur, and K Saraswat. "3-D ICs: a novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration." *Proceedings of the IEEE*, **89**(5):602 – 633, 2001.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambrdige University Press, 2004.

[CFB09] L Cadix, A Farcy, C Bermond, C Fuchs, P Leduc, M Rousseau, M Assous, A Valentian, J Roullard, E Eid, N Sillon, B Flechet, and P Ancey. "Modelling of Through Silicon Via RF performance and impact on signal transmission in 3D integrated circuits." *3D System Integration, 2009.*, pp. 1 – 7, 2009.

[Che99] W Chew. "Waves and fields in inhomogenous media." *Recherche*, Jan 1999.

[CYF91] Y Chow, J Yang, and D Fang.... "A closed-form spatial Green's function for the thick microstrip substrate." *Microwave Theory and* ..., Jan 1991.

[DKC01] H. Darabi, S. Khorram, Hung-Ming Chien, Meng-An Pan, S. Wu, S. Moloudi, J.C. Leete, J.J. Rael, M. Syed, R. Lee, B. Ibrahim, M. Rofougaran, and A. Rofougaran. "A 2.4-GHz CMOS transceiver for Bluetooth." *Solid-State Circuits, IEEE Journal of*, **36**(12):2016–2024, Dec 2001.

[DLR12] T.O. Dickson, Yong Liu, S.V. Rylov, Bing Dang, C.K. Tsang, P.S. Andry, J.F. Bulzacchelli, H.A. Ainspan, Xiaoxiong Gu, L. Turlapati, M.P. Beakes, B.D. Parker, J.U. Knickerbocker, and D.J. Friedman. "An 8x 10-Gb/s Source-Synchronous I/O System Based on High-Density Silicon Carrier Interconnects." *Solid-State Circuits, IEEE Journal of*, **47**(4):884 –896, april 2012.

[ES06] N. Een and N. Sorensson. "Translating pseudo-boolean constraints into SAT." *JSAT*, **2**:1–26, 2006.

[Has71] Hasegawa. "Properties of Microstrip Line on Si-SiO2 System." *Microwave Theory and Techniques*, **19**(11):869 – 881, 1971.

[Hay00] S. Haykin. *Communication Systems*. John Wiley and Sons, 2000.

[HCL99] L He, N Chang, S Lin, and O.S Nakagawa. "An efficient inductance modeling for on-chip interconnects." *Custom Integrated Circuits, 1999. Proceedings of the IEEE.*, pp. 457–460, 1999.

[HL01] H. Huang and E. K. F. Lee. "Design of Low Voltage CMOS Continuous-Time Filter with On-Chip Automatic Tuning." *IEEE J. Solid-State Circuits*, 2001.

[HST07] M. Hashimoto, J. Siriporn, A. Tsuchiya, Haikun Zhu, and Chung-Kuan Cheng. "Analytical Eye-diagram Model for On-chip Distortionless Transmission Lines and Its Application to Design Space Exploration." *Custom Integrated Circuits Conference, 2007. CICC '07. IEEE*, pp. 869–872, Sept. 2007.

[HWM05] P.K. Hanumolu, Gu-Yeon Wei, and Un-Ku Moon. "Equalizers for high-speed serial links." *International Journal of High Speed Electronics and Systems*, **15**(2):429–458, 2005.

[KHK05] Kyung Ki Kim, Jing Huang, Yong-Bin Kim, and F. Lombardi. "On the modeling and analysis of jitter in ATE using Matlab." *Defect and Fault Tolerance in VLSI Systems, 2005. DFT 2005. 20th IEEE International Symposium on*, pp. 285–293, Oct. 2005.

[KLS06] Jingook Kim, Jiwang Lee, Eakhwan Song, Jeonghyeon Jo, and Joungho Kim. "Compensation of undesired channel effects by frequency domain optimization of pre-emphasis filter for over Gbps signaling." *Electromagnetic Compatibility, 2006. EMC 2006. 2006 IEEE International Symposium on*, **3**:721–726, Aug. 2006.

[KNR08] Kai Kang, Lan Nan, S.C. Rustagi, K. Mouthaan, Jinglin Shi, R. Kumar, Wen-Yan Yin, and Le-Wei Li. "A Wideband Scalable and SPICE-Compatible Model for On-Chip Interconnects Up to 110 GHz." *Microwave Theory and Techniques, IEEE Transactions on*, **56**(4):942 – 951, april 2008.

[KPC11] J Kim, J.S Pak, J Cho, E Song, H Kim, T Song, J Lee, H Lee, K Park, and S Yang. "High-frequency scalable electrical model and analysis of a through silicon via (TSV)." *Components, Packaging and Manufacturing Technology, IEEE Transactions on*, **1**(2):181–195, 2011.

[KRF05] A. Kuo, R. Rosales, T. Farahmand, S. Tabatabaei, and A. Ivanov. "Crosstalk bounded uncorrelated jitter (BUJ) for high-speed interconnects." *Instrumentation and Measurement, IEEE Transactions on*, **54**(5):1800–1810, Oct. 2005.

[LL08] A. C. Y. Lin and M. J. Loinaz. "A Serial Data Transmitter for Multiple 10Gb/s Communication Standards in 0.13um CMOS." *ISSCC, 2008 IEEE International*, 2008.

[LLF12] Yong Liu, Wing Luk, and D. Friedman. "A compact low-power 3D I/O in 45nm CMOS." *ISSCC, 2012 IEEE International*, pp. 142 –144, 2012.

[LWT04] M. Li, S. Wang, Y. Tao, and T. Kwasniewski. "FIR filter optimisation as pre-emphasis of high-speed backplane data transmission." *Electronics Letters*, **40**(14):912–913, July 2004.

[MB04] B. Murmann and B. Boser. "Digitally Assisted Analog Integrated Circuits." *Queue*, **2**(1):64–71, 2004.

[Mil81] Kenneth S Miller. "On the inverse of the sum of matrices." *Math. Mag.*, **54**(2):67–72, 1981.

[MK02] M. Mansuri and Chih-Kong Ken. "Jitter optimization based on phase-locked loop design parameters." *Solid-State Circuits, IEEE Journal of*, **37**(11):1375–1382, Nov 2002.

[MOD98] V. Milanovic, M. Ozgur, D.C. DeGroot, J.A. Jargon, M. Gaitan, and M.E. Zaghloul. "Characterization of broad-band transmission for coplanar waveguides on CMOS silicon substrates." *Microwave Theory and Techniques, IEEE Transactions on*, **46**(5):632 –640, may 1998.

[MSO06] M. Mani, A. K. Singh, and M. Orshansky. "Joint design-time and postsilicon minimization of parametric yield loss using adjustable robust optimization." In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, 2006.

[MTH03] G. Miller, M. Timko, H.-S.Lee, E. Nestler, M. Mueck, and P. Ferguson. "Design and modeling of a 16-bit 1.5MSPS successive approximation ADC with non-binary capacitor array." *Proc. IEEE/ACM International Great Lakes Symposium on VLSI*, 2003.

[OFK04] N. Ou, T. Farahmand, A. Kuo, S. Tabatabaei, and A. Ivanov. "Jitter models for the design and test of Gbps-speed serial interconnects." *Design & Test of Computers, IEEE*, **21**(4):302–313, July-Aug. 2004.

[Pau06] C. R. Paul. *Introduction to electromagnetic compatibility - second edition.* Wiley Interscience, 2006.

[PBW08] Kevin Powell, Stephen Burgess, Tony Wilby, Rhonda Hyndman, and John Callahan. "3D IC Process integration challenges and solutions." *Interconnect Technology Conference, 2008. International*, pp. 40 – 42, 2008.

[PCK10] Jun So Pak, Jonghyun Cho, Joohee Kim, Junho Lee, Hyungdong Lee, Kunwoo Park, and Joungho Kim. "Slow wave and dielectric

quasi-TEM modes of Metal-Insulator-Semiconductor (MIS) structure Through Silicon Via (TSV) in signal propagation and power delivery in 3D chip package." *Electronic Components and Technology Conference (ECTC), 2010 Proceedings 60th*, pp. 667 – 672, 2010.

[PDW89] M. Pelgrom, A. Duinmaijer, and A. Welbers. "Matching properties of MOS transistors." *IEEE J. Solid-State Circuits*, 1989.

[PKC11] Jun So Pak, Joohee Kim, Jonghyun Cho, Kiyeong Kim, Taigon Song, Seungyoung Ahn, Junho Lee, Hyungdong Lee, Kunwoo Park, and Joungho Kim;. "PDN Impedance Modeling and Analysis of 3D TSV IC by Using Proposed P/G TSV Array Model Based on Separated P/G TSV and Chip-PDN Models." *Components, Packaging and Manufacturing Technology*, **1**(2):208 – 219, 2011.

[PKT07] V. Popescu, B.S. Kirei, M. Topa, and C. Munteanu. "Analysis of Lossless Differential Microstrip Line." *Electronics Technology, 30th International Spring Seminar on*, pp. 551–554, May 2007.

[PMM06] G. Plaza, R. Marques, and F. Medina. "Quasi-TM MoL/MoM approach for computing the transmission-line parameters of lossy lines." *Microwave Theory and Techniques, IEEE Transactions on*, **54**(1):198 –209, jan. 2006.

[Pre] Predicted Technology Model. "http://ptm.asu.edu/.".

[PRK07] Jun So Pak, Chunghyun Ryu, and Joungho Kim;. "Electrical characterization of trough silicon via (TSV) depending on structural and material parameters based on 3D full wave simulation." *Electronic Materials and Packaging, 2007. EMAP 2007. International Conference on*, pp. 1 – 6, 2007.

[Raz95] B. Razavi. *Principles of Data Conversion System Design*. John Wiley and Sons, 1995.

[Raz96] B. Razavi. *Monolithic Phase-Locked Loops and Clock Recovery Circuits: Theory and Design*. John Wiley & Sons, 1996.

[RB73] A Ruehli and P Brennan. "Efficient Capacitance Calculations for Three-Dimensional Multiconductor Systems." *Microwave Theory and Techniques*, **21**(2):76 – 82, 1973.

[RFC12] J. Roullard, A. Farcy, S. Capraro, T. Lacrevaz, C. Bermond, G. Houzet, J. Charbonnier, C. Fuchs, C. Ferrandon, P. Leduc, and B. Flechet. "Evaluation of 3D interconnect routing and stacking strategy to optimize high speed signal transmission for memory on logic." In *Electronic Components and Technology Conference (ECTC), 2012 IEEE 62nd*, pp. 8 –13, 29 2012-june 1 2012.

[Rue72] Ruehli. "Inductance Calculations in a Complex Integrated Circuit Environment." *IBM Journal of Research and Development*, **16**(5):470 – 481, 1972.

[SCA10] Fengda Sun, A Cevrero, P Athanasopoulos, and Y Leblebici. "Design and feasibility of multi-Gb/s quasi-serial vertical interconnects based on TSVs for 3D ICs." *VLSI System on Chip Conference (VLSI-SoC), 2010 18th IEEE/IFIP*, pp. 149 – 154, 2010.

[SH03] V. Stojanovic and M. Horowitz. "Modeling and analysis of high-speed links." *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003*, pp. 589–594, Sept. 2003.

[SLK00] S. Sidiropoulos, Dean Liu, Jaeha Kim, Guyeon Wei, and M. Horowitz. "Adaptive bandwidth DLLs and PLLs using regulated supply CMOS buffers." *VLSI Circuits, 2000. Digest of Technical Papers. 2000 Symposium on*, pp. 124–127, 2000.

[SNS08] S. Sen, V. Natarajan, R. Senguttuvan, and A. Chatterjee. "Pro-VIZOR: Process tunable virtually zero margin low power adaptive RF for wireless systems." In *Proc. Design Automation Conf. (DAC)*, June 2008.

[SRG09] A. Sayag, D. Ritter, and D. Goren. "Compact Modeling and Comparative Analysis of Silicon-Chip Slow-Wave Transmission Lines With Slotted Bottom Metal Ground Planes." *Microwave Theory and Techniques, IEEE Transactions on*, **57**(4):840 –847, april 2009.

[SRI11] K Salah, H Ragai, Y Ismail, and A El Rouby. "Equivalent lumped element models for various n-port Through Silicon Vias networks." *Design Automation Conference (ASP-DAC), 2011 16th Asia and South Pacific*, pp. 176 – 183, 2011.

[SYZ08] R. Shi, W. Yu, Yi Zhu, Chung-Kuan Cheng, and E.S. Kuh. "Efficient and accurate eye diagram prediction for high speed signaling." *Computer-Aided Design, 2008. ICCAD 2008. IEEE/ACM International Conference on*, pp. 655–661, Nov. 2008.

[TBP05] Yuming Tao, W. Bereza, R.H. Patel, S. Shumarayev, and T. Kwasniewski. "A signal integrity-based link performance simulation platform." *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005*, pp. 725–728, Sept. 2005.

[TN92] T.K. Tang and M.S. Nakhla. "Analysis of high-speed VLSI interconnects using the asymptotic waveform evaluation technique." *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, **11**(3):341–352, Mar 1992.

[TPH07] L.F. Tiemeijer, R.M.T. Pijper, R.J. Havens, and O. Hubert. "Low-Loss Patterned Ground Shield Interconnect Transmission Lines in Advanced IC Processes." *Microwave Theory and Techniques, IEEE Transactions on*, **55**(3):561 –570, march 2007.

[VMA03] J. VIta, A. Marques, P. Azevedo, and J. Franca. *Design Considerations for a Retargetable 12b 200MHz CMOS Current-Steering DAC*. Springer US, 2003.

[WKC92] Ruey-Beei Wu, Chien-Nan Kuo, and K.K. Chang. "Inductance and resistance computations for three-dimensional multiconductor interconnection structures." *Microwave Theory and Techniques, IEEE Transactions on*, **40**(2):263 –271, feb 1992.

[WWX08] F. Wang, X. Wu, and Y. Xie. "Variability-Driven Module Selection with Joint Design Time Optimization and Post-Silicon Tuning." In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, 2008.

[XBZ10] Z Xu, A Beece, D Zhang, Q Chen, K Chen, K Rose, and J.Q Lu. "Crosstalk evaluation, suppression and modeling in 3D through-strata-via (TSV) network." *3DIC, 2010 IEEE International*, pp. 1–8, 2010.

[XKS11] Chuan Xu, V. Kourkoulos, R. Suaya, and K. Banerjee. "A Fully Analytical Model for the Series Impedance of Through-Silicon Vias With Consideration of Substrate Effects and Coupling With Horizontal Interconnects." *Electron Devices, IEEE Transactions on*, **58**(10):3529 –3540, oct. 2011.

[XL11] Zheng Xu and Jian-Qiang Lu;. "Through-Strata-Via (TSV) Parasitics and Wideband Modeling for Three-Dimensional Integration/Packaging." *Electron Device Letters, IEEE*, **32**(9):1278 – 1280, 2011.

[XLS10] Chuan Xu, Hong Li, R Suaya, and K Banerjee. "Compact AC Modeling and Performance Analysis of Through-Silicon Vias in 3-D ICs." *Electron Devices*, **57**(12):3405 – 3417, 2010.

[YKL09] Kihyun Yoon, Gawon Kim, Woojin Lee, Taigon Song, Junho Lee, Hyungdong Lee, Kunwoo Park, and Joungho Kim;. "Modeling and analysis of coupling between TSVs, metal, and RDL interconnects in TSV-based 3D IC with silicon interposer." *Electronics Packaging Technology Conference, 2009. EPTC '09. 11th*, pp. 702 – 706, 2009.

[YLN08] Yun Ye, F. Liu, S. Nassif, and Yu Cao. "Statistical modeling and simulation of threshold variation under dopant fluctuations and line-edge roughness." *Proc. Design Automation Conf. (DAC)*, June 2008.

[YNM03] H. Ymeri, B. Nauwelaers, K. Maex, and D. De Roest. *"Broadband impedance parameters of symmetric coupled coplanar CMOS interconnects"*. Electrotechnical Review, 2003.

[ZCL09] X. Zhang, T.C. Chai, J.H. Lau, C.S. Selvanayagam, K. Biswas, Shiguo Liu, D. Pinjala, G.Y. Tang, Y.Y. Ong, S.R. Vempati, E. Wai, H.Y. Li, E.B. Liao, N. Ranganathan, V. Kripesh, Jiangyan Sun, J. Doricko, and C.J. Vath. "Development of through silicon via (TSV) interposer technology for large die (21x21mm) fine-pitch Cu/low-k FCBGA package." In *Electronic Components and Technology Conference, 2009. ECTC 2009. 59th*, pp. 305 –312, may 2009.

[ZHT00] Ji Zheng, Yeon-Chang Hahm, V.K. Tripathi, and A. Weisshaar. "CAD-oriented equivalent-circuit modeling of on-chip interconnects on lossy silicon substrate." *Microwave Theory and Techniques, IEEE Transactions on*, **48**(9):1443 –1451, sep 2000.

[ZK02] G. Zhong and C.-K. Koh. "Exact closed form formula for partial mutual inductances of on-chip interconnects." In *Computer Design: VLSI in Computers and Processors, 2002. Proceedings. 2002 IEEE International Conference on*, pp. 428 – 433, 2002.

[ZTW01] Ji Zheng, V.K. Tripathi, and A. Weisshaar. "Characterization and modeling of multiple coupled on-chip interconnects on silicon substrate." *Microwave Theory and Techniques, IEEE Transactions on*, **49**(10):1733 –1739, oct 2001.