MANAGING SIGNAL, POWER, AND THERMAL INTEGRITY FOR THREE-DIMENSIONAL INTEGRATED CIRCUITS AND SYSTEMS

A Dissertation
Presented to
The Academic Faculty

by

Sung Joo Park

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

Georgia Institute of Technology
August 2016

MANAGING SIGNAL, POWER, AND THERMAL INTEGRITY FOR THREE-DIMENSIONAL INTEGRATED CIRCUITS AND SYSTEMS

Approved by:

Dr. Madhavan Swaminathan, Advisor
School of Electrical & Computer
Engineering
Georgia Institute of Technology

Dr. David Keezer
School of Electrical & Computer
Engineering
Georgia Institute of Technology

Dr. Sung-Kyu Lim
School of Electrical & Computer
Engineering
Georgia Institute of Technology

Dr. Muhannad S. Bakir
School of Electrical & Computer
Engineering
Georgia Institute of Technology

Dr. Yogendra K. Joshi
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: July 13, 2016

To my family and friends

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This dissertation would not have been possible without invaluable support and help from many people. I would like to express my sincere appreciation to all of them. First, I would like to express my appreciation to my advisor, Professor Madhavan Swaminathan, for his guidance and consistent support. I may not have accomplished and finished the Ph.D. course without him. I also would like to extend my gratitude to the Ph.D. dissertation reading committee, Dr. David Keezer and Dr. Sung-Kyu Lim, for their review about my dissertation and valuable comments for the dissertation and proposal. I also would like to thank the other dissertation committee members, Dr. Muhannad S. Bakir, and Dr. Yogendra K. Joshi, for their valuable comments and contribution to the improvement of my dissertation.

I would like to give my special thanks to current and former members of the Mixed Signal Design group, Mr. Colin Pardue, Dr. Jianyong Xie, Dr. Biancun Xie, Dr. Ming Yi, Mr. Rishik Bazaz, Dr. Kyu Hwan Han, Dr. Jae Young Choi, Dr. Myunghyun Ha, Dr. Suzanne Huh, and Mr. Sang Kyu Kim, and visiting researchers, Dr. Jun Ki Min, Mr. You Keun Han, Dr. Sebastian Mueller, and Dr. Anto K Davis. In addition, I would like to extend my thanks for interesting and helpful discussions to Satyan Tellikepali, David Zhang, and Nitish Natu.

I would like to extend my gratitude to my friends at Georgia Tech and I also would like to thank Samsung Electronics for providing funds for my graduate study at Georgia Tech.

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

3-D	Three-Dimension
IC	Integrated Circuit
SI	Signal Integrity
PI	Power Integrity
TI	Thermal Integrity
TSV	Through-Silicon Via
CMOS	Complementary Metal Oxide Semiconductor
CDN	Clock Distribution Network
PDN	Power Distribution Network
SSN	Simultaneous Switching Noise
PCB	Printed Circuit Board
ESL	Equivalent Series Inductance
ESR	Equivalent Series Resistance
VRM	Voltage Regulator Module
PTC	Positive Temperature Coefficient
FPGA	Field-Programmable Gate Array

CHAPTER 1   INTRODUCTION

1.1 Background and Motivation

The perpetual growth of information technology has demanded that the electronics industry continues to improve in performance, density, functionality, and reliability. Many technical innovations have overcome technical hurdles to enable and achieve these improvements. One of the improvements, transistor density in the last several decades, has increased with Moore's law [1] accompanied by increased operation frequency of electronic systems. In addition, advancements in integration have led to smaller size devices, as shown in Figure 1 [2].
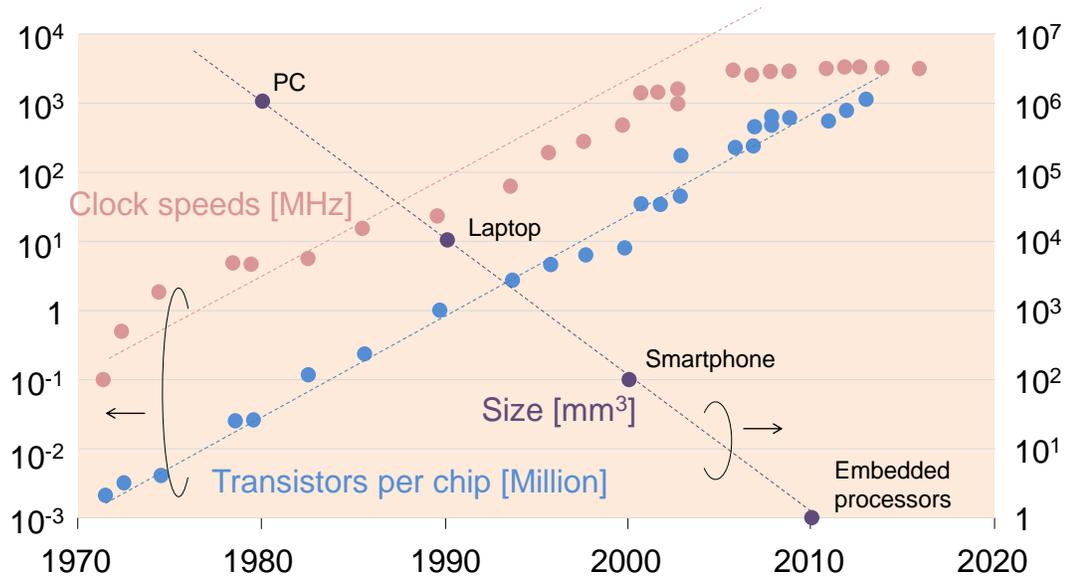
Figure 1   **Electronic system trend in million transistors per chip, size of computing system in mm³ and clock speed in MHz; modified from [2].**

However, this trend of continuous growth challenges the electrical integrity of electronic systems. Integrity, a nomenclature which means an undiminished status, is

widely used to guarantee the performance of electronics systems through simulations. The goal of signal integrity (SI) is to ensure perfect timing and minimum reflection without any kind of coupling or distortion. In contrast, power integrity (PI) estimates non-ideal effects of the power/ground network with minimum switching noise. Both are crucial to meet the operational specifications of digital systems and guarantee better performance, as shown in Figure 2.
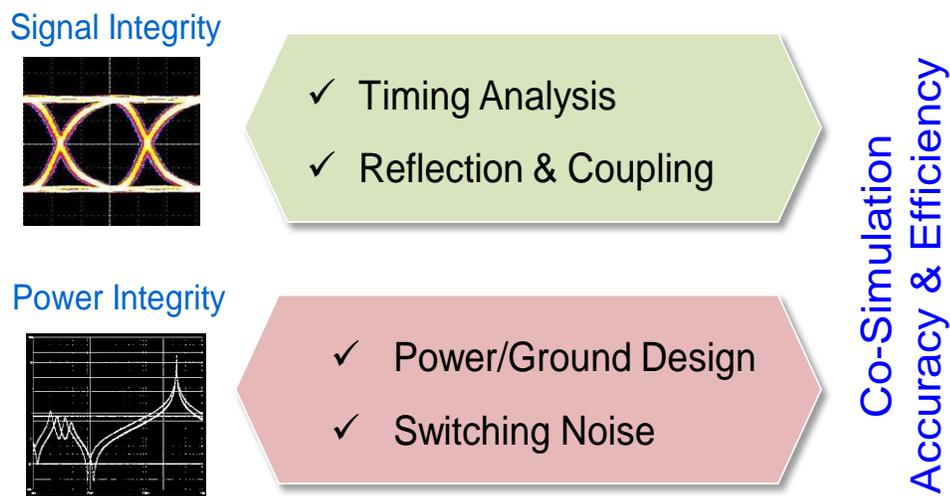
Figure 2   **Signal integrity and power integrity for high speed designs.**

One of the most significant discussions in electronic system packaging is three dimensional integration. Many design challenges, such as performance, form factor, and density, have been overcome using this technique. Therefore, three-dimensional integrated circuits (3-D ICs) have many advantages, such as dramatically improved transistor density, electrical performance, and functionality [3]. However, 3-D technology also has some engineering related issues in modeling, design, fabrication, testing, and thermal management. Figure 3 shows the most widely used application options of TSV-based 3-D integration with and without an interposer, which are referred to as 2.5-D and 3-D integration.
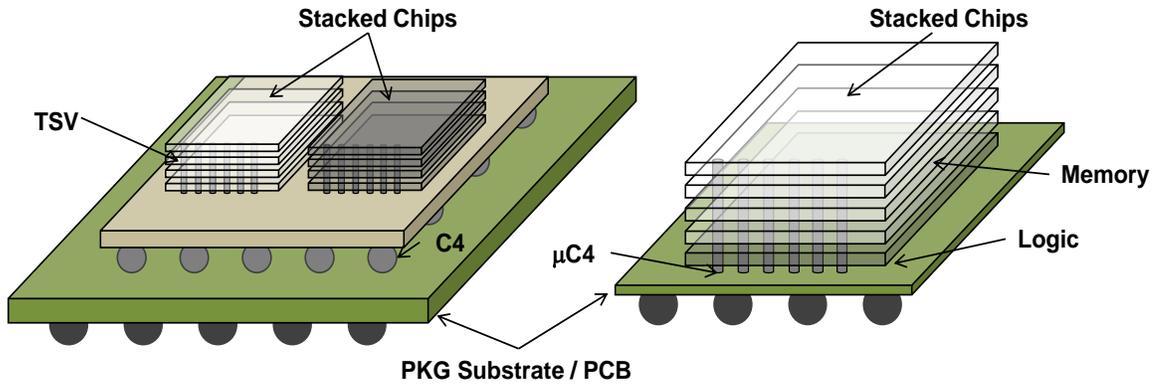
Figure 3　　　**3-D integration options. (a) 2.5-D with silicon or glass interposer and (b) 3-D integration with TSVs.**

Increased circuit and system density result in increased power and thermal density of 3-D ICs which requires consideration of thermal management solutions. Hot spots and temperature gradients, as shown in Figure 4, significantly affect signal quality and degrade system performance. Those temperature-induced problems originate from increased power density in 3-D ICs [4]. Previous literature has indicated that temperature gradients in 3-D ICs are three times more severe than conventional ICs [5] and highest on-chip temperature gradients in some systems can be over 50 degrees Celsius [6], which pose significant problems for ensuring system reliability [7].
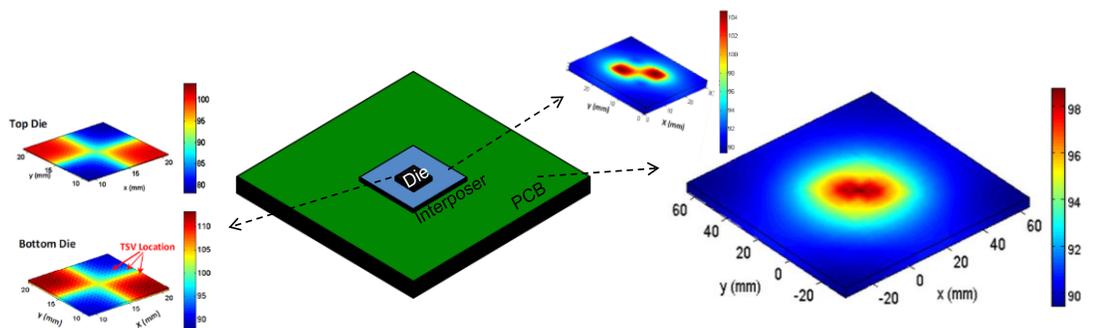
Figure 4　　　**Temperature gradient in a 3-D integrated system [8].**

3

For example, consider a 3-D system shown in Figure 4 comprised of two stacked dies with different power maps. They are stacked on top of a silicon interposer which is in turn mounted on a printed circuit board. The hot spots in various parts of the system shown in Figure 4 affect active as well as passive circuits due to changes in resistivity, mobility, and threshold voltages.

Temperature-induced problems affect the integrity of signals and system performance which can cause reliability issues. With higher operation frequencies and tighter timing budgets, the timing margin of system operation is continually shrinking. Furthermore, higher frequency operation of globally distributed clock signals is more sensitive to temperature variations across the chip. Because the clock signal synchronizes data signals, design of the clock distribution network is very critical in modern digital systems [11].

The interaction between thermal and electrical integrity is becoming increasingly difficult to ignore because they closely interact and the amount of interaction can become non-negligible, as described in Figure 5. In this research, interactions between signal integrity (SI), power integrity (PI), and thermal integrity (TI) have been investigated. Power integrity and signal integrity are both affected by temperature gradients on the clock distribution network (CDN) in 3-D ICs. This poses unique problems for both modeling and design.
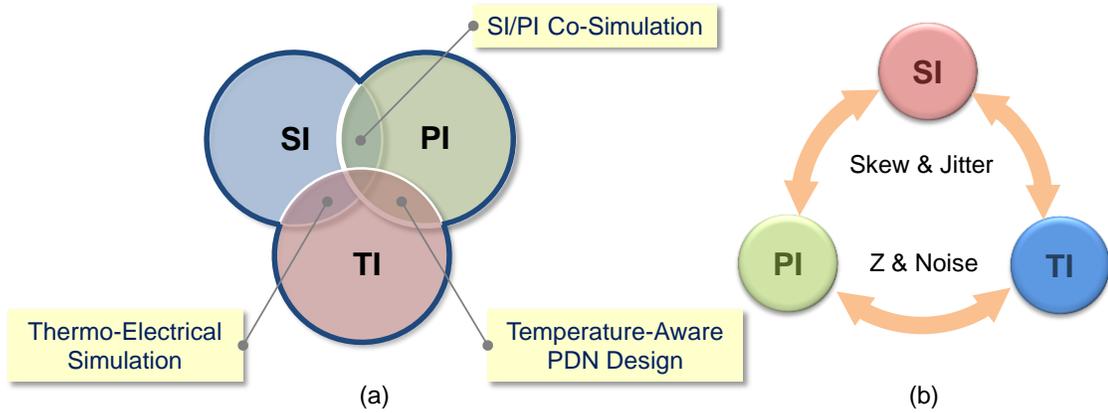
Figure 5   **(a) Inclusion of thermal integrity to electrical design and (b) interactions of signal, power and thermal integrity.**

1.2 Prior Work

In this thesis, (1) development of a flow for estimating signal, power, thermal integrity, (2) thermal skew compensation methods, and (3) optimization approaches are described and proposed. Several researchers have investigated and presented on these topics before. The differences between this research and prior work are discussed below.

First, most conventional circuit simulations have been using the global temperature for PVT (process, voltage, and temperature) variations to reflect thermal effects instead of using the local temperature in a die. With respect to electrical and thermal simulations, previous studies have shown that, (1) co-design of signal and power integrity [12] and (2) electro-thermal simulation approach for estimating temperature and voltage drop in a PDN [13]. In contrast, this dissertation provides a full system model for inclusion of thermal integrity into signal and power integrity and quantifies thermal effects on them.

Second, researchers have proposed temperature control approaches to mitigate the effect of temperature in electronic circuits and systems. These techniques are based on controlling physical heat to reduce maximum temperature [14], managing power using

optimized voltage and frequency operation, also known as dynamic voltage and frequency scaling (DVFS) [15], and performing thermal-aware floor planning of the chip layout for uniform heat distribution [16].

In comparison, this dissertation investigates that circuit-based approaches can be specialized for unique target parameters such as delay, noise, and jitter. The focus of this dissertation is to propose and optimize the efficiency of a few methods which have specific compensation performance and overhead associated with them, as discussed in chapter 4. Compensation methods presented in previous literature have been implemented using temperature-variable supply voltage level converters [17], multiple-threshold voltage design [18], dynamic body biasing [19], adaptive body biasing with V_T detection [20], and tuning pull-up and pull-down transistors with a programmable temperature compensation device [21]. In addition, previous work has used a dynamically-modified clock tree using tunable delay [22], adjusting the drive strength of buffers [23], using a reversal of temperature dependency [24], [25], and adapting temperature-dependent gate delay [26].

Last but not the least, several studies have proposed statistical methods such as worst case and Monte Carlo analyses [27] to optimize a large number of design parameters. Due to the large number of simulation cases and costly calculation overhead of these methods, others have proposed approaches that reduce the number of simulations, using design of experiments (DOE) [28]. However, the DOE approach has restrictions such as 1) interactions between parameters need to be minimal and 2) the number of levels is normally limited to below three. Moreover, some electrical and thermal parameters in electronic systems have continuous range, so quantization error can create problems during optimization. Other approaches for global optimization are available with multiple

algorithms [29]. However, global optimization algorithms require a lot of computing

resources because of the number of data sets and compute time for each data set.

Optimization of a non-linear function using the global optimization algorithms available in

MATLAB is compared with a machine-leaning method, namely Bayesian optimization, as

illustrated in Figure 5. Table 1 also compares the number of iterations and optimization

results by using global optimization algorithms and Bayesian Optimization which will be

further discussed in Chapter 5. As Figure 6 illustrates, global optimization algorithms

require up to 2500 iterations to converge, however machine-learning based algorithms can

find an optima in less than 100 iterations. As shown in Chapter 5, the optimized value using

Bayesian optimization was superior to the value obtained using a more conventional global

search algorithm, for a fixed number of iterations.
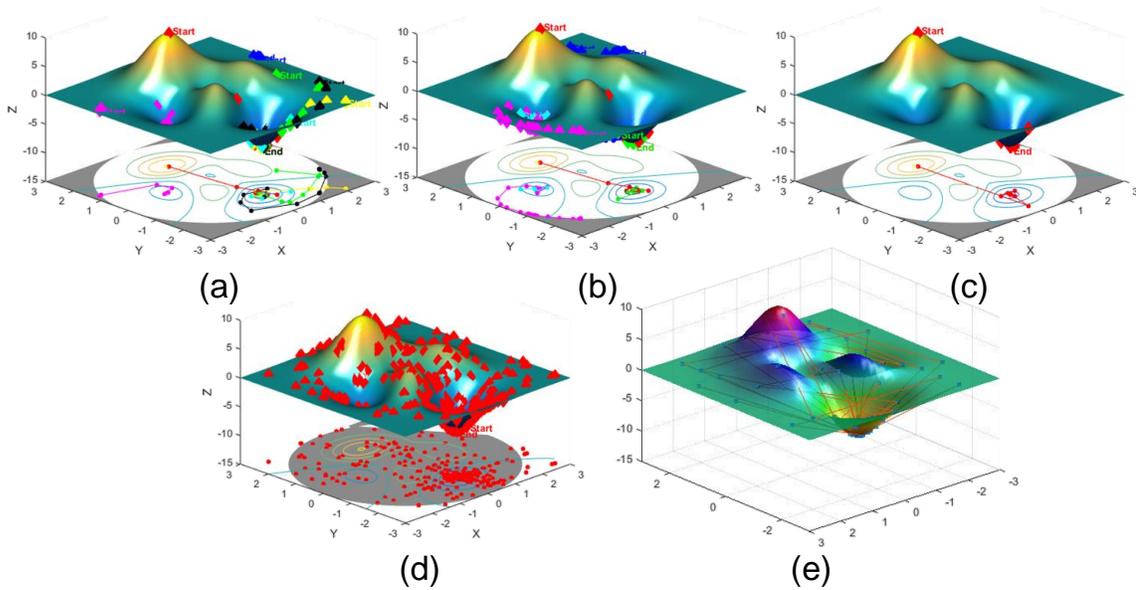
Figure 6 **Comparison of global optimization algorithms available in MATLAB with Bayesian optimization for a predefined function; (a) multi-start, (b) global search, (c) pattern search, (d) genetic algorithm, and (e) Bayesian optimization [81].**

Table 1 **Comparison of the optimization algorithms (X_1 and X_2 represent optimized values).**

	Algorithm	X_1	X_2	Iterations
(a)	Multi-start	-1.6255	0.2283	345
(b)	Global-search	0.2045	-1.3474	273
(c)	Pattern-search	-1.6255	0.2283	272
(d)	Genetic-algorithm	-1.6255	0.2283	2650
(e)	Bayesian opt. (this work)	-1.6251	0.2283	100

Recently, machine-learning (ML) algorithms have become popular and have been applied for various applications [30]-[33]. These techniques have enabled machines to learn from the training sequences by accumulating data sets through automated learning algorithms. Some of these methods have been adapted to electromagnetic problems [30], static timing analysis [31], high speed interconnect systems [32], and time domain performance estimation [33]. In this dissertation, we apply machine learning algorithms for design optimization of 3-D ICs and systems.

1.3 This Work

As discussed in the previous section, this dissertation presents a flow for electrical-thermal simulation, thermal skew compensation methods and optimization approaches. Three-dimension integrated systems require temperature-aware electrical simulations and analysis due to increased thermal density and operating frequency. This dissertation presents a flow for electrical-thermal simulations (Figure 7) and quantifies the impact of temperature in 3-D systems. To mitigate the temperature effects, this dissertation also proposes three energy-efficient mitigation methods to compensate for thermally induced

skew. These methods include adaptive voltage, controllable delay, and variable strength. These methods have been discussed with measurements using two test vehicles.
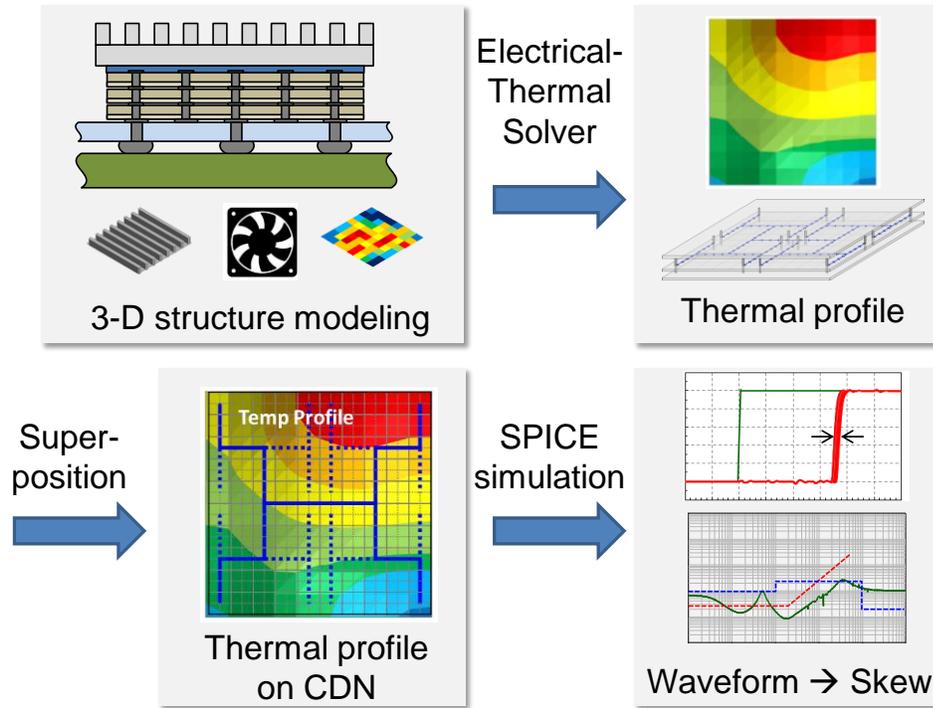
Figure 7   **Flow of electrical-thermal analysis for 3-D system design.**

Simulating 3-D systems requires large computing resources due to the large number of design parameters and multi-scale structure. To improve efficiency, a machine-learning algorithm is applied for the optimization of 3-D systems. The algorithm shows fast convergence with a large number of parameters for the examples considered.

1.4 Contributions

The objective of this dissertation is to develop a co-design methodology for 3-D ICs that can be used in electronic systems. This study also quantifies the effects of temperature on 3-D ICs and systems having high thermal density, thermal hot spots, and

temperature gradient. Temperature-related issues degrade not only signal quality but also the power distribution network (PDN) characteristics.

The work completed as part of this research include the following:

1. Simulation model of the clock distribution network (CDN) and power distribution network (PDN) in a 3-D system has been built and electrical-thermal simulations were performed. From the results of these thermo-electrical simulations, thermal effects such as increased skew and noise in 3-D systems were quantified.

2. To reduce and to compensate for temperature-induced skew in a clock distribution network, three skew compensation methods were proposed. The proposed methods use adaptive voltages, controllable delays, and variable strengths for clock repeaters and interconnect. The concepts were implemented with a recent technology node (45 nm) and their performance and design overhead were compared.

3. As a verification procedure, two types of test vehicles were designed. One was an FPGA-based test vehicle, and the other is a custom IC-based test vehicle. An H-tree CDN was implemented in the test vehicle. Measurement results using the designed test vehicles were correlated with electrical-thermal simulation results such as temperature profile and delay including temperature and compensation methods.

4. Design optimization using the proposed electrical-thermal simulation approach requires considerable computing resources because of many parameters involved and due to the multi-scale geometries. In this work,

machine learning methods were applied for design optimization. Bayesian optimization using Gaussian process shows that this approach is beneficial for optimization of 3-D systems.

In summary, the key contributions of this dissertation are as follows:

1) A temperature-aware signal and power integrity simulation method for 3-D ICs and systems with high thermal density and temperature gradient was introduced.

2) The delay, PDN impedance, and noise induced by temperature in 3-D ICs was simulated and analyzed, and optimal delay compensation methods were developed.

3) Test vehicles using FPGA and custom IC were designed and fabricated and the proposed methods were validated using measurements.

4) A machine-learning (ML) methods was applied for 3-D IC and system optimization, and the results quantified.

1.5 Organization of the Dissertation
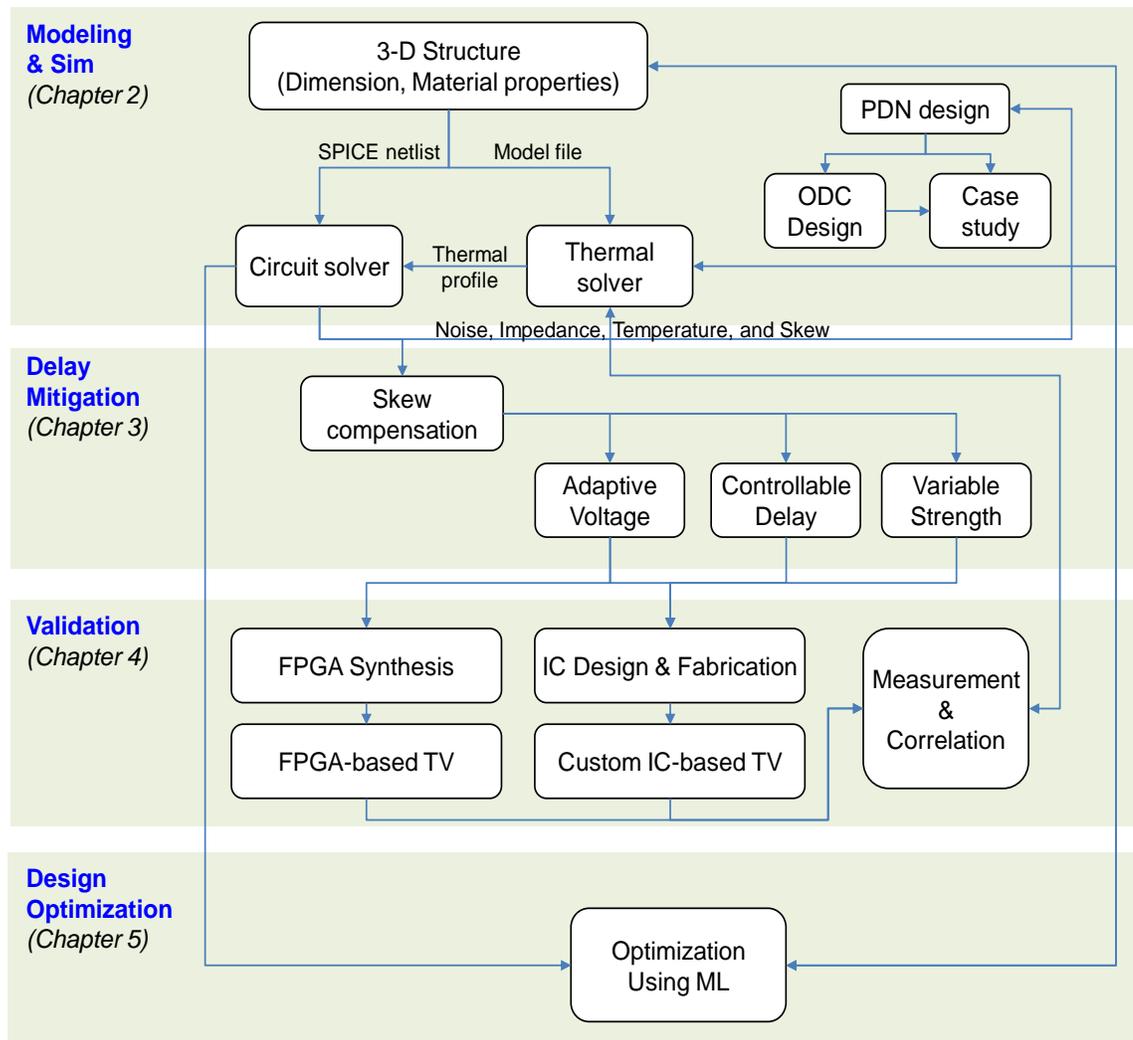
Figure 8   Organization of the dissertation.

As described in Figure 8, the rest of this dissertation is organized as follows. Electrical-thermal modeling with the target system configuration, the proposed electrical-thermal simulation flow and results including power delivery network design considering thermal effects are addressed in Chapter 2. The following chapters, Chapter 3 and Chapter 4, provide details on thermal-skew mitigation methods and measurement results for

verification and correlation, respectively. Chapter 5 presents a new optimization approach using machine-learning followed by summary in Chapter 6.

CHAPTER 2   ELECTRICAL-THERMAL MODELING AND

SIMULATION

2.1 Introduction

First, this chapter defines system configuration comprised of the 3-D IC stack with power maps, clock distribution network, and system details. 3-D integrated systems, as in the example shown in Figure 9 (a), having interposer and stacked dies, have configuration options with respect to number, type, and layout of the integrated dies [8]. This dissertation examines a TSV-based 3-D system comprised of three dies stacked and mounted on an interposer as shown in Figure 9 (b). The assumed configuration with an interposer covers feasible integration options available. The clock distribution network is placed inside of the stack, between two dies of logic, and are integrated and connected to a PCB through the interposer by using TSVs.

(a)   (b)
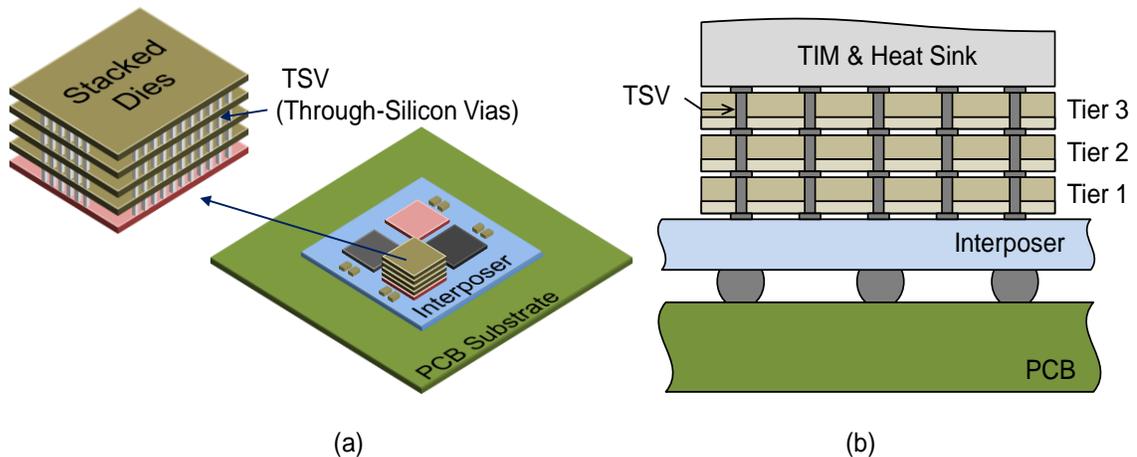
Figure 9   System configuration with a 3-D IC; (a) comprised of multiple dies, interposer, and PCB and (b) cross-sectional view of the target system.

2.2 System Configuration

In this chapter, clock distribution network and power distribution network, especially for timing skew in the clock distribution network and power supply fluctuation in the power distribution network (PDN), are the main focus for signal and power integrity evaluation. First, the clock was chosen because it is one of the most critical signals in synchronous digital systems because the signal is the timing reference for other signals. Furthermore, the clock signal becomes more sensitive with tighter timing margin under faster operation and temperature variation conditions [11]. Second, the power/ground network provides electrical sources, distribution rails, and reference of the voltage. Non-ideal effects of parasitics, such as inductance of planes and packages, return path discontinuity, and resonance, cause power supply fluctuation during simultaneous switching of the circuits. Even though simultaneous switching noise analysis has different consideration with specific types as shown in [34], it becomes more critical with faster switching rates during high speed operation.

The structure of the clock distribution network (CDN) in 3-D integration is important since it affects thermal and electrical characteristics. Previous approaches to the CDN architecture in 3-D systems are discussed in [35], [36]. The first method shown in [35] was implemented with clock routing on an interposer, and distributed through the TSVs. The second method in [36] uses multiple but symmetrical tree-structure TSVs to distribute loads.
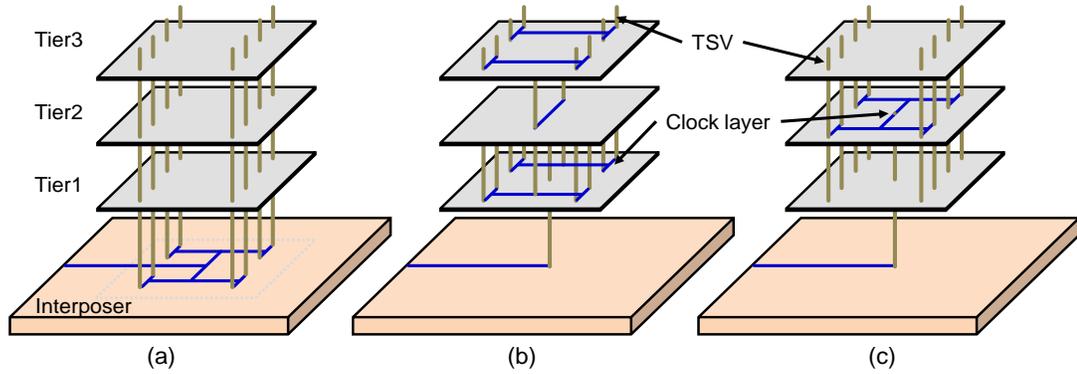
Figure 10   **3-D CDN structures: (a) CDN on an interposer [35], (b) CDN with tree-structure TSV [36], and (c) centered CDN; modified from [37].**

The structure proposed in this study uses a CDN structure constructed in the centered die, which requires fewer TSVs and routing of interposer layer and minimizes the effect of die to die variations. Electrical modeling shown in Chapter 3 will describe a more detailed structure containing cascaded chains of clock repeaters and metal interconnects. Figure 10 shows the three options for the CDN in 3-D ICs.

This chapter provides geometrical information and parameter values for the configuration in Figure 10 (c). Sizes of the die, interposer, and substrate PCB are 10 mm x 10 mm, 30 mm x 30 mm, and 100 mm x 100 mm, respectively. Layers of the PCB, the interposer, and the die are connected with C4 bumps and micro-bumps. In addition, the interposer and stacked dies are vertically connected using through-silicon vias. The detailed dimensions of the system configuration are listed in Table 2.

Table 2 **Geometrical dimension of the system configuration [37].**

	Unit	Value	Note
Die	mm³	10 x 10 x 0.2	Height x Length x Thickness
Interposer	mm³	30 x 30 x 0.2	Height x Length x Thickness
PCB	mm³	100 x 100 x 0.2	Height x Length x Thickness
C4 Bump	μm	300	Diameter
Micro-bump	μm	100	Diameter
TSV (interposer)	μm	30 | 100 | 100	Diameter | Height | Pitch
TSV (die)	μm	5 | 50	Diameter | Height

2.3 Electrical and Thermal Modeling

Most electrical parameters are affected by temperature. For example, the electrical resistivity of metal is linearly proportional to the temperature as shown in (1). Metals mainly used for interconnects in a chip (either aluminum or copper) have a temperature coefficient of resistivity (α_R) of 0.0039 degree Celsius. Because temperature can cause an almost 3.9% resistance change with a 10-degree increase, its effect on resistance is not negligible:

$$R(T) = R_0 [1 + \alpha_R (T - T_0)]$$ **(1)**

Operation of a complementary metal-oxide semiconductor (CMOS) transistor, one of the most widely used active devices, is mainly affected by temperature. Therefore, temperature has been one of the main variations for estimating performance of electronic systems with process and voltage, known as PVT (process-voltage-temperature) variations. As temperature increases, semiconductor mobility decreases by lattice scattering, both for

17

n-type or p-type semiconductors. Temperature also varies the threshold voltage of transistor and its drain current, as shown in Equations (1) – (4), where α_μ, α_{VT}, and α are temperature coefficients of mobility, threshold voltage, and drain current, respectively [9], [10].

$$\mu\,(T)=\mu_0\,(T/T_0)^{\alpha_\mu} \tag{2}$$

$$I_d(T)=\mu(T)\frac{W}{L_{eff}}\,P_l\,[V_{gs}-V_T(T)]^{\alpha/2}V_{ds} \tag{3}$$

$$V_T(T)=V_{T0}+\alpha_{V_T}\,(T-T_0) \tag{4}$$

The key objective of this electrical simulation modeling is to estimate signal and power integrity effects, specifically skew and power ground noise, in a single clock domain for 3-D ICs with temperature gradients. The CDN has cascaded buffers and interconnects also called as a "buffered clock tree." The shape of the clock is a symmetrical H-tree with the clock buffers placed every 500 μm on the metal interconnect. The symmetrical structure and cascaded buffer maximize the temperature effect. Figure 11 illustrates the buffered CDN structure with TSVs to connect clock distribution into other tiers. On-chip interconnects are in the resistance dominant range, therefore recent finer technology nodes having higher resistance show a significant effect on the delay of the signal.
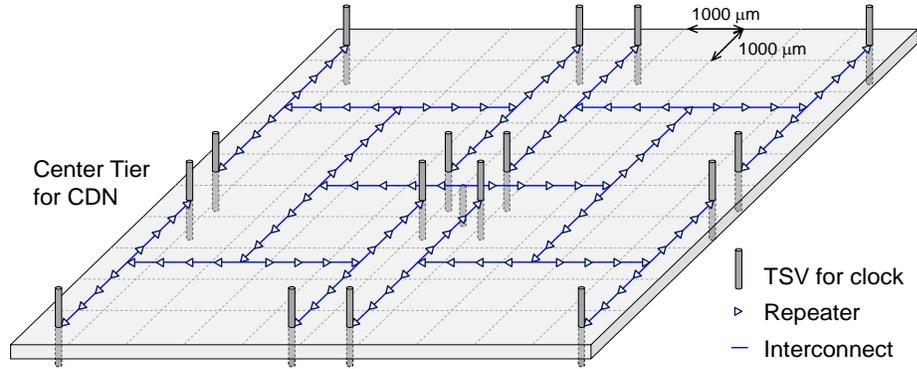
Figure 11   An H-tree CDN structure in the centered die [37].

2.3.1 Electrical Modeling

To build the electrical model for the entire structure, a BSIM4 model of the 45-nm technology node [38], [39] was used for clock repeaters having 2-stage inverter. Initial widths of the buffer are 315/95 nm 630/195 nm with the length of 50 nm referring to a standard buffer-size profile described in [8]. The initial size of the buffer is optimized to have the width of the unit repeater for fan-out 1 (FO1) (for PMOS/NMOS) and fan-out 2 (FO2) was 2.52/0.8 μm (for PMOS/NMOS) to ensure more immunity to variation. The optimized size has a ratio of 3-4 for PMOS and NMOS size and 4-5 and 8-10 times larger than initial sizing for FO1 and FO2, respectively. Simulation results with buffer sizing will be discussed in Chapter 4. Cascaded buffers are connected through metal interconnects, as shown in Figure 12 (a). Lumped RLC values from [40] for the metal interconnect are 30 ohm and 200 fF per mm.
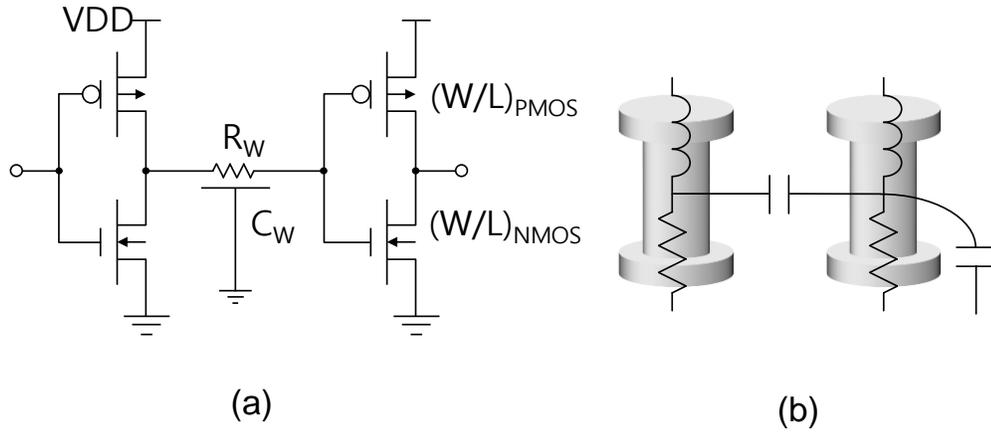
Figure 12   **Unit schematic of the simulation model: (a) CDN having buffers and interconnect and (b) TSV.**

RLC values of clock interconnect in Figure 13 show inductance with RC values, and mutual values of capacitance and inductance. As indicated earlier, clock is the most critical signal in a digital system, thus more room with approaches with wider interconnect width and spacing are required. Optimization of RLC values of clock interconnect helps to reduce temperature effect of the delay.

Dimensions	RLC
W = 1 um	R = 15.068 Ohm
s = 1 um	L = 0.686 nH
l = 500 um	M12 = 0.521 nH
t = 0.73 um	M13 = 0.452 nH
h = 1 um	M14 = 0.412 nH
K = 3.9	(k12 = 0.759
	k13 = 0.658
	k14 = 0.6)
	Cground = 31.5765 fF
	Ccouple = 16.4545 fF
	Ctotal = 96.062 fF

Dimensions	RLC
W = 1 um	R = 15.068 Ohm
s = 3 um	L = 0.686 nH
l = 500 um	M12 = 0.452 nH
t = 0.73 um	M13 = 0.384 nH
h = 1 um	M14 = 0.344 nH
K = 3.9	(k12 = 0.658
	k13 = 0.559
	k14 = 0.501)
	Cground = 40.1445 fF
	Ccouple = 3.4245 fF
	Ctotal = 87.138 fF
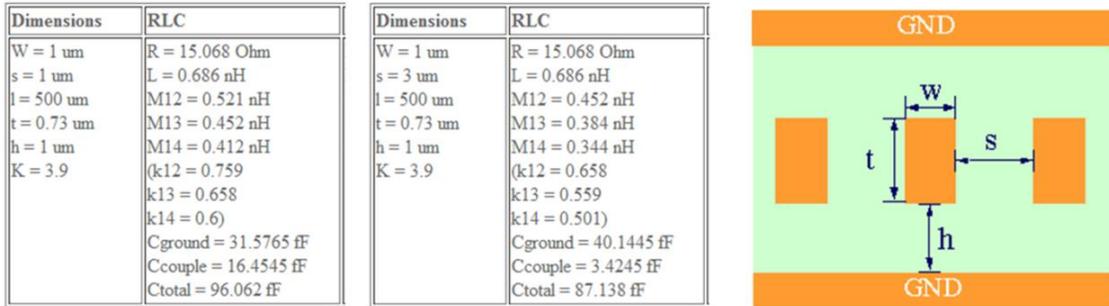
Figure 13   **Dimension and RLC values of clock interconnects [40].**

Resistance of TSVs have been modeled using DC term and AC term including skin effect, as shown in Equation (5) – (8) [41], [42], and [43]. The value of self-inductance, shown in Equation (9), is used rather than mutual values because pitch of TSVs are much longer than its height.

$$R_{TSV} = \sqrt{R_{DC,TSV}^2 + R_{AC,TSV}^2} \; [\Omega] \qquad (5)$$

$$R_{DC,TSV} = \frac{\rho\,l}{A} = \frac{\rho_{TSV}\,h_{TSV}}{\pi\,(d_{TSV}/2)^2} \; [\Omega] \qquad (6)$$

$$R_{AC,TSV} \approx \frac{\rho\,l}{\pi(D-\delta)\delta} = \frac{\rho_{TSV}\,h_{TSV}}{\pi\,(d_{TSV}-\delta_{TSV})\,\delta_{TSV}} \; [\Omega] \qquad (7)$$

$$\delta_{TSV} = \sqrt{\frac{2\rho}{\omega\mu}} = \sqrt{\frac{\rho_{TSV}}{\pi f \mu_{TSV}}} \; [m] \qquad (8)$$

$$L_{Self} = \frac{\mu l}{2\pi}\left[\ln\left(\frac{2l}{R}\right) - \frac{3}{4}\right] = \frac{\mu h_{TSV}}{2\pi}\left[\ln\left(\frac{2h_{TSV}}{d_{TSV}/2}\right) - \frac{3}{4}\right] \qquad (9)$$

W_{PDN} S_{PDN}

W_{PDN}: 10 um, S_{PDN}: 50 um
t_{Metal1}: 1.0 um, t_{Metal2}: 0.8 um
$t_{Insulator1}$: 0.6 um, $t_{Insulator2}$: 4.0 um

(a)   (b)
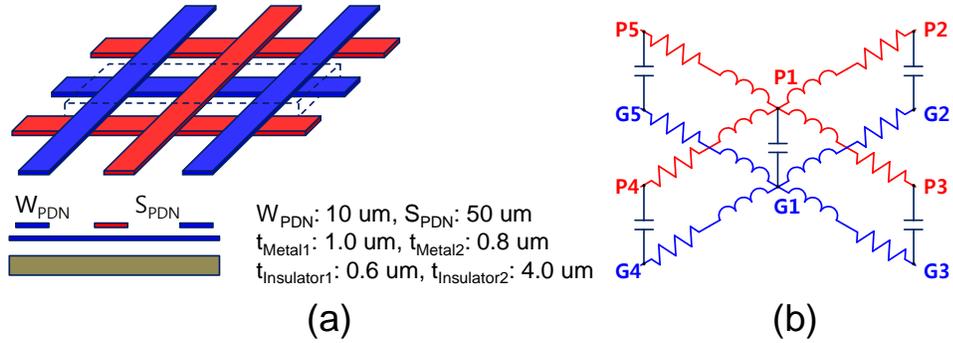
Figure 14   Dimension and unit schematic of the simulation model of meshed-PDN.

Because of the increased power density of 3-D ICs, accurate power distribution network modeling is required for power ground noise estimation. A meshed on-chip PDN model, shown in [44], was used with an on-chip decoupling capacitor model from [45]. To mimic a noisy chip condition, the output buffers of a 64-bit data bus were distributed on

other tiers as simultaneous switching noise sources. With the noise sources, a full model consisting of CDN, TSV, and PDN with decoupling capacitors was used for simulations.

Table 3 shows the dimension and equivalent parasitic values of the CDN, the TSV, and the PDN. For more accurate power integrity estimation, the PDN model includes a voltage regulator module (VRM) providing power to the three-dimensional stacked dies through the PCB, decoupling capacitors, package, and TSVs, which will be further described in the following sections.

Table 3   Dimension and electrical parasitics of unit components [8].

	Width	**Length**	**Height**	**R**	**L**	**C**	**Note**
CDN	1 μm	1 mm	N/A	30.1 Ω	1.4 nH	192 fF	per mm
TSV	5 μm	5 μm	50 μm	61.4 mΩ	29.4 pH	4.0 fF	per TSV
PDN	10 μm	5 μm	1 μm	430 mΩ	22.3 pH	1.7 pF	per mm^2

The objective of the PDN is to provide ideal and stable power for the load circuits. To design a PDN, complete modeling from voltage generator to load circuits, as shown in Figure 15 is required. Input impedance from load circuits and voltage fluctuation at the load points have been simulated for the PDN design.

(a)

(b)
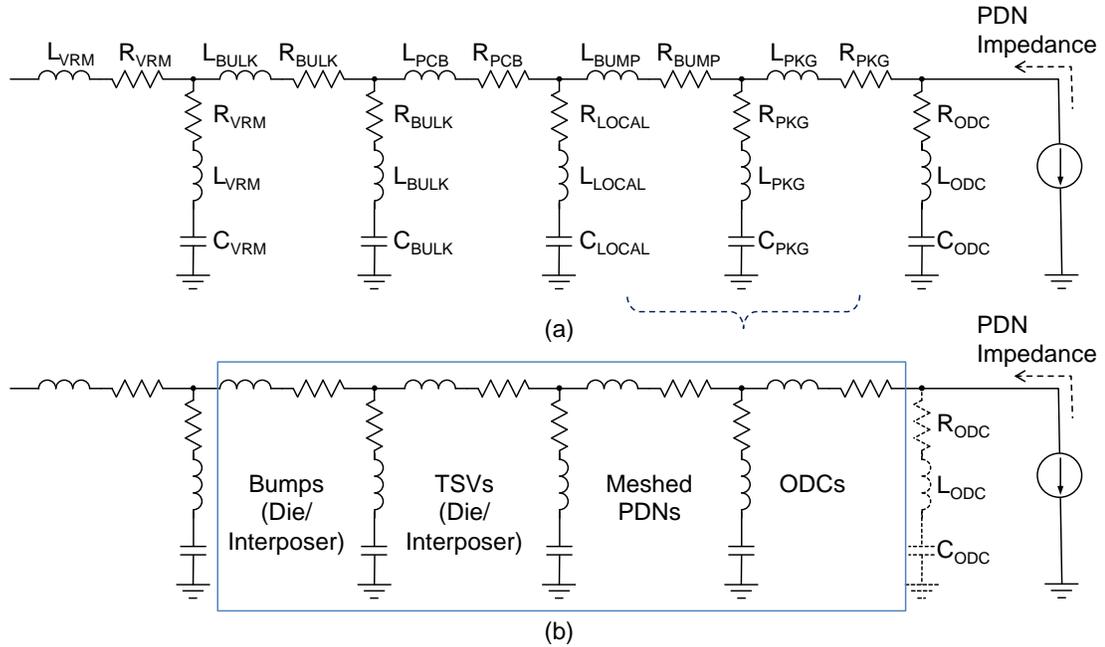
Figure 15   Schematic of the full system power model; (a) conventional model and (b) new model having TSVs and bumps.

An accurate full-system PDN model is critical for the analysis. In this research, the off-chip decoupling parasitics for the PDN model were chosen from [46]. Table 4 shows electrical parasitics of the system-level PDN. From a thermal-effect perspective in PDN design, equivalent series resistance (ESR) parasitics are more temperature dependent than decoupling capacitor parasitics because temperature-dependent resistance. The manufacturing class also describes temperature dependency of capacitances of off-chip decoupling capacitors. The temperature dependency for ESR and capacitance changes the PDN impedances. Resistance and transistors characteristics also have temperature dependency as previously described in Chapter 1.

Component	C [F]	R [Ω]	L [H]	Quantity	Note
VRM	3.3m	60m	17n	1	
Bulk	4.7u	16m	2.1n	2	
PCB Plane		s-parameters			
Decap	100n	20m	0.8n	10	
Package	1p	50m	1.5n		
TSV	4.0f	61.4m	29.4p	40	
Meshed PDN	1.7p	430m	22.3p	100	mm^2
On-die decap	3p	100	N/A	16	unit

2.3.2 Thermal Modeling

The thermal modeling for electrical systems was initiated with the thermal resistance model shown in Figure 16. The model exploits conduction heat transfer to address thermal dissipation. The heat resistances from heat sources to ambient are in parallel as shown in Figure 16 (b).

(a)   (b)
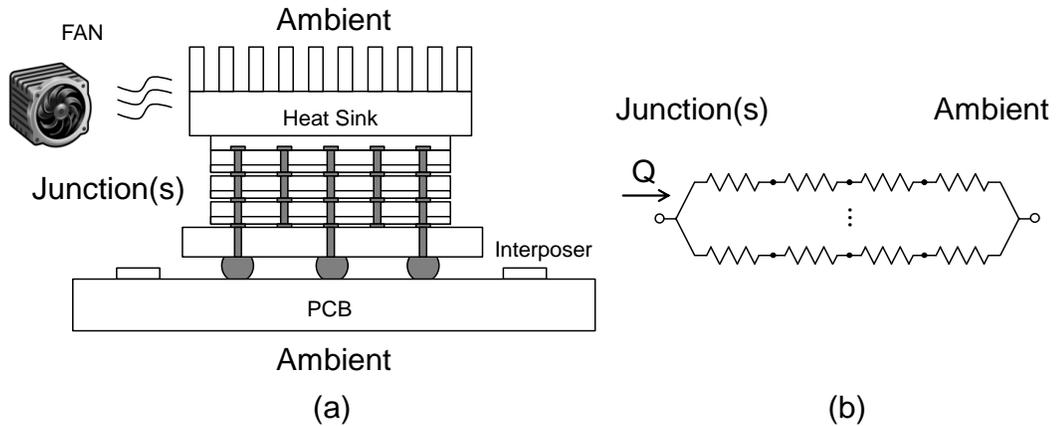
Figure 16 **Equivalent thermal resistance of the structure.**

In this study, convection condition with air flow from a fan was assumed for the system level cooling. The solver used [13] requires not only electrical excitations in the form of voltage and current sources, but also for thermal that includes power sources and boundary conditions to enable electrical-thermal simulations. The detailed material parameters including thermal conductivity and coefficients are listed in Table 5. Ambient temperature of 25 degree Celsius and fan speed for air convection condition are also shown in the table.

<center>Table 5   Material parameters for thermal model.</center>

	Unit	Value	Note
Ambient temperature	°C	25	T_A
Air convection	W/(m² ·K)	20	fan speed
TIM conductivity	W/(m ·K)	1.2	
Under-fill material	W/(m ·K)	4.3	

To analyze the system structure and obtain thermal profiles, a solver using the finite volume method [13] was used. The finite volume formulation assumes nodes surrounded by finite volume cells and applies the finite difference approximation, by solving the voltage distribution equation and heat equation. Therefore, the solver can be used to compute distributions of voltage, current, and temperature with Joule heating effect. For steady-state thermal analysis, heat equation for solid medium can be expressed as:

$$\nabla \cdot [k(x, y, z)\nabla T(x, y, z)] = -P(x, y, z) \tag{10}$$

where k and T are thermal conductivity of medium and temperature distribution. Using Equation (10) based on conduction, the finite volume formulation along with the convection boundary condition can be expressed as in Equation (11) and (12).

$$\frac{T_{i,j} - T_{i,-1\,j}}{\frac{\Delta x_1}{kd}} + \frac{T_{i,j} - T_{i+1,j}}{\frac{\Delta x_2}{kd}} +$$

$$\frac{T_{i,j} - T_{i,j-1}}{\frac{\Delta y_1}{kw}} + \frac{T_{i,j} - T_{i,j+1}}{\frac{\Delta y_2}{kw}} = P_{total} \qquad (11)$$

$$k\frac{\partial T}{\partial n}\bigg|_{convection} = -h_c\,(T - T_a) \qquad (12)$$

where P_{total} is the total heat source in a cell, and T_a and h_c are ambient temperature and convention coefficient for the convection boundary, respectively.

The solver is capable of stretching the mesh between multiple regions with multiple materials such as die, interposer, and package by using a non-uniform grid. Figure 17 shows the flow of the electrical-thermal solver used.
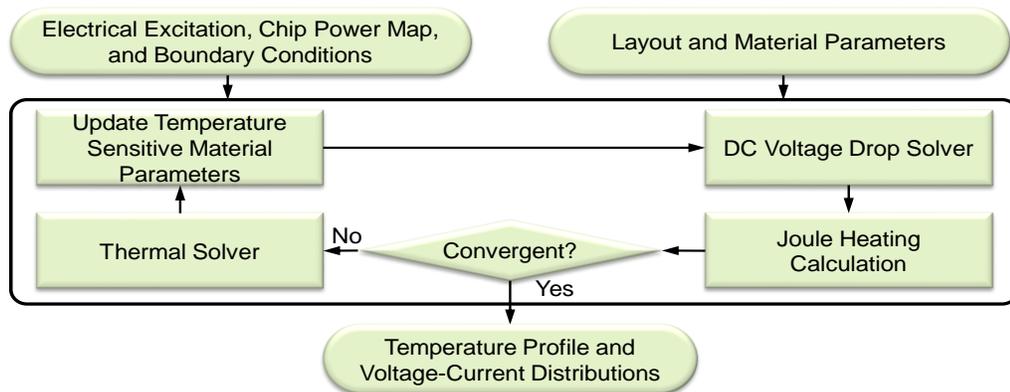
Figure 17 Flow of electrical-thermalsolver [13].

26

The solver was used to simulate the 3-D configuration with randomly distributed power maps on the two layers (top and bottom die). Total power of the three dies was 50 W which is 20 W for the bottom and the top die respectively, and 10 W for the center die for CDN. The used sample power maps were generated to reflect varying temperature gradient cases. Figure 18 shows the generated power maps for electrical-thermal simulations.
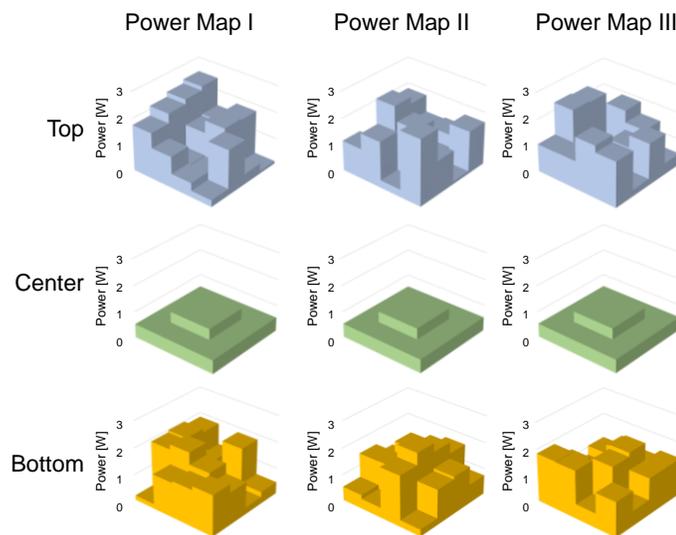
Figure 18 Power maps for thermal simulations.

2.4 Electrical-thermal Simulation

2.4.1 Simulation Results

Electrical simulations to estimate clock skew and PDN noise are feasible with electrical-thermal modeling described in the previous section. The operating condition assumption for initial simulation shown in Figure 19 uses 1.1 V operation. The initial results show that the base condition of an ideal PDN has a skew of 30.7 ps. The addition

of PDN and temperature effects increases the skew by 19.2 ps (62.5%) and 143.6 ps (467.8%), respectively. Additional skew of 20.3 ps (68.2%) occurs with temperature gradient superimposed on the PDN. The amount of total skew shown in the results present summation of additional skew portions by PDN noise and thermal [8].
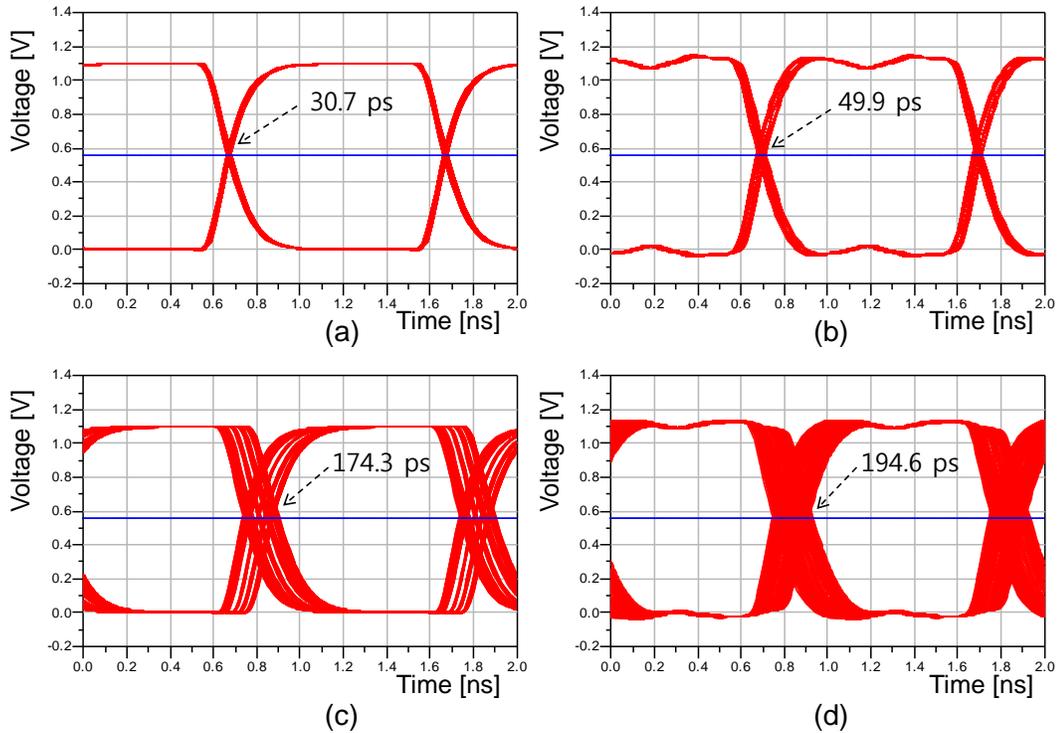
Figure 19 **Transient simulation waveforms at clock ends (a) without the PDN model, (b) with the PDN model, (c) with the thermal gradient and (d) with both the PDN model and the temperature gradient [8].**

The initial simulation results show the thermal and PDN impact. The simulation condition was further updated for reflecting more recent technology node, as shown in Figure 19. Repeater sizing described in Section 2.3.1 show a reasonable range of skew and thermal variation as shown in Figure 20. Delay of the clock tree varied with the global temperature (25 degrees to 125 degrees), as shown in Figure 20 (a), but thermal-induced skew, shown in Figure 20, caused by temperature gradient in a single die reached up to 96 ps, which is not a negligible number for global skew.
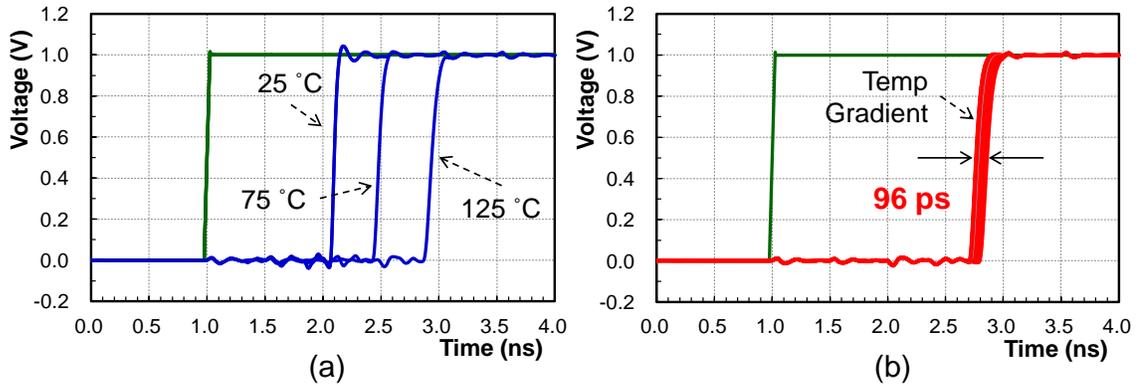
Figure 20 Transient simulation waveforms at clock ends: (a) without the temperature gradient (@ 25˚C, 75˚C, and 125˚C) and (b) with the temperature gradient [37].

2.4.2 Analysis of Delay

Capacitor components in CMOS integrated circuits include gate-drain, diffusion, wire, and gate structures [47]. Both intrinsic capacitance and overlap/fringing capacitance need to be analyzed because they are all related with body bias, short channel, drain-induced barrier lowering (DIBL), overlap, and fringing. As Figure 21 indicates, many components of capacitances are present in the buffer and interconnect model. Table 6 shows parasitic capacitance components and values of the model used.
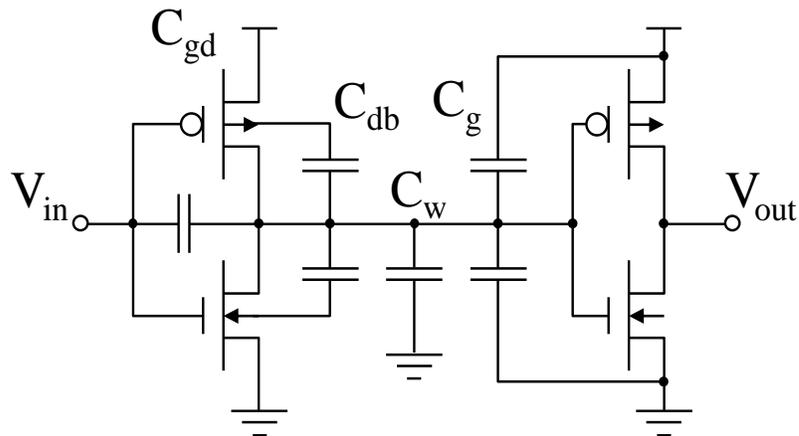
Figure 21 Capacitance of buffer and interconnect model.

Table 6   **Capacitance components in a CMOS inverter.**

Parasitics		Expression	NMOS	PMOS	Note
Gate-Drain Cap	C_{gd}	2*CGDO*W	0.23	0.61	Overlap cap
Diffusion Cap (H-L)	C_{db}	k_{eq}AD*CJ + K_{eqsw}PD*CJSW	0.66	1.50	Junction cap (Bottom, Side-wall)
Diffusion Cap (L-H)	C_{db}	k_{eq}AD*CJ + K_{eqsw}PD*CJSW	0.90	1.15	Junction cap (Bottom, Side-wall)
Wire Cap	C_w	Extraction from layout	0.12	0.12	Negligible
Gate Cap	C_g	(CGDO+CGSO)* W + C_{ox}*W*L	0.76	2.28	
Total Load Cap	C_L		1.77 / 2.01	4.51 / 4.16	H-L / L-H

In this research, analytical analysis based on the Elmore delay model has been performed to estimate propagation delay of RC interconnects [47]. It calculates propagation delay based on Equation (14).

This section shows analytical calculations used to compute the propagation delay. The calculation comes from the Elmore delay model, which has proven to be an efficient method for analyzing RC interconnects [47]. The propagation delay, which is comprised of inverters and interconnects, can be divided into four parts, as shown in Figure 22.

$$t_p = 0.69 R_{DR} C_W + 0.69 R_{DR}(C_O + C_I) + 0.38 R_W C_W + 0.69 R_W C_I \qquad \textbf{(13)}$$

Combinations of resistance and capacitance from the inverter and the wire create resistor-capacitor (RC) delay components. The input/output capacitance (C_I and C_O) and wire resistance (R_W) have relatively smaller values than the capacitance of the wire (C_W) and the resistance of the driver (R_{DR}). Therefore, the latter significantly increases delay but it is closely dependent on temperature. The resistance of copper interconnects has a linear

dependency on temperature while the capacitance built on a silicon substrate has negligible dependency and is comparatively stable with temperature variations.

$$0.69R_{DR}C_W$$

(a)

$$0.69R_{DR}(C_I+C_O)$$

(b)

$$0.38R_WC_W$$

(c)

$$0.69R_WC_I$$

(d)
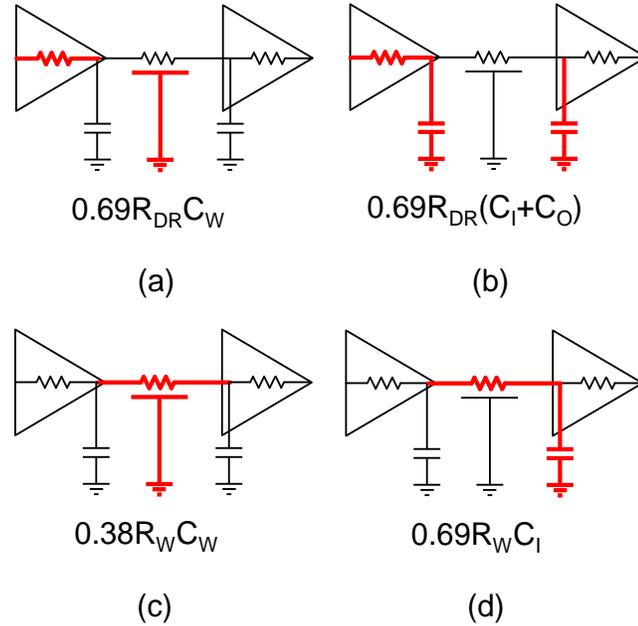
Figure 22   Four parts of the propagation delay of a CDN: (a) an inverter driving wire cap, (b) an inverter, (c) wire delay, and (d) a wire driving an inverter [8].

RC delay of the CDN has been divided into four parts. As shown in Figure 22, each part is defined as a combination of resistance and capacitance with an inverter or a wire. Capacitance of the wire (C_W) and resistance of the driver (R_{DR}) provide the largest impact on delay, based on the calculations shown in Table 7.

Table 7   Delay values with buffers and wire [8].

	Formula			Values			Delay [ps]	Note
	Network	**R**	**C**	**Coeff.**	**R [Ω]**	**C [F]**		
(a)	Lumped	R_{DR}	C_W	0.69	1.25k	100f	86.3	82.7 %
(b)	Lumped	R_{DR}	C_O+C_I	0.69	1.25k	20f	17.3	16.6 %
(c)	Distributed	R_W	C_W	0.38	15	100f	0.6	0.6 %
(d)	Lumped	R_W	C_I	0.69	15	10f	0.2	0.2 %

31

The values for the delay shown in Table 7 also vary with temperature. Figure 23 shows temperature dependency of the four parts shown above. Resistance of copper has a linear dependency on temperature with a coefficient of 0.0039 per degree Celsius. However, thermal dependency of capacitance built on a silicon substrate is negligible and the value is very stable with a small temperature coefficient. From the figure it can be concluded that RC delay has a linear relationship with temperature.
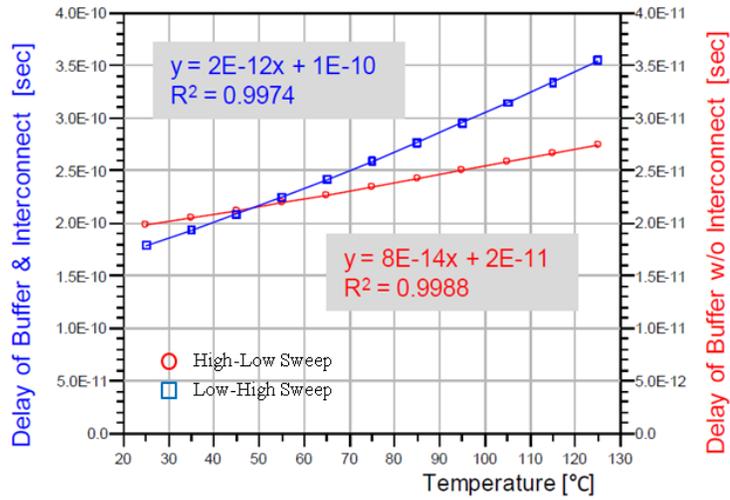
Figure 23   Temperature effect on delay [8].

Temperature coefficient of interconnect resistance and lumped electro-thermal modeling with thermal effects have been discussed in [48]. The formula shown in [48] has thermal variation for interconnect. In this research, the formula is extended to include driver resistance along with capacitance of the buffer, as shown in Equation (15).
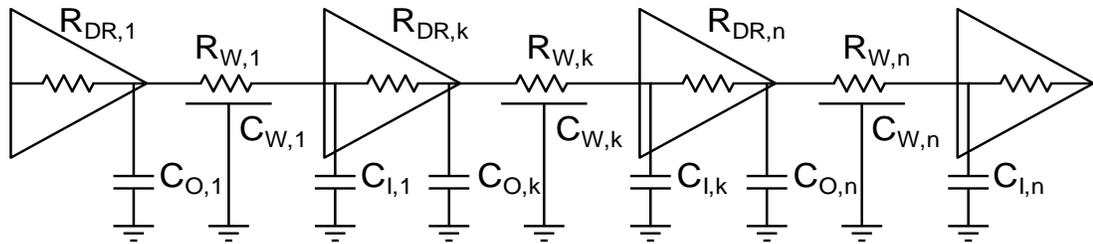
Figure 24   Cascaded buffers and interconnects [8].

Therefore, Equation (15) provides a delay (D) for the clock network comprised of cascaded buffers and interconnects. Equations (16) and (17) capture the effects of temperature on buffers and interconnects, respectively. In Equation (15), the constants for the RC delay are 0.69 ("a" for lumped networks) and 0.38 ("b" for distributed networks), respectively, for calculating the propagation delay [39].

$$D = a \sum_{i=1}^{n} R_{DR,i} \left(C_{O,i} + C_{W,i} + C_{I,i} \right) + \sum_{i=1}^{n} R_{W,i} \left(b C_{W,i} + a C_{I,i} \right) \tag{14}$$

$$R_{DR,k} = R_{DR,0} \left(1 + \beta_{DR} \left(T_k - T_0 \right) \right) \tag{15}$$

$$R_{W,k} = R_{W,0} \left(1 + \beta_W \left(T_k - T_0 \right) \right) \tag{16}$$
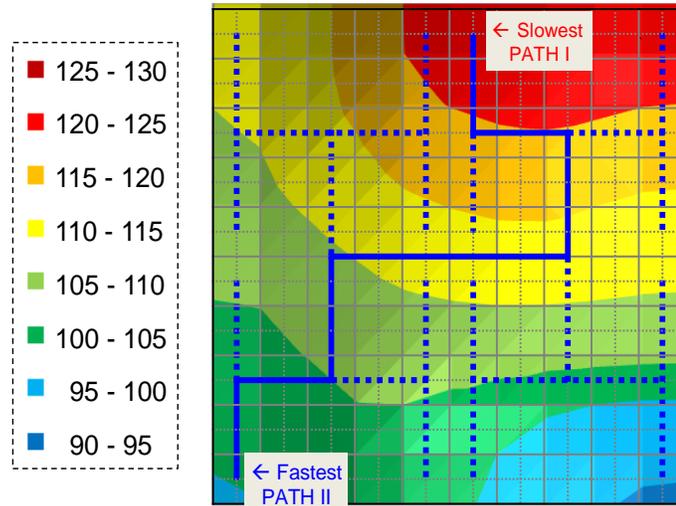
Figure 25   A clock tree with a temperature gradient [37].

Thermal simulation provides a temperature profile, as shown in Figure 25, which can impact the CDN based on location. The simulated temperature profile is superimposed on the electrical model to generate a delay profile. Figure 26 (a) shows temperature along

with the CDN position from center to four edges. Equation (15) captures temperature dependency. Using Equation (15), correlations between calculations and simulations from the previous section are shown in Figure 26 (b).
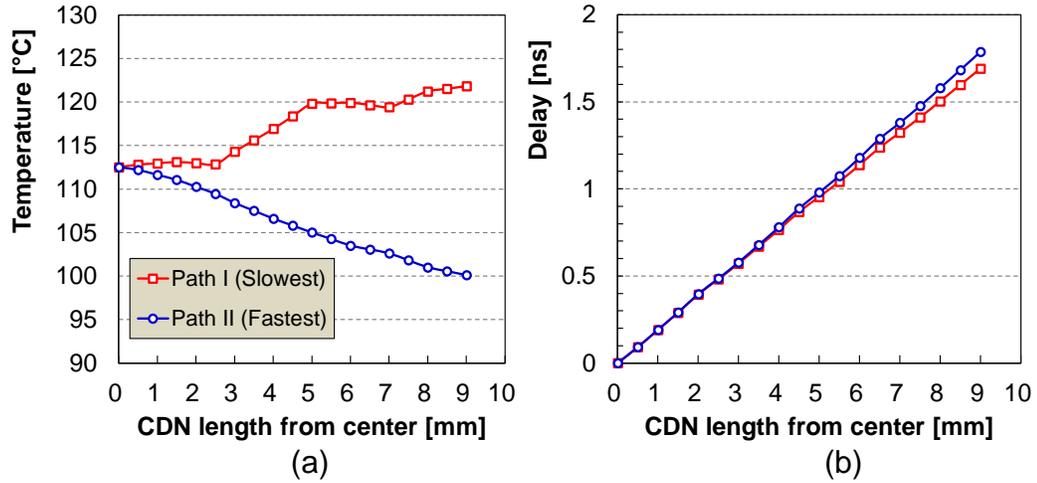
Figure 26   **(a) The temperature profile for delay calculation. (b) Comparison of path delay with calculation [34].**

2.5 Thermal Impact on Power Delivery Network Design

The focus of this chapter is on the impact of temperature effect on power integrity (PI). PI estimates the quality of power delivery, including the non-ideal effects of power distribution network (PDN). The goal of PDN design is to lower PDN impedance below a target value, also called as target impedance. In real applications the actual impedance is increased by parasitics such as resistance and inductance at higher operating frequencies [49], as shown in Figure 27. As discussed in the previous chapter, because of higher temperature, increase of ohmic loss (ESR; Equivalent Series Resistance) increases impedance of the on-chip decoupling capacitors. Therefore, variations of resistance are critical for the design of the power distribution network.
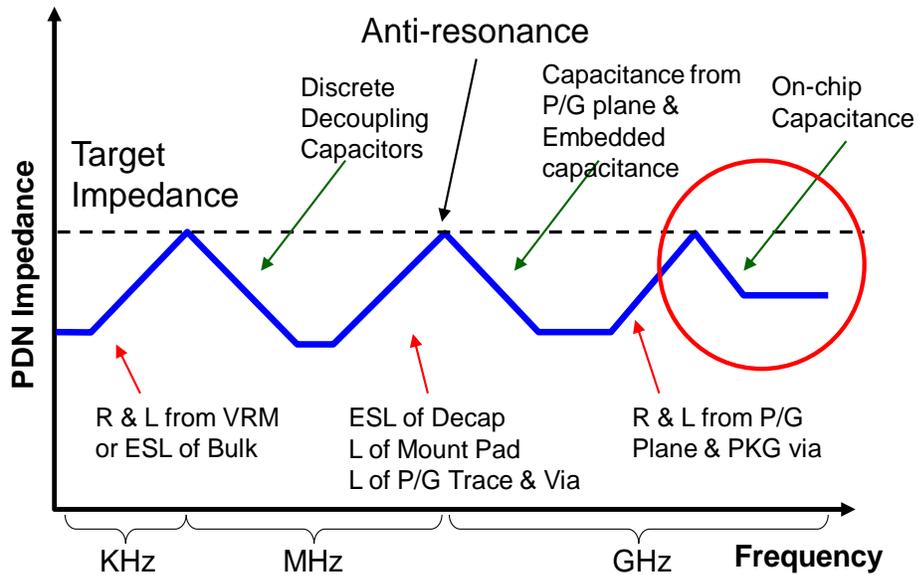
Figure 27   Target impedance and PDN response; modified from [49].

For PDN simulation, both frequency and time domain simulations were performed for calculating impedance profile and voltage transient, where the same simulation flow discussed in the previous chapter was used. With the same operating condition described in the previous chapter, a clock signal operating at 500 MHz with supply voltage of 1.0 V was used for transient simulations using a PDN model connected to an off-chip voltage regulator module. The results of the simulations illustrate that the PDN impedance increases as the temperature rises.
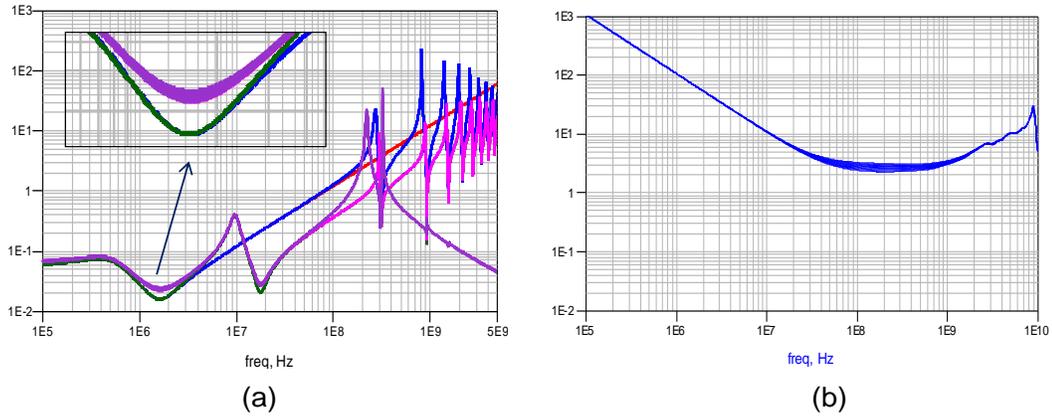
Figure 28   **Impedance of PDN with temperature variation; (a) off-chip and (b) on-chip.**

Delay variation caused by temperature gradient in 3-D ICs were investigated in the previous section. PDN impedance also has a linear dependency with temperature because the resistance of PDN changes linearly with temperature. Resistance of PDN affects ESR (Equivalent Series Resistance) of a full PDN. The PDN impedance of the full PDN model in Figure 28 (a) shows variation with temperature. The lowest point in impedance curve is determined by the ESR of decoupling capacitors which is affected by temperature. It is important to note that the ESR is dependent upon not only temperature gradient but chip global temperature, as compared to the skew effect which is affected by temperature gradient condition only.

In this section, temperature dependency of electrical resistance is the main focus. On-chip PDN and on-chip decoupling capacitors (ODCs) have larger series resistance as compared to decoupling capacitor in the package. A larger number of off-chip decoupling capacitors, such as Bulk or SMT capacitors, are used for achieving small series inductance and resistance. However, the absolute value of ESR of ODC is much larger and the number of ODC is often limited for overcoming the ESR values. Therefore, temperature dependency of ODC's ESR is more dominant than off-chip decoupling capacitors, as

shown in Figure 28 (b). Variations of simulated impedances illustrate that ODCs are operating in the ESR dominant range. Transient simulation results in Figure 29 show how the effect ESR on noise waveform. The two waveforms (with and without ODC) show a different amount of noise, but they show different temperature dependency.
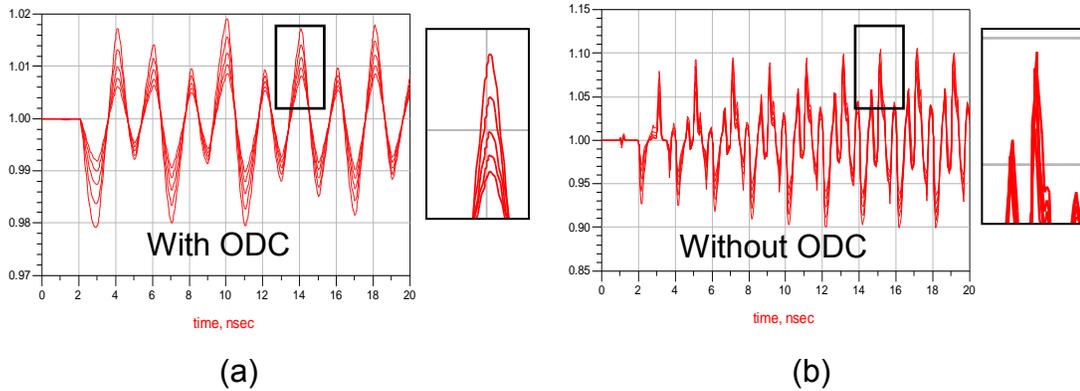
Figure 29   **PDN noise with temperature variation (a) With ODC. (b) Without ODC.**

2.5.1 On-Chip Decoupling Capacitors

Important elements in a PDN design are the various types of decoupling capacitors (off-chip or on-chip) that can be used. Traditionally, off-chip decoupling capacitors have been used to optimize PDN designs. By contrast, on-chip decoupling capacitors are mainly design-dependent because multiple types of ODCs are available [50]. The various types of ODCs have implementation variations [51]. In this chapter, the effects of temperature on PDN design are estimated and quantified since thermal variations of PDN using various decoupling capacitors have not yet been fully analyzed in the open literature [52].

In PDN design, decoupling capacitors are used to lower impedance looking into the power supply. In this chapter, on-chip decoupling capacitors are the main focus because ODCs are more resistance (ESR) dominant because of lower inductance. The values of

ESR mainly depend on the types of ODC and their implementation. Regardless of the implementation of ODCs, the values of resistance are mainly dependent on temperature. This section discusses the various types of ODCs, their parasitics for various structures, for example metal-oxide semiconductor (MOS), metal-insulator-metal (MIM), or poly-insulator-poly (PIP) [51], as shown in Figure 30.
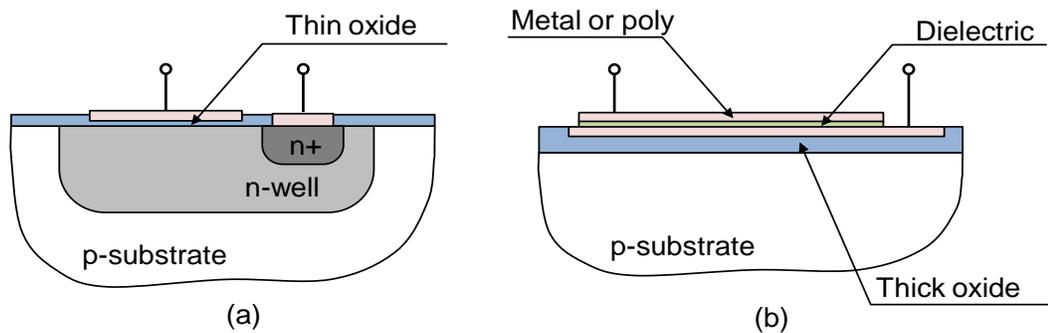
Figure 30   **Example structures of ODCs: (a) MOS, (b) MIM, and PIP [51].**

The most widely used types of ODCs are MOS-based capacitors because they are fully compatible with CMOS circuits and have larger capacitance with simpler implementation. With the trend toward low power consumption, the leakage of MOS capacitors becomes an important design parameter. The leakage decreases with increase in oxide thickness. Thus, the thickness of a MOS capacitor can be optimized for less leakage with the use of thin- and thick-oxide capacitors [53]. In addition, gated- and active-decoupling capacitors have been presented in the open literature to reduce the leakage current and power-ground noise [54], [55], and [56]. Figure 31 shows schematics of the various MOS capacitors.
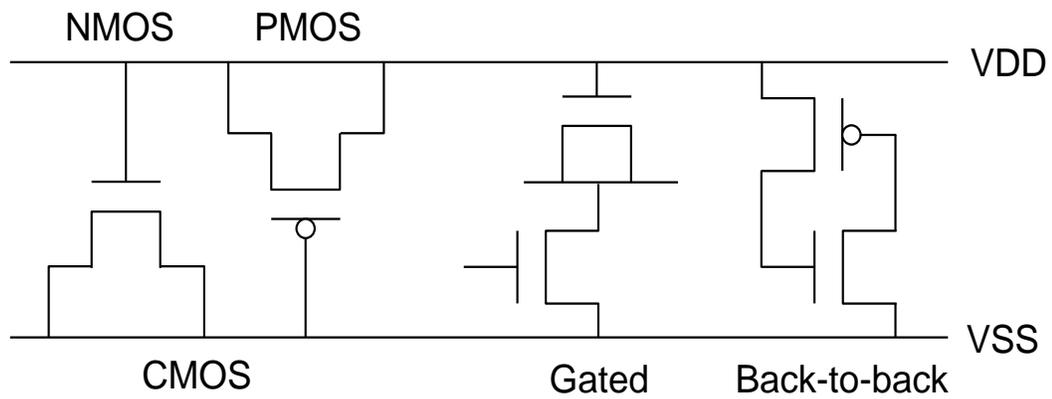
Figure 31   Variations of the MOS capacitor [52].

Other types such as MIM and PIP are also available for ODCs, as shown in Table 8. MIM capacitors provide precise capacitance with a relatively lower temperature coefficient [57], and larger capacitance is achievable with a higher permittivity of the dielectric [58]. Deep-trench capacitors with a vertical structure are also available for ODCs using TSV technology [59]. In contrast to MIM and PIP capacitors, various design variations of MOS capacitors have been presented in [50]. The purpose of these approaches is for improved area-efficiency and power-efficiency and less leakage. Optimization of design, including number of fingers, is feasible by modifying the layout of decoupling capacitors [60], with various design implementations and verifications presented in [61].

Table 8   Types of on-chip decoupling capacitors (ODCs).

Type	Description
NMOS	NMOS decap
NMOS_TH	NMOS with thick oxide
NMOS_LVT	NMOS with low Vt
NMOS_ACC	NMOS with accumulation mode
PMOS	PMOS decap
CMOS	NMOS + PMOS
NMOS_GATE	Gated NMOS decap
CMOS_B2B	Back-to-back decap
MIM	MIM decap
PIP	PIP decap

Previous studies have provided modeling of parasitic resistance in MOS-based decoupling capacitors [62] which included gate-oxide leakage [63] and substrate effect [64]. Layout of decoupling capacitors determines detailed values of their parasitics. However, the parasitics are also determined by the structure of decoupling capacitors. Characteristics of materials provide various range of parasitics, as shown in [51]. Figure 32 shows a comparison of parasitics of the three types of decoupling capacitors, namely MOS, MIM, and PIP.
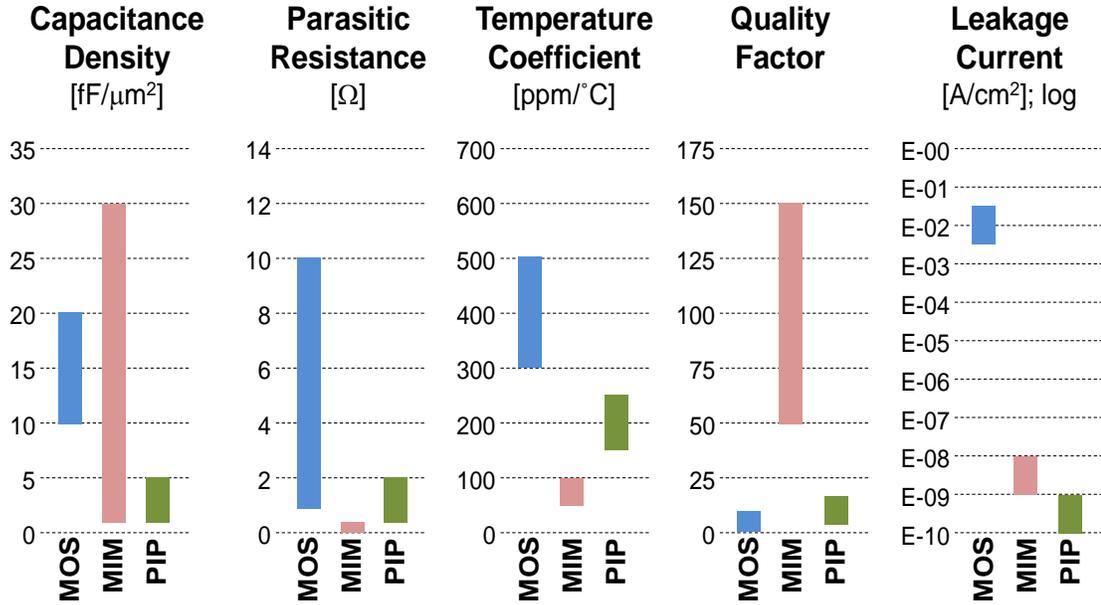
Figure 32   Comparison of on-chip decoupling capacitors [52].

The three types of MOS decoupling capacitors have been chosen in this section because MOS-based capacitors have variations in capacitance density and leakage current. In addition, MIM type capacitor was also chosen for comparison, because MIM and PIP have similar structure. The values of the parasitics, such as capacitance, ESR [65], and temperature coefficients [51], were chosen to estimate the PDN response. Parasitics have various resistance and temperature dependencies. Each type of decoupling capacitor has capacitance density and parasitic resistance, so impedance of decoupling capacitors varies with the type used.

Using the temperature variations of the ODCs, the temperature-aware PDN design for 3-D ICs and systems was investigated. Temperature-aware power integrity simulations were used to quantify the degradation caused by high temperature. Sensitivity analysis was used to verify the impact of various components such as meshed PDN, TSVs, and on-chip decoupling capacitors in 3-D structures.

2.5.2 PDN Simulation with Temperature Effects

To simulate the PDN impedance and voltage transient, a 3-D PDN model using segmentation method [66] was constructed, as shown in Figure 33. All components of the model are segmented and cascaded with parasitics included, as presented in the previous section. The parasitics of decoupling capacitors and current excitation points are distributed across the entire chip area using the temperature profile in [8].
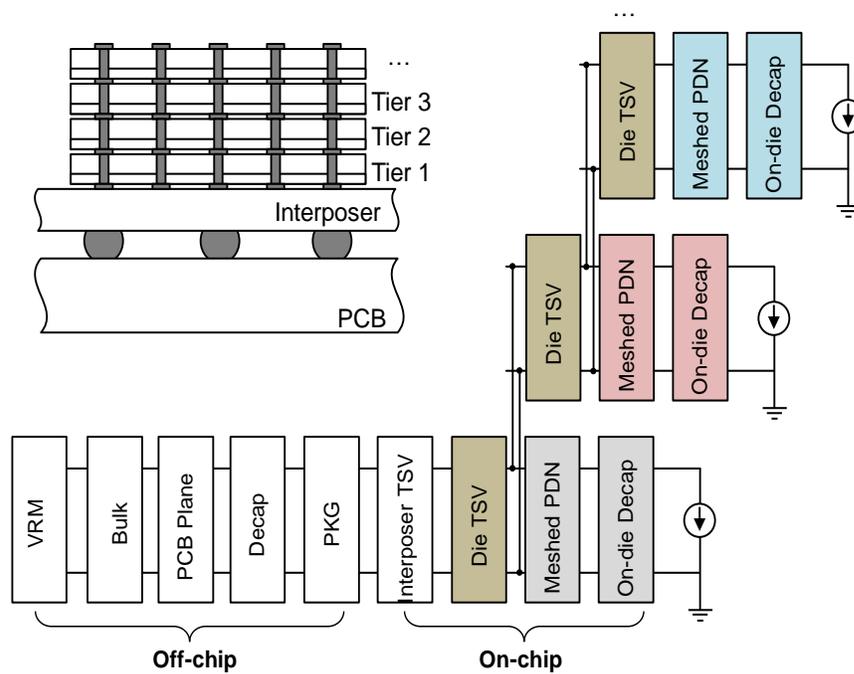
Figure 33   Block diagram of a 3-D PDN.

In addition to signal integrity being affected by temperature, power integrity, especially impedance of the PDN is affected by temperature and gradients as well, as shown in Figure 34 (a). Variations in impedance plot represent the location of probe points across a die. Figure 34 (b) shows impedances at probe points modeled at 1 GHz. Figure 34 (c) also illustrates the difference of impedance measured at a frequency of 1 GHz. The maximum difference is ≤ 0.1 ohm, which is caused by temperature changes.
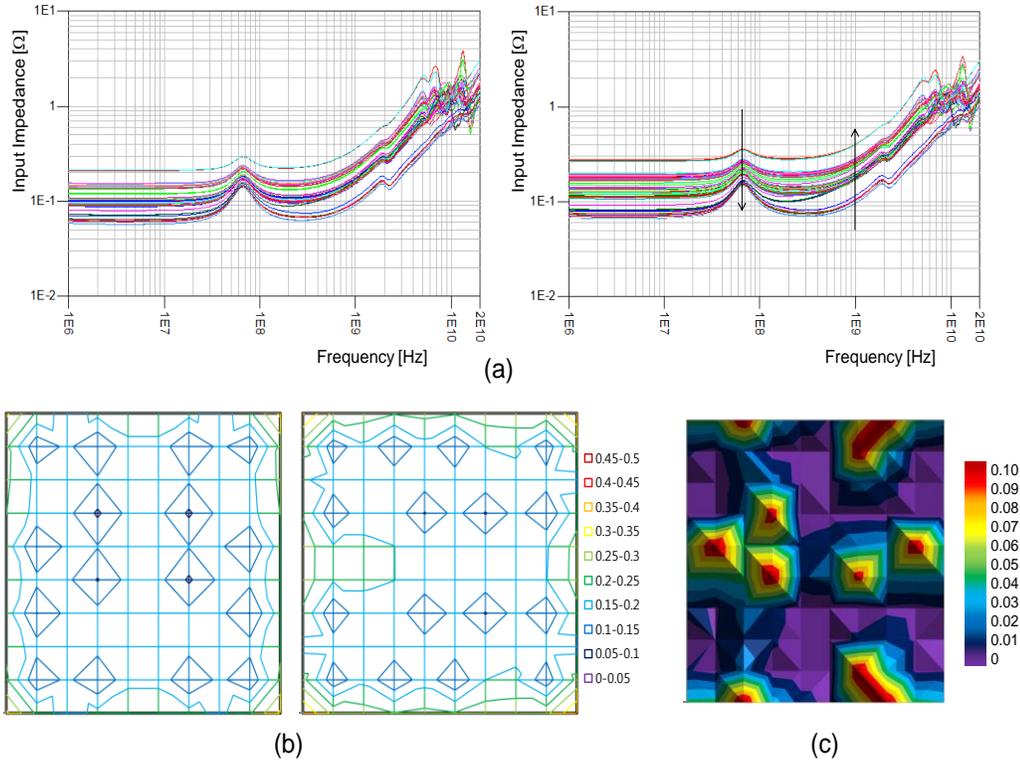
Figure 34 **Simulated PDN impedances with and without temperature gradient: (a) impedance, (b) impedance distribution measured at 1 GHz, and (c) impedance difference measured at 1 GHz [52].**

To estimate voltage transient, current sources were distributed to represent current excitation points. A piece-wise linear (PWL) current source with a linear approximation [51] was used with a SPICE simulated load current of a sample buffer. In Figure 35, I_{max} and I_{min} show the maximum and the minimum current of the excitation points, respectively. In the figure, T_r, T_f, and T_c represent rise, fall, and cycle time, respectively, where T_r and T_f depict current changes vs time based on driver and loading capacitance, respectively, and T_c related to the operation frequency of the circuit. In this simulation, 50 ps, 150 ps, and 1 ns were used for T_r, T_f, and T_c, respectively, for 1GHz operation.
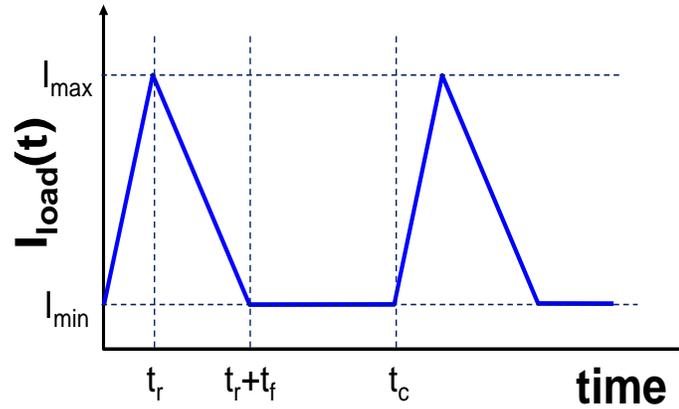
Figure 35   Current excitation for transient simulation [52].

Time transient simulations using the model and current excitation at 1 GHz show PDN noise with temperature effect, as shown in Figure 36. The difference of noise, which is 15.4 mV peak-to-peak (10.8% of voltage swing without noise), represents the effect of temperature on the noise of the power supply.
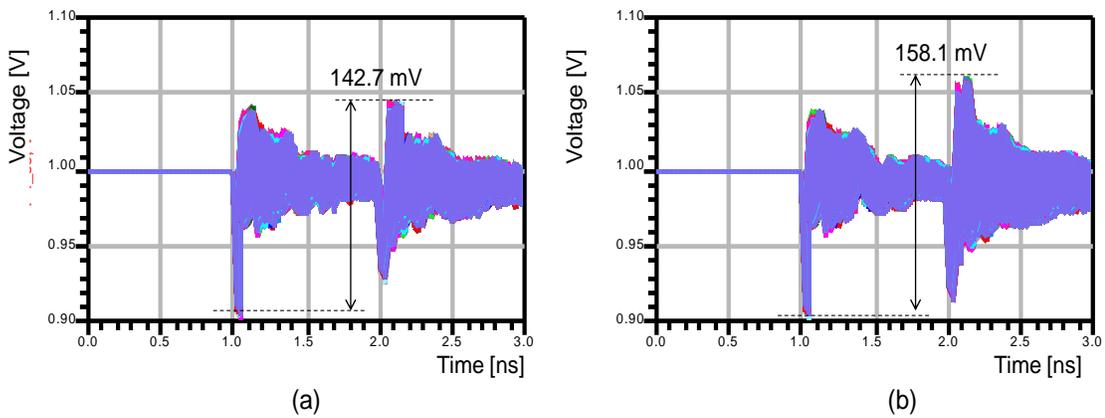
Figure 36   Simulated PDN noise: (a) without temperature gradient and (b) with temperature gradient [52].

2.5.3 Case Study

To quantify the impact of temperature on PDN noise, two 3-D applications were chosen [67]-[68]. The chosen applications are for high-frequency memory and include a 3-

D-stacked double data rate generation 3 (DDR3) memory and high-bandwidth memory

(HBM). Both applications shown in Figure 37 are emerging applications.
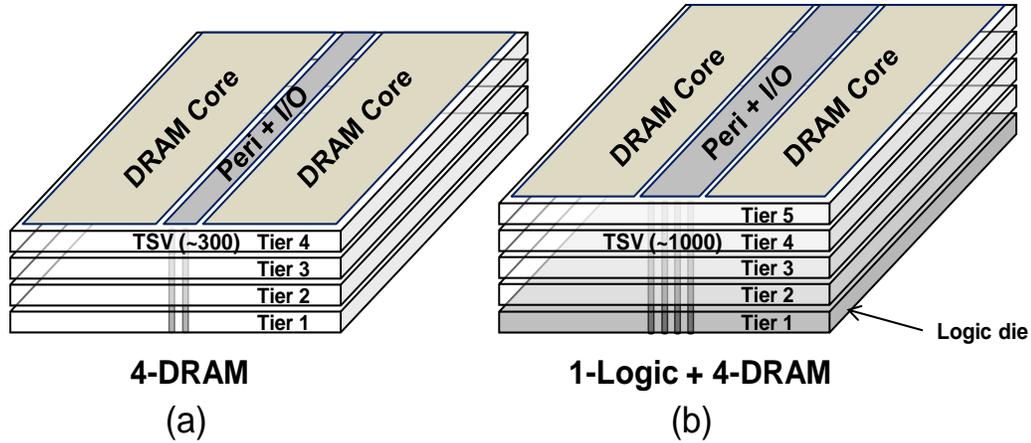
Figure 37   **3-D IC applications for high-speed memory products: (a) DDR3 [67] and (b) high bandwidth memory (HBM) [68]; Modified from [52].**

These applications have combinations of organization, floor planning, decoupling

condition, operation voltage and frequency, as described in Table 9. With the variations of

designs, this case study can cover organization (homogeneous dies or heterogeneous dies)

and constraints (floor planning constraint, area and power constraint).

Table 9   **Comparison of PDN design cases [52].**

	Case I [26]	**Case II** [27]	**Note**
Application	DDR3	HBM	
Publication Date	2010	2014	
Process	50 nm	29 nm	
Organization	4-DRAM	1-logic + 4-DRAM	
# of TSVs	~ 300 + P/G	~ 1000 I/O + P/G	
VDD	1.5 V	1.2 V	
Data Rate	1.6 Gbps	1.0 Gbps	
Chip size	10.9 x 9.0	6.9 x 5.1	mm^2

Both have similar structures such as a main area for the DRAM core, and center area for input/output (I/O) and peripheral circuits. With the different design constraints, decoupling area were allocated for DRAM dies and logic, and distributed current excitation points were also used for both cases. Figure 38 shows the locations of decoupling capacitors and TSVs for the simulations.
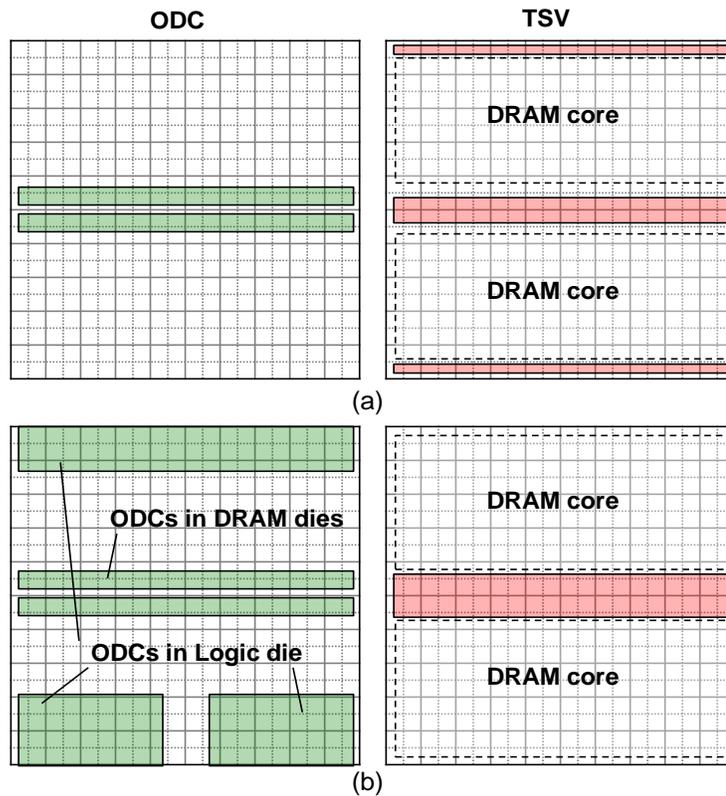
Figure 38   Locations of on-chip decoupling capacitors for: (a) DDR3 and (b) HBM [52].

Four types of decoupling capacitors for these two cases were implemented and the impedances were simulated and compared. The PDN impedances were changed with types of decoupling capacitors and temperature-dependent parasitics, as shown in Figure 39. It can therefore be concluded that temperature-aware PDN design is required for 3-D systems that have high temperature density. Figure 39 (a) shows total capacitance varied with type

of decoupling capacitors, and Figure 39 (b) shows temperature dependency for each type of decoupling capacitors.
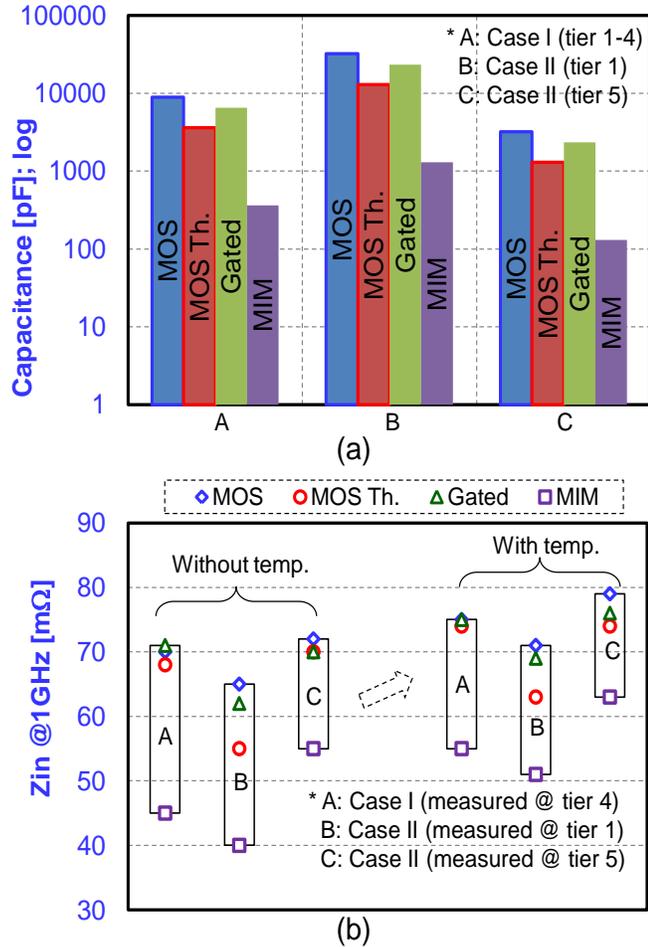
Figure 39   **A comparison of four ODC types: (a) total decoupling capacitance and (b) PDN impedance @ 1GHz [52].**

2.6 Summary

This chapter describes and discusses the system configuration for the entire dissertation. The target system has physical structure with stacked dies and interposer on a PCB. The system also has electrical and thermal parameters, power maps, and target circuits such as clock and power distribution. Regarding modeling and simulation, differences from prior published work are as follows: First, the amount of degradation

caused by temperature on SI/PI of 3-D ICs and systems has not been fully quantified in previous work. Previous work tried to merge signal and power integrity analysis or add thermal integrity into either signal or power integrity analysis to analyze the effect. In contrast, this dissertation proposes inclusion of thermal integrity (TI) into merged SI and PI analysis, which has been quantified and analyzed using the system level simulation model in this work.

This chapter focused on the undesired effects of thermal gradients on the clock distribution networks (CDN) in a three-dimensional (3-D) IC. The state-of-the art integrated circuit boasts of more than a billion transistors on a single die. The advancement is achieved through technologies like System-on-Chip and System-in-Package which feature heterogeneous integration, improved power consumption, a small form factor and reduced production cost. However, heat management remains a concern and leads to hotspots and thermal gradients that can affect the performance of CDN, especially for 3-D ICs.

This work was done using a full design flow with three dies stacked on each other with the CDN built in the center die. A new CDN structure was proposed for 3-D ICs where a full set of temperature-aware simulations were performed to show the effect of varying temperature on the CDN. Though this chapter focuses on the clock skew simulation with temperature effect and power distribution design, the following chapter discusses solutions for compensation for mitigating thermal-induced clock skew.

This chapter also quantifies change of impedance and noise in a power distribution network that is affected by temperature. At 1 GHz operation, the results show increase of 10% of PDN noise and impedance as compared to the results without thermal effect. The

results depict that temperature-aware PDN design is required because temperature increases impedance and noise in systems having high power- and thermal-density. This chapter also presents design considerations for several types of decoupling capacitors for high-density 3-D applications. The case study with chosen applications confirm the practicality of the. This chapter shows that PDN design for 3-D ICs needs to consider multiple types of on-chip decoupling capacitors and their temperature dependencies.

CHAPTER 3   COMPENSATION OF THERMALLY INDUCED DELAY

3.1 Introduction

This chapter builds on existing compensation techniques and introduces new methods that modify and implement them. The compensation is done mainly through active compensation techniques. Meanwhile, passive compensation techniques using resistance and capacitance of clock distribution network (CDN) components shown in Figure 40 can be explored for countering the thermal effects.
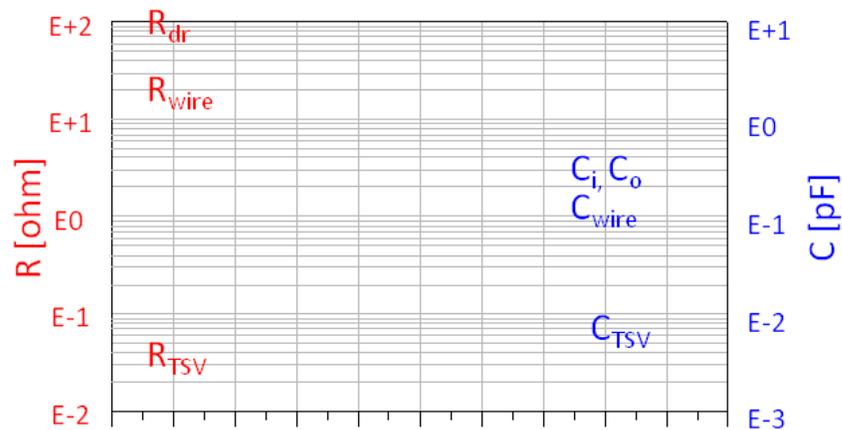
Figure 40   Resistance and capacitance of buffer, interconnects, and TSV.

RC values of TSV are smaller than wire interconnects (10^{-2} - 10^{-3}) and buffer area of TSVs are too dominant to compensate for RC delay. However, resistance and capacitance of TSV will grow as industry uses fine fabrication processes with small dimension of TSV and spacing. Passive methods with TSV or interconnect parasitics has limit to compensate large delay.

New methods proposed in this chapter focus on three parameters for the buffer namely– bias voltage, delay, and drive strength. Additional parameters like power supply noise at the buffer and crosstalk may be used to create new compensation techniques, since these parameters are also temperature sensitive.

The thermal gradients shown in the previous chapter cause delay variations which are compensated for by adjusting various parameters associated with the buffer and interconnects. These parameters were used to develop three methods namely: adaptive supply voltage, controllable path delay, and variable drive strength. The basic concept of the three methods are shown in Figure 41. Three approaches used in this chapter are discussed with variations of voltage, delay, and strength. To adjust these three components, additional circuitry for temperature sensor and control are necessary.
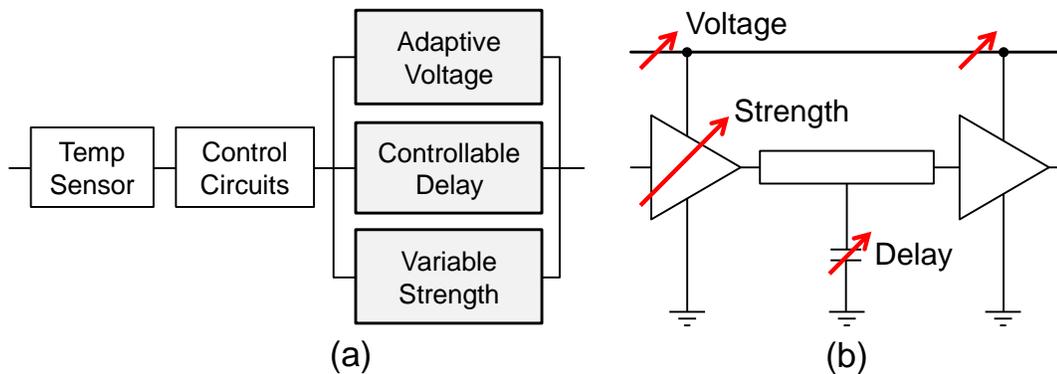
Figure 41   Concepts of the compensation methods: (a) a block diagram and (b) a simple circuit diagram [37].

3.2 Adaptive Supply Voltage Using Variable Reference Voltage

Threshold voltage and mobility of the transistor can be varied with temperature, as described in Chapter 2. The effect caused by temperature can be compensated using a change in bias voltage. Figure 42 shows sensitivity of clock delay with bias voltage, which is the concept used behind the first approach. Delay with temperature shows linear

dependency, however bias voltages (for drain and base) shows different linearity and temperature dependency.
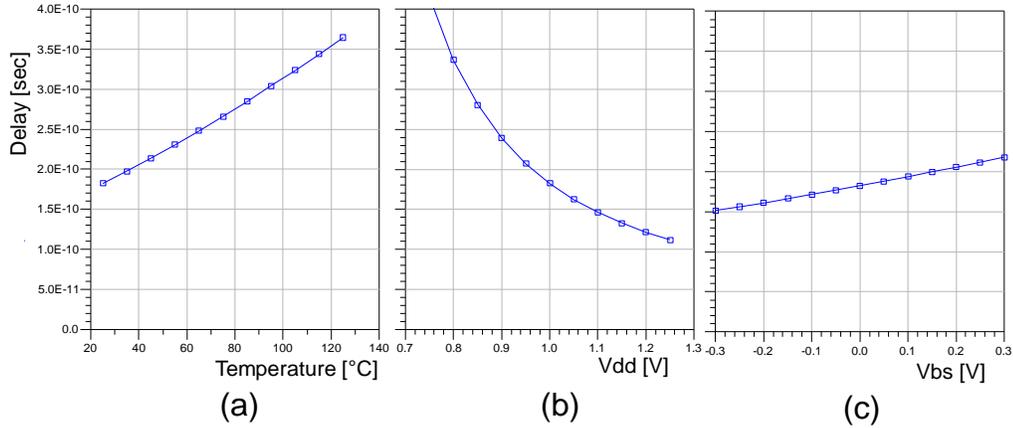
Figure 42   **Sensitivity with bias voltage (VDD and VBB) compared with temperature variation.**

The first approach is the adaptive voltage method which is based on the temperature dependency of mobility and threshold voltage of buffers. This method does not require additional circuits to control the other parameters including bias voltages, such as those for the drain or the body of transistors. These requirements of temperature sensors and level converters in previous methods have impeded scaling down of the design. For more efficient implementation for power and area efficiency, this work presents a new implementation of the circuits derived from the adaptive voltage method [17]. Previous methods have design and power overhead using additional circuitry such as temperature sensors, level shifters, and control circuits [17], as compared to the method being proposed.
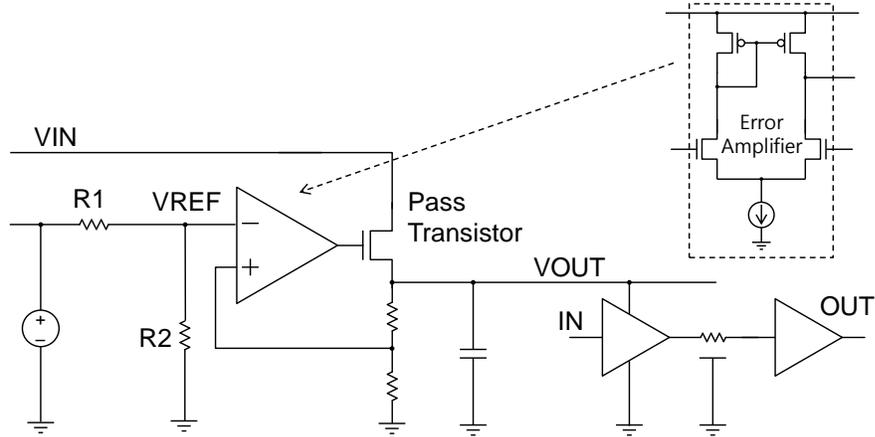
Figure 43   Schematic of adaptive supply voltage using variable reference voltage [37].

Instead of using an internal reference voltage generator, on-chip linear voltage regulators with temperature variable reference voltages, as shown in Figure 43, can be used. The voltage regulator generates a temperature-variable reference voltage with temperature dependency on the voltage divider. Implementation options are available with types of resistors (passive or active) and various temperature dependency. The circuitry of voltage divider consists of resistors R1 and R2, as shown in Figure 43.
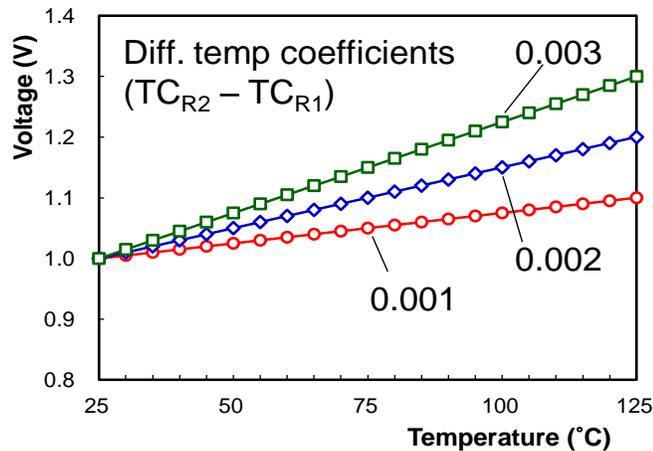
Figure 44   Generated reference voltages with differences of temperature coefficients of R2 and R1 [37].

To provide adaptive voltages, resistors, R1 and R2, have three values of temperature coefficients: positive temperature coefficient (PTC), zero-TC (ZTC), and

53

negative-TC (NTC). The on-chip resistor (R2) can be implemented using metal or silicon, and its nominal temperature coefficients have positive values, such as several hundred ppm (poly-silicon) and thousands of ppm (well-silicon). Resistor R1 can be implemented with either an on-chip resistor or SMD resistor, with the latter having a temperature coefficient resistance of 0.00015 per degree Celsius. A variety of differences in coefficients demonstrates the flexibility of this method because various materials enable multiple design choices. The temperature coefficient of resistor R1 is less than that of on-chip resistor, R2. Furthermore, because it is located on the PCB, R1 is far from the heat source.

From the implementation perspective, the external resistor R1 is similar to an alternate power distribution method comprising power transmission lines and series resistors, which have shown to reduce switching noise and provide design scalability [69]. The proposed method is less complex and more area-efficient because of the resistor based reference voltage generator and linear voltage regulator. Compensation of the temperature induced delay with this method is shown in Figure 45.
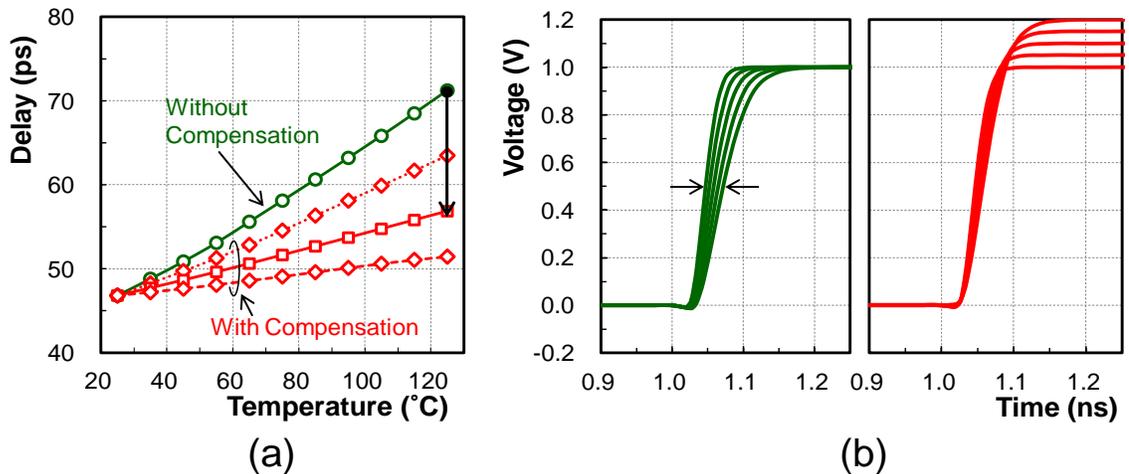
Figure 45 **Simulated skew of a repeater unit of the CDN using the adaptive voltage method: (a) variations of delay compensation with temperature coefficients and (b) waveforms [37].**

Table 10 shows comparison of this method with the previous published method. As the table indicates, the method proposed in this work has small die size overhead and is easier to control than the previous method, even though both methods have slightly larger power overhead than other methods, which consume additional dynamic power.

Table 10   **Comparison of the adaptive voltage method with the previous method; modified from [37].**

Component	[17]	This work
Implementation	Voltage regulator	Adaptive reference voltage
Additional circuits	Temp sensor, level shifter	Regulator and R divider
Compensation Range	Intermediate Range	**Intermediate range**
Power Consumption	Large (Static)	Large (static)
Die Size Overhead	Intermediate (TS, level shifter)	**Small (regulators)**
Controllability	Complex	**Easy (No temp sensor)**
Stability/Reliability	Stable	Stable

3.3 Controllable Path Delay with Redundant Paths

The second approach compensates the thermal-induced delay using additional adjustable loads, which were previously implemented with gate capacitance [22] for the controllable delay units. In this chapter, large values of delay components, such as redundant paths controlled by switches, providing larger delay values than the methods used in [22], are used. The previous method [22] showed that the delay units adjust and control the amount of delay, but require additional control logic due to small and varying

amounts of capacitance. Large delay values for redundant paths without the use of complex

logic is a key feature of the controllable delay method presented in this section.
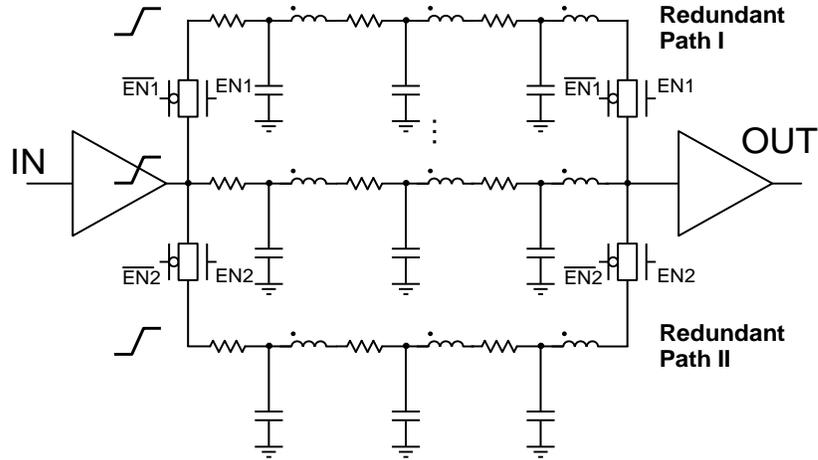
Figure 46 **Schematic of controllable delay compensation using redundant interconnect [37].**

The redundant interconnect design in Figure 46 provides a larger number of delay

options than other capacitor loads consisting of gate and junction capacitance [22]. For

example, resistance, capacitance, and inductance of the used 1 μm x 1 μm x 500 μm trace

are 15.1 ohm, 100 fF, and 0.69 nH, respectively, and mutual capacitance and inductance

are 33.0 fF and 0.52 nH, respectively.

$$TD_{even} = \sqrt{L_{even}C_{even}} = \sqrt{(L_{11}+L_{12})(C_{11}-C_{12})} \qquad \textbf{(17)}$$
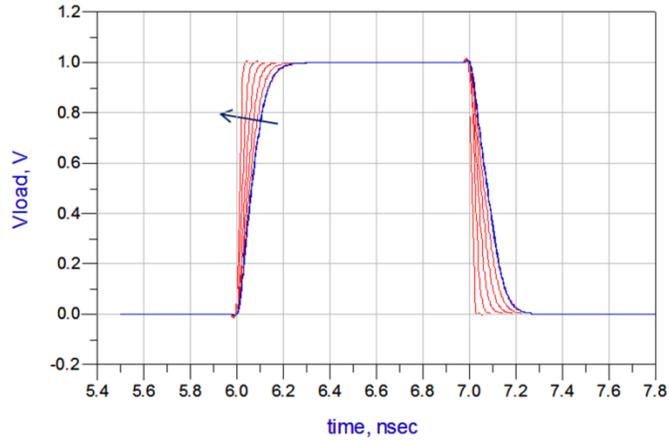
Figure 47 Delay effect by odd mode coupling (Lm: 0.7, Cm: 0.5).

With increase of crosstalk due mutual inductance and capacitance, this method with multiple line options and crosstalk effects described in (14) can provide larger delay, due to larger mutual capacitance and inductance condition. Furthermore, the method is simple and added interconnects added for delay compensation can act as shields. Figure 48 shows simulated skew and its reduction using the method.

(a)   (b)
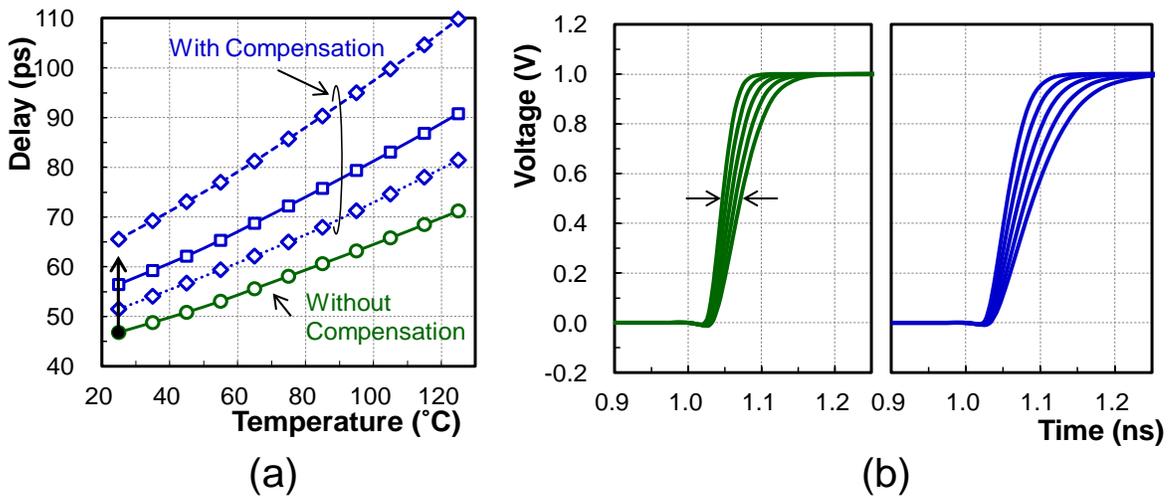
Figure 48   Simulated skew of a repeater unit of the CDN using the controllable delay method: (a) variations of delay compensation with redundant paths and (b) waveforms [37].

Comparison of the proposed method with the method in [22] is shown in Table 11. As compared in the table, the controllable delay method has a wider compensation range

than the previous method, small die size overhead, and easiness for control. A detailed comparison is provided in Table 11.

Table 11   Comparison of the controllable delay method with the previous method; modified from [37].

Component	[22]	This work
Implementation	Interconnect and gates	Redundant interconnects
Additional circuits	Xgates and TRs	Xgates and traces
Compensation Range	Small range	**Wide range**
Power Consumption	Small (dynamic)	Small (dynamic)
Die Size Overhead	Intermediate (Xgates, TRs)	**Small (Xgates and traces)**
Controllability	Complex	**Easy**
Stability/Reliability	More stable	More stable

3.4 Variable Driving Strength for Clock Repeaters

The third method uses variable drive strengths for clock repeaters. These drive strength options have been conventionally used in I/O buffers for obtaining better signal integrity or power efficiency with optimal output impedance. The strength option from a typical data buffer is applied for clock buffers instead of the complex clock buffer shown in [23].
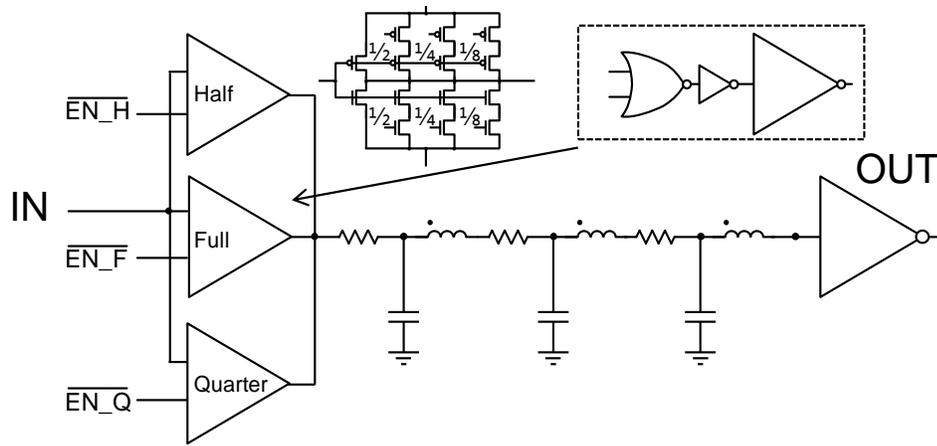
Figure 49 Schematic of variable driving strength buffer.

Simulated skew of a repeater unit of the CDN is reduced with additional buffers of half, quarter, and octant sizes, which are conventional scaling numbers in transistor sizing. The buffer size option is easily implemented, however the additional buffer directly induces power consumption with the addition of current capacity. Skews of Figure 50 (a) and (b) show decreased clock skew using the method.


(a)   (b)
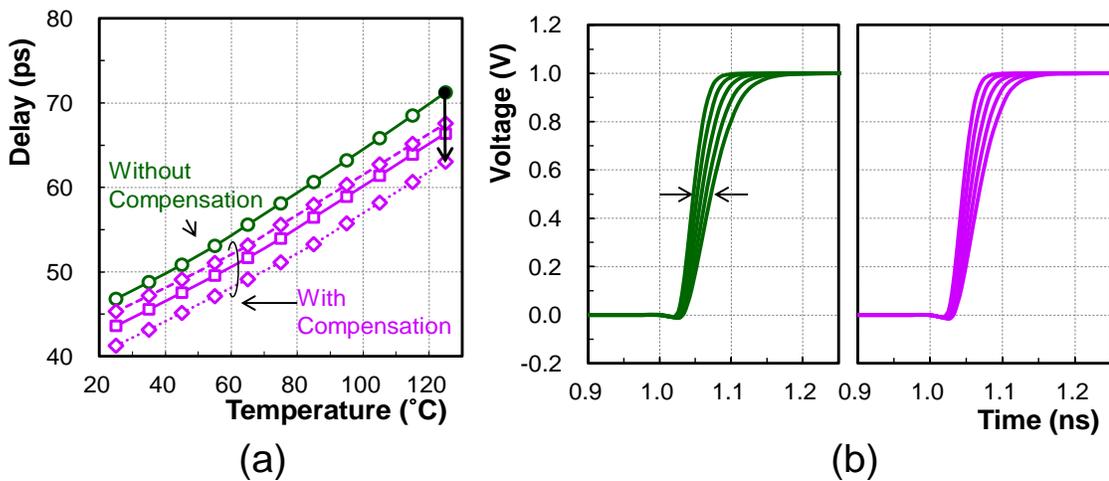
Figure 50 Simulated skew of a repeater unit of the CDN using the variable strength method: (a) variations of delay compensation with drive strengths and (b) waveforms [37].

Table 12 compares the proposed variable strength method with the method shown in [23]. As indicated in the table, the method has a smaller compensation range, but has a small die size overhead and is easy to control.

Table 12   **Comparison of the variable strength method with the previous method; modified from [37].**

Component	[23]	This work
Implementation	Adjusting driving strengths	Conventional parallel buffer
Additional circuits	Control logic + additional TR	Control logic + additional TR
Compensation Range	Intermediate range	**Small Range**
Power Consumption	Intermediate	Intermediate (buffer)
Die Size Overhead	Intermediate	**Small**
Controllability	Complex	**Easy**
Stability/Reliability	Stable	Stable

3.5 Comparison of Methods

The proposed methods show different delay compensation capabilities. Figure 51 shows delay change with temperature increase from 25 degrees to 125 degrees Celsius. Slope of the graph in Figure 51 indicates change of delay with temperature and the change of delay shows the range of delay compensation. The controllable delay method is capable of the largest delay compensation, and the variable strength method shows the smallest delay compensation among the three methods presented.
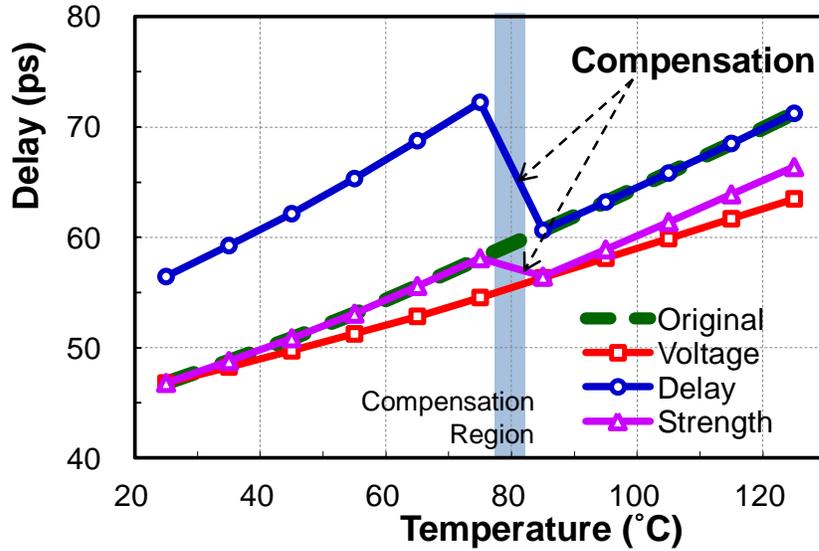
Figure 51   Simulated delay of the repeater unit using the three methods [37].

Each method has its advantages and disadvantages. As indicated earlier, the adaptive voltage method can be implemented with smaller area overhead and simpler control; however the signal integrity of the clock, such as duty cycle and crossover-point, are degraded with this method as described in the previous section. Therefore, negative impacts require optimization of the amount of delay compensation using this method. On the contrary, the controllable delay method is simple but provides a larger amount of delay compensation; furthermore, it is stable for thermal compensation. Delay compensation using variable strength is proportional to the addition of transistor area and power, but the method can be implemented using simple control logic. These methods can be combined to compensate for thermal variations in CDN. Figure 52 shows the comparison of these techniques with previous methods.
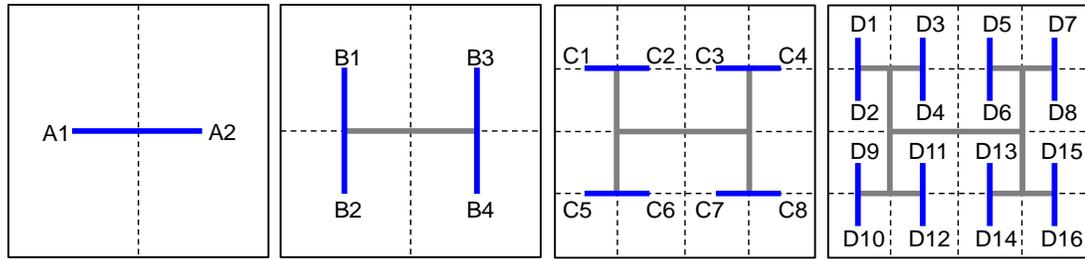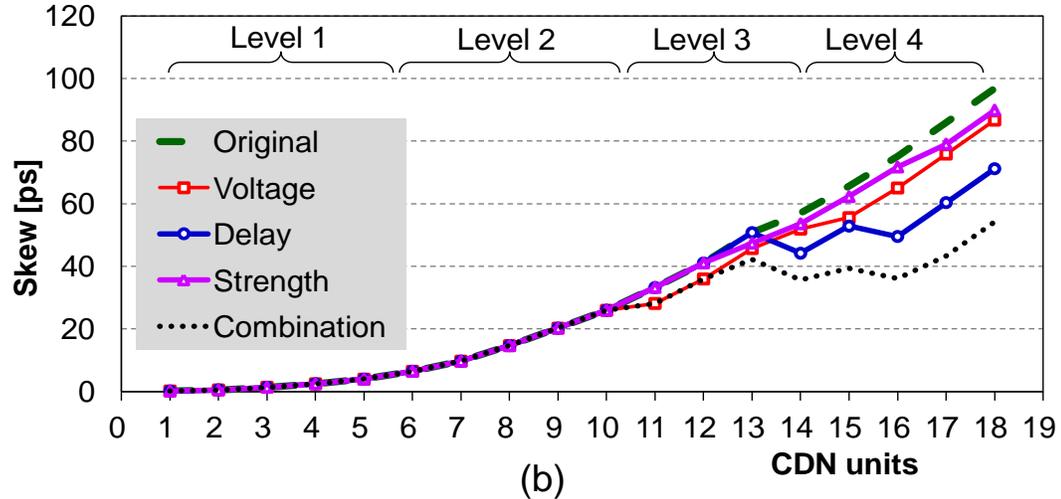
Figure 52 **(a) Level of an H-tree clock. (b) Delay compensation with circuit allocation for compensation implementation [37].**

With different compensation range and design overhead, the three methods can be used at different levels of the H-tree clock structure. Figure 52 (a) shows an example of a cascaded level for the H-tree, and Figure 52 (b) illustrates the delay compensation profile with different levels of the H-tree.

The performance improvement with the implementation of the compensation techniques depicts a decrease in delay from 96.2 ns to 54.2 ns, which represent a 43.5 % improvement using these methods. The skew improvement also causes degradation in power and area as shown and compared in Figure 53. This overhead increases temperature immunity by using the compensation techniques, but the control circuit implementation requires more die space and consumes additional power to activate. Power and design

overhead with the proposed methods are 11.6% and 9.7%, respectively, as compared to the without compensation circuits.
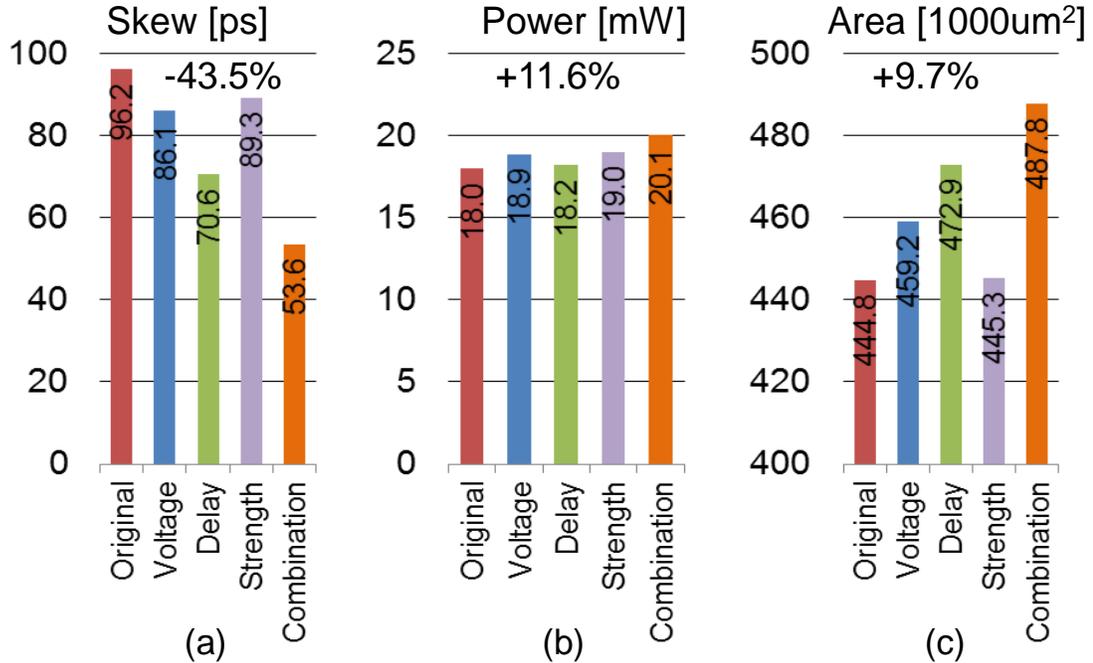
Figure 53   **Comparison of (a) skew, (b) power consumption, and (c) area [37].**

Variation in the supply voltage has an effect on the total power consumption of the system. These proposed methods increase leakage power. Distribution of power has been estimated with the PowerPC estimator tool. Total power increased from 18 mW to 20 mW, however, the power overhead will be reduced when these concepts are optimized using a customized chip design.

3.6 Summary

This chapter discusses approaches to mitigate thermal induced skew simulated in a CDN. Proposed methods are different from previous methods described in [17], [22], and [23]. In summary, (1) the first method is able to provide an adaptive voltage without temperature sensors and reference voltage generator in an on-chip voltage regulator. The

reference voltage generator-free low dropout (LDO) voltage regulator compensates the temperature-induced skews by generating self-adapted bias in an efficient manner; (2) The second method uses redundant paths which provide a larger delay compared to previous methods using gate capacitance of transistors without any leakage; (3) The third method uses strength-based skew-tuning which has been used for data buffer's impedance control, instead of using complex buffers. These proposed methods are more energy-efficient and innovative when their various performances and overhead are compared.

There is also scope for improvement in the implementation schemes. The designs in this work were developed with the objective of verifying the functionality of the compensation methods. However, they displayed excessive degradation in terms of power and area. Performance always trades off with power and area but at least one of them can be optimized while the others can be kept at a level where it provides satisfactory performance. Power efficient techniques can be constructed using gating and scaling schemes that guarantee that the right amount of compensation is provided at the right time. Area efficient designs would require advanced logic optimization of the control units and circuit modifications.

This chapter also presented simple but efficient compensation methods to reduce the degradation and showed detailed implementation of the proposed methods. These methods have different capabilities for skew compensation with power and area overhead. Therefore, the effectiveness and efficiency of the methods are implemented with optimal combination for CDN design. Design results using a 45 nm process showed that 43.5 % performance improvement is possible with 11.6% and 9.7 % power and area overhead,

respectively. The proposed concepts are validated and correlated with test vehicles in the

following chapters.

CHAPTER 4   VALIDATION USING MEASUREMENTS

4.1 Introduction

As a verification procedure to verify the proposed concepts and simulation results, two types of test vehicles were designed. One test vehicle was developed using a field-programmable gate array (FPGA), which was configured with programmable logic blocks using a hardware description language (HDL). As a prototype for validation, FPGA is one of the best options. The test vehicle confirms the delay mitigation concepts and presents delay compensation performance described in Chapter 4.

The other test vehicle is based on a customized integrated circuit which contains a clock network, compensation circuit, and power delivery network. The designed chip was fabricated with a multi-project wafer (MPW) project using 180 nm process. This test vehicle was used to correlate the electrical-thermal solver and PDN simulation, described in Chapter 2 with measurements.

4.2 FPGA Implementation

To validate the suggested methods prior to custom chip design, a test vehicle using FPGA was developed. This first test vehicle was designed to attribute various thermal conditions using additional heaters. The positive temperature coefficient (PTC) heater is one of the best candidates for external heater because of safety, longevity, and cost. External PTC heaters attached on the FPGA mimic the condition for measurement. Figure 54 (a) shows the FPGA test vehicle with additional heaters, which has SMA (Sub-

Miniature version A) connector that can be used to connect to test equipment such as automatic testers and digital sampling oscilloscopes.
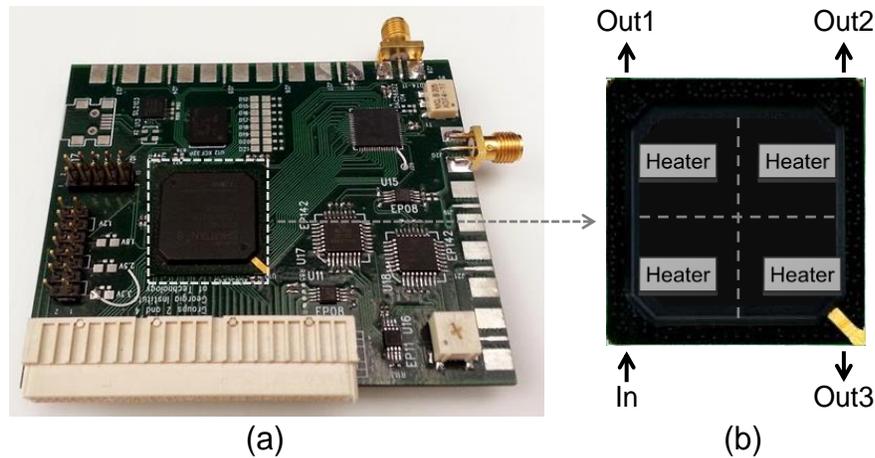
Figure 54   **(a) FPGA (Spartan 6)-based test vehicle. (b) Placement of four PTC heaters on the FPGA to mimic the thermal condition [70].**

As a first step, the Spartan 6 FPGA test vehicle, which is fabricated using 45 nm technology node with an operation temperature range between 85°C and 125°C, was set with constraints to replicate the conditions for thermal and electrical analysis. External PTC heater components create artificial temperature gradients, however the number of heaters was limited to four because of the physical dimension.

The CDN was synthesized with Verilog coding using the Xilinx ISE Design Suite. The Plan Ahead tool confirmed that the CDN had been implemented in the form of H-Tree architecture, and the iSim waveform analyzer was used to verify the design. To mimic a real symmetric H-tree structure of the clock, location constraints in FPGA synthesis were used. With the location constraint, Figure 55 shows an implementation of a CDN structure in FPGA, which is similar to the H-tree structure described earlier.
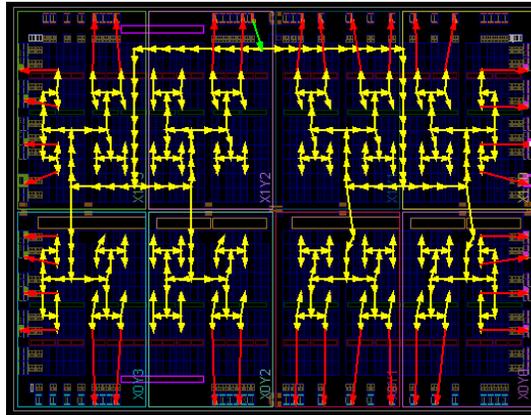
Figure 55   A CDN structure in FPGA [37].

First, the VDD of buffers to implement the adaptive voltage method shown in Figure 56 (a) were adjusted. Buffers were constructed using switch models in Verilog, so external ports were used to feed supply voltages to them. The voltages on the IO port were optimized to speed up the buffers [71].

(a)   (b)
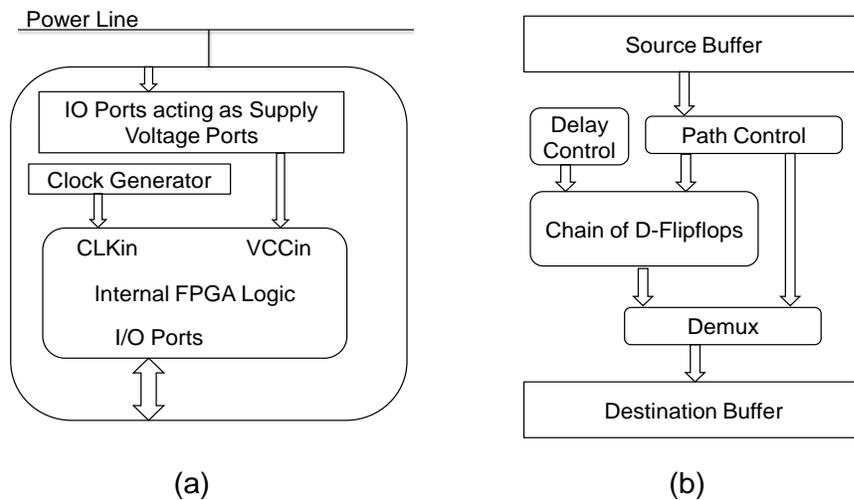
Figure 56   FPGA implementation of compensation techniques with (a) Adaptive supply voltage and (b) Controllable path delay [71].

Second, Figure 56 (b) shows the implementation structure in the FPGA-based test vehicle for using the controllable path delay method. Delay components in the method are redundant interconnects, but the component in the test vehicle was implemented with D

68

flip-flops. A control unit was used to select the number of flip-flops between the source and destination buffers as an interconnect delay along the path. An algorithm for the combination of these methods starts with an estimation of the temperature across the CDN and fetches data from the memory of the FPGA in the test vehicle. The fetched data indicate the predefined temperature maps that provide the delays across various paths. The algorithm compares the delays to a threshold and determines whether a correction is needed or not. If necessary, it selects the correct compensation technique and continues. The test vehicle including the corrective techniques, the primary CDN, and a control unit was coded with RTL to implement the algorithm shown in Figure 56.

4.3 Custom IC Design

4.3.1 Design Concept

The aim of this test chip is to validate the temperature gradient simulated with the electro-thermal simulation with temperature sourcing and monitoring circuits in a real chip. In addition, an analysis of the temperature effects on high speed circuits, especially clock distribution network (CDN) was performed. The objectives of this chip design are:

- Design of a test chip which is capable of temperature sourcing for creating temperature gradient in the chip and temperature monitoring using temperature sensor circuitry.
- Correlation of electro-thermal simulation results with measurement using a test chip having temperature sourcing and monitoring circuits and analysis of the effects of the temperature on high speed circuits.

- Additional compensation circuits for adjusting the delay mismatch caused by temperature gradient on a global clock tree and verification and hardware measurements of the designed circuit.

- Validation of temperature effects of the temperature on the power delivery network (PDN) including on-chip decoupling capacitors (ODCs)

The goal of the design was to mimic the temperature effects of real chips including temperature gradients and their measurements. For temperature sourcing and monitoring, on-chip heaters implemented with polysilicon resistors and MOS diodes were used. Dimension of on-chip heaters was 100μm x 100μm and these heaters were implemented on the poly-silicon layer because the layer has reasonable resistance. Sixteen pairs of heaters and temperature sensors were placed on a 4 x 4 grid, as shown in Figure 57 (a). Also, MOS diodes were included as heat monitors. This provided the necessary baseline features for electrical-thermal analysis of high density ICs, which was used to correlate with simulations. This combination of measuring the thermal effects with the resulting electrical effects will provide sufficient data to validate the electrical-thermal solver along with other circuit simulators. Figure 57 (b) shows layout of the heater and temperature monitoring diode.

(a)	(b)
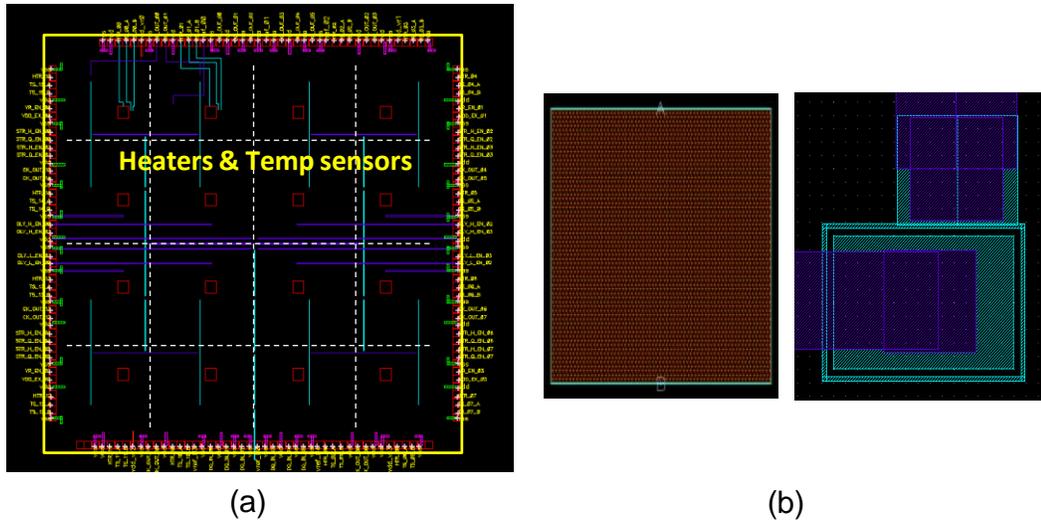
Figure 57   **(a) Location of designed heaters and temperature sensors. (b) Layout of an n-poly-based heater and a temperature monitoring diode.**

Die size was assumed as 10 mm x 10 mm for simulation though only 3.8 mm x 3.8 mm die area was available for the design, as shown in Figure 58. The size of the unit grid for both cases (simulation and design) were similar and around 1 mm x 1 mm, but the number of heater and temp sensor grid decreased from 10 x 10 to 4 x 4 because of the limitation of die area, as shown in Figure 58.

(a)	(b)
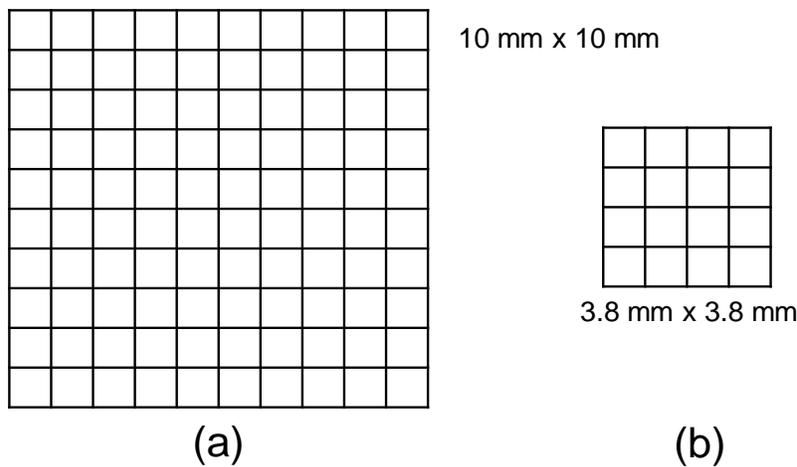
Figure 58   **Die area for the heater and the temperature sensor. (a) 10 mm x 10 mm for simulations and (b) 3.8 mm x 3.8 mm for the chip design.**

To validate the methods, a customized IC was designed and a test vehicle fabricated. Figure 59 shows a full schematic of the CDN including sub-blocks and compensation circuits. Some of the clock repeaters use four different adaptive voltages and delay options implemented using transmission gates.
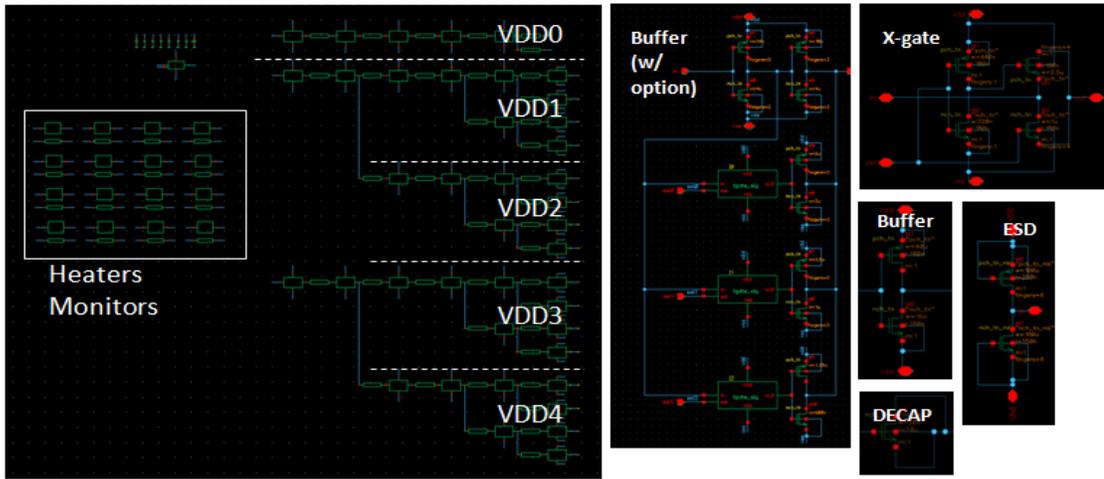
Figure 59   Schematic of CDN circuits.

Layouts of the main sub-circuits are shown in Figure 60. Same voltage regulator shown in chapter 4 was used for adaptive voltage generation. The delay and buffer strengths were controlled with optional switches.
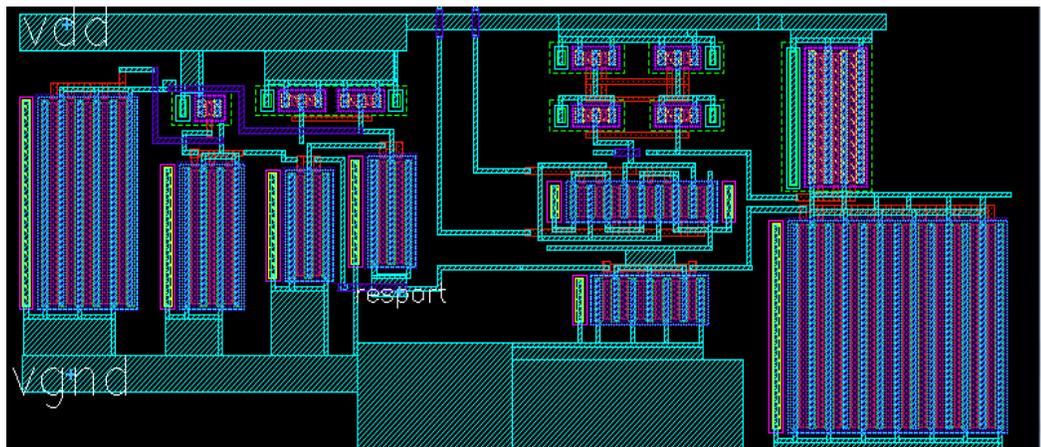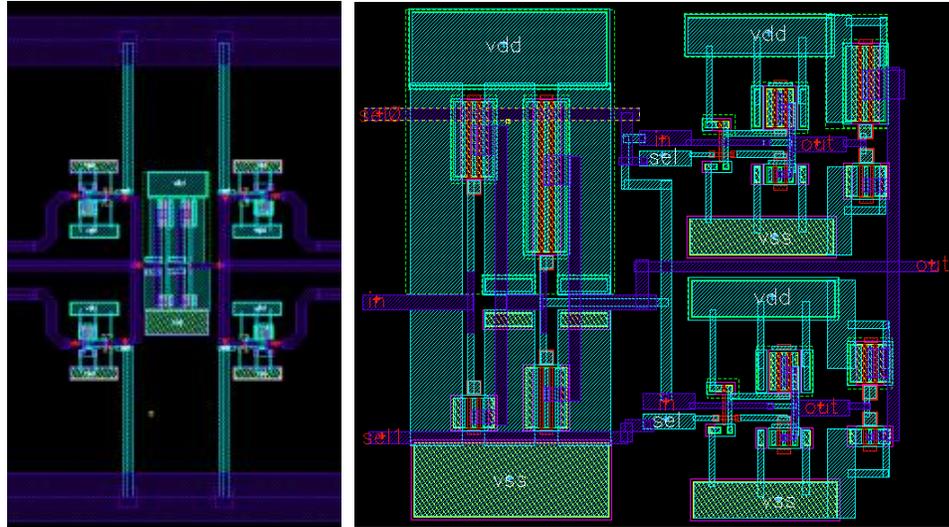
Figure 60   Layout results of the compensation circuits.

This chip was fabricated using a low-cost multi-project wafer (MPW) IC prototyping service. It used the 0.18 um process of Magna Chip/SK Hynix and consists of six metal layers. The circuits were implemented on 0.18μm CMOS process with the specifications shown in Table 13.

Table 13   Fabrication process specifications.

Process node	0.18μm CMOS
Layers	1 poly-silicon layer, 6 metal layers
Devices	1.8V (thin oxide)/3.3V (thick oxide)/5V
Minimum gate length	0.18μm for 1.8V, 0.30/0.35μm for 2.5V
Substrate	P-substrate with N-wells

The size of the chip including output pads was 3.8 mm x 3.8 mm. An H-tree CDN with compensation circuits and on-chip voltage regulator were placed. The CDN consists of inverter chains with two different repeaters for FO1 (Fan-Out 1) and FO2 (Fan-Out 2) and 500 um metal interconnect. The on-chip voltage regulator generates 1.8 V for internal voltage. Figure 61 shows the full layout with meshed power distribution network.
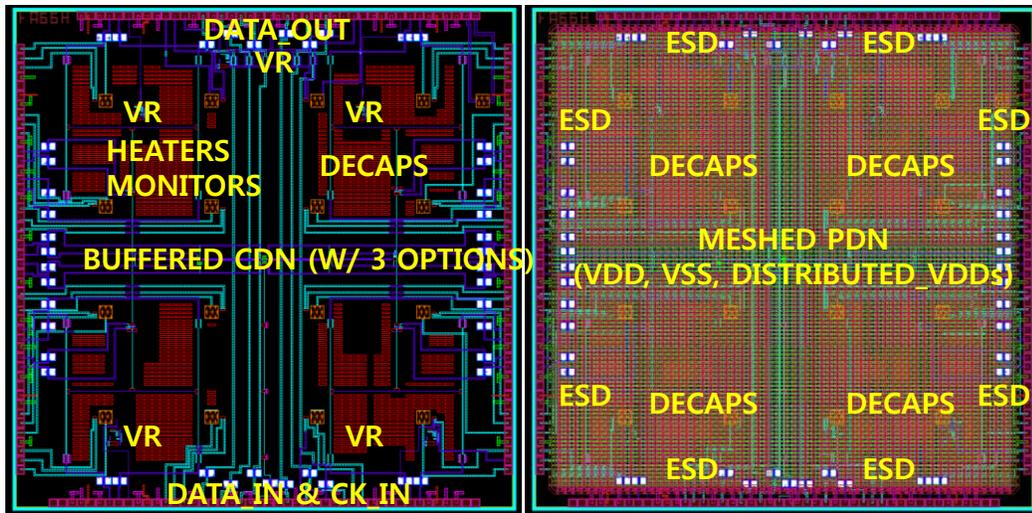
Figure 61   Overview of layout and power distribution network [81].

Figure 61 shows the layout of the complete chip including the power distribution network and input/output pin location. Figure 62 shows fabricated chips.
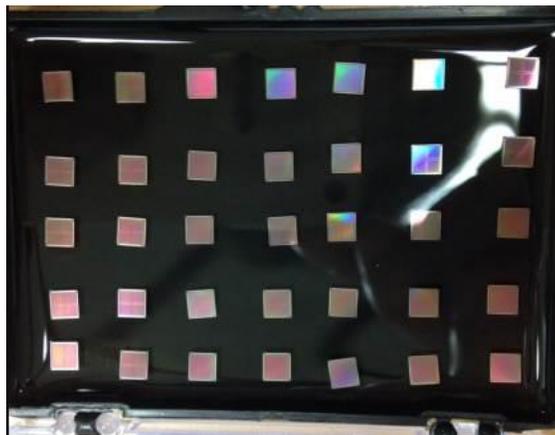
Figure 62   Fabricated chips using 180 nm process.

4.3.2 IC Test Board

A 10 mm x 10 mm size PCB was used for the IC-based test vehicle to validate the methods discussed in the Chapter 4. Figure 63 shows the concept of the test vehicle. The chip, either packaged or not, was connected to an FPGA for input and output control.

The results that can be measured with the test vehicle below are as follows:

1) Measure temp gradient on the chip and correlation with thermal simulation (power map, in °C)

2) Timing effects, especially delay variation, caused by temperature gradients (in ps)

Pin Assignment

Pin Name	Pin Description
CLK_IN	Clock Input
CLK_OUT [0:n]	Clock Output
Data_IN [0:n]	Data Input
Data_OUT [0:n]	Data Output
HS [0:n]	Heat Source Input
TS [0:n]	Temp Sensor Control
VDD	Power voltage
VREF [0:n]	Reference voltage
VSS	Ground voltage
TP [0:n]	Test Point
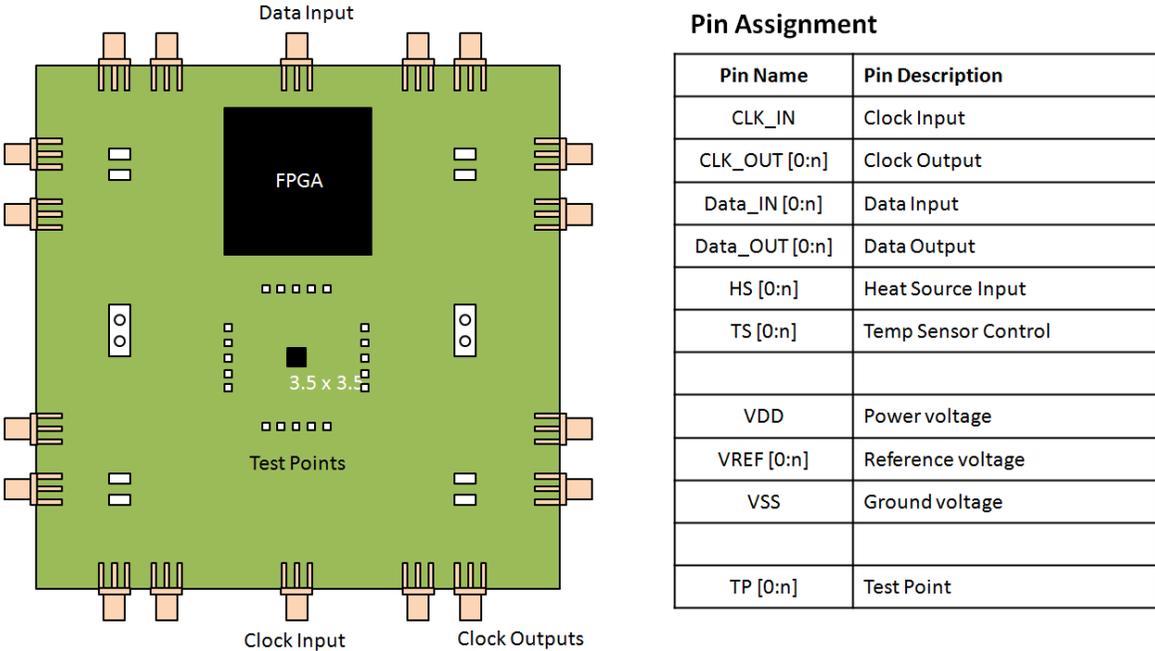
Figure 63   **Concept of the test vehicle and pin assignment.**
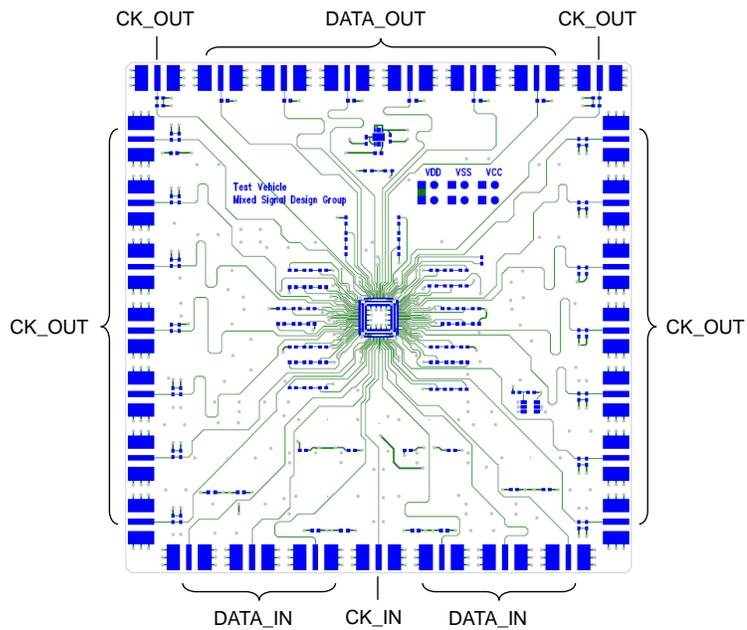
Figure 64   **PCB design and input/output port location.**

A 10 mm x 10 mm size PCB was used to validate the operation of an LDO and compensation methods discussed in the previous section. Figure 65 shows the designed PCB of the test vehicle and a wire bonded chip at the center of the test vehicle.

(a)   (b)
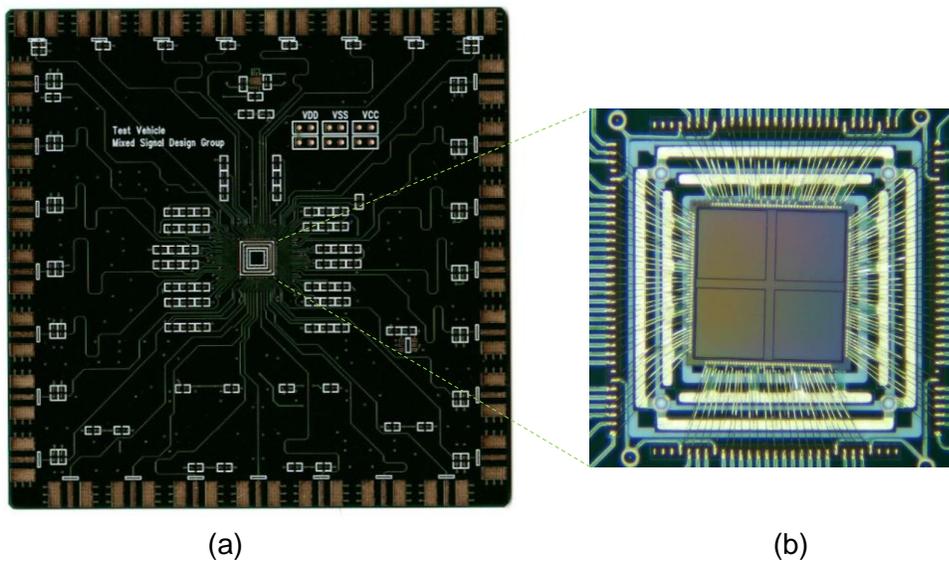
Figure 65   **(a) Fabricated PCB and (b) the wire-bonded chip [81].**

The fabricated chip has 184 pins to estimate the effect of temperature on the clock distribution network and power distribution network. Table 14 shows the pinout of the fabricated chip. The chip was bonded as a chip-on-board (CoB) type package.

Table 14   Pinout of the fabricated chip.

TOP				RIGHT				BOTTOM				LEFT			
No.	Name	No.	Name	No.	Name	No.	Name	No.	Name	No.	Name	No.	Name	No.	Name
1	NC	24	VDD_VR0	47	HTR_03	70	DLY_L_EN_03	93	NC	116	CK_IN	139	HTR_12	162	DLY_H_EN_01
2	NC	25	VSS	48	TS_03_A	71	DLY_L_EN_02	94	NC	117	VSS	140	TS_12_A	163	DLY_H_EN_00
3	NC	26	DQ_OUT_02	49	TS_03_B	72	VSS	95	VSS	118	DQ_IN_03	141	TS_12_B	164	VSS
4	VDD	27	VDD	50	VDD	73	TS_10_B	96	VDD	119	VSS	142	VSS	165	TS_05_B
5	VDD_VR1	28	DQ_OUT_01	51	VR_EN_02	74	TS_10_A	97	VDD_VR4	120	DQ_IN_04	143	HTR_08	166	TS_05_A
6	VSS	29	VSS	52	VDD_VR2	75	HTR_10	98	VSS	121	VDD	144	TS_08_A	167	HTR_05
7	CK_OUT_00	30	DQ_OUT_00	53	VSS	76	VSS	99	CK_OUT_15	122	DQ_IN_05	145	TS_08_B	168	VSS
8	CK_OUT_01	31	VSS	54	HTR_07	77	CK_OUT_10	100	CK_OUT_14	123	VSS	146	VDD	169	CK_OUT_05
9	VREF_01	32	TS_02_A	55	TS_07_A	78	CK_OUT_11	101	VREF_04	124	TS_13_B	147	VDD_VR3	170	CK_OUT_04
10	VSS	33	TS_02_B	56	TS_07_B	79	VDD	102	VSS	125	TS_13_A	148	VR_EN_03	171	VDD
11	VDD_VR1	34	HTR_02	57	VSS	80	STR_H_EN_07	103	VDD_VR4	126	HTR_13	149	VSS	172	STR_Q_EN_01
12	VDD	35	VDD	58	STR_H_EN_03	81	STR_Q_EN_07	104	VDD	127	VDD	150	STR_Q_EN_05	173	STR_H_EN_01
13	HTR_01	36	VDD_VR2	59	STR_Q_EN_03	82	VSS	105	HTR_14	128	VDD_VR3	151	STR_H_EN_05	174	VSS
14	TS_01_A	37	VSS	60	VDD	83	VR_EN_04	106	TS_14_B	129	VSS	152	VDD	175	TS_04_B
15	TS_01_B	38	VREF_02	61	CK_OUT_07	84	VDD_VR4	107	TS_14_A	130	VREF_03	153	CK_OUT_08	176	TS_04_A
16	VSS	39	CK_OUT_02	62	CK_OUT_06	85	VDD	108	VSS	131	CK_OUT_13	154	CK_OUT_09	177	HTR_04
17	DQ_OUT_05	40	CK_OUT_03	63	VSS	86	TS_11_B	109	DQ_IN_00	132	CK_OUT_12	155	VSS	178	VSS
18	VDD	41	VSS	64	HTR_06	87	TS_11_A	110	VSS	133	VSS	156	HTR_09	179	VDD_VR1
19	DQ_OUT_04	42	VDD_VR2	65	TS_06_A	88	HTR_11	111	DQ_IN_01	134	VDD_VR3	157	TS_09_A	180	VR_EN_01
20	VSS	43	VDD	66	TS_06_B	89	VSS	112	VDD	135	VDD	158	TS_09_B	181	VDD
21	DQ_OUT_03	44	VSS	67	VSS	90	TS_15_B	113	DQ_IN_02	136	VSS	159	VSS	182	TS_00_B
22	VSS	45	HTR_A	68	DLY_H_EN_02	91	TS_15_A	114	VSS	137	NC	160	DLY_L_EN_00	183	TS_00_A
23	VREF_00	46	HTR_B	69	DLY_H_EN_03	92	HTR_15	115	VDD	138	NC	161	DLY_L_EN_01	184	HTR_00

4.4 Measurement Setup

To measure the developed test vehicles, time domain and frequency domain equipment were used. A digital sampling oscilloscope (DSO) and vector network analyzer (VNA) were used for the measurements. For measuring time transient waveforms, signal generators, power supply units, and digital voltage meter were used, as shown in Figure 66 (a). A vector network analyzer and a probe station shown in Figure 66 (b) were used to measure scattering parameters of the power distribution network. The setup is used for both FPGA and IC-based test vehicles.
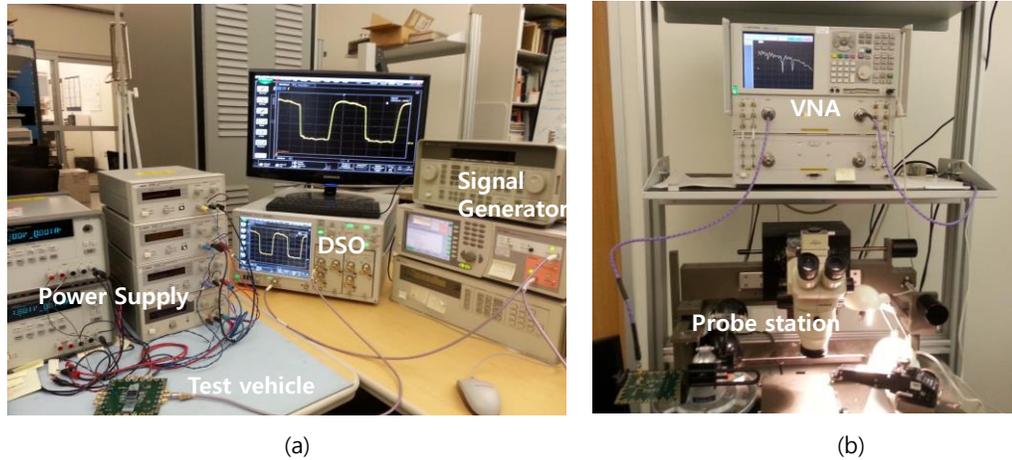
Figure 66   **Test setup for (a) digital sampling oscilloscope (DSO) and (b) vector network analyzer (VNA).**

The measurements of the test vehicles were two fold, namely: 1) Measurement of temperature gradient on the chip and correlation with thermal simulation with power maps (in degree Celsius), 2) Measurements of timing effects, especially delay variation, caused by temperature gradients (in ps) and its correlation with the electrical-thermal solver and circuit simulator.

4.5 Measurement Results

4.5.1 FPGA-based Board

Figure 67 and Figure 68 show measurement results using the FPGA-based test vehicle. Figure 67 (a) shows delay change due to temperature without using any compensation techniques. Alternatively, Figure 68 (b) shows a better response as the slope of delay decreased by using the adaptive voltage technique. The graph in Figure 67 shows that the adaptive voltage technique improves the propagation delay from 410 ps to 143 ps, when the temperature is increased to predefined values by heaters.
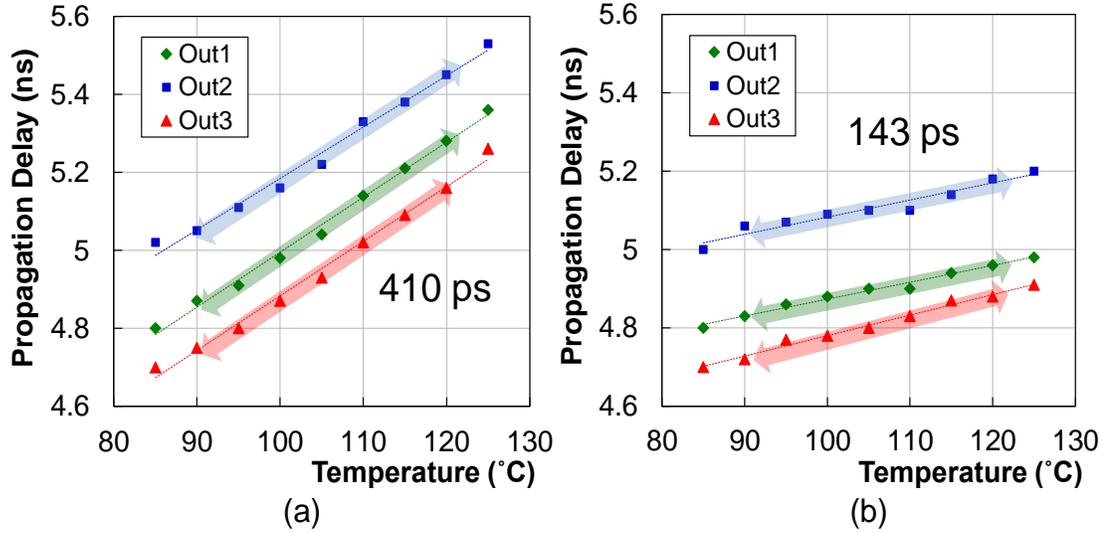
Figure 67 Propagation delay with temperature using the adaptive voltage method (a) without compensation and (b) with compensation [70].

Measurement with the tunable delay method are shown in Figure 68 (a) where the skew is stable across the CDN, as shown in Figure 68 (b). The compensation method reduces the skew from 1155 ps to 531 ps in the presence of temperature effects. The amount of skew compensation shows the delay compensation capability of the methods, and the results are well correlated with the methods compared in Chapter 3.
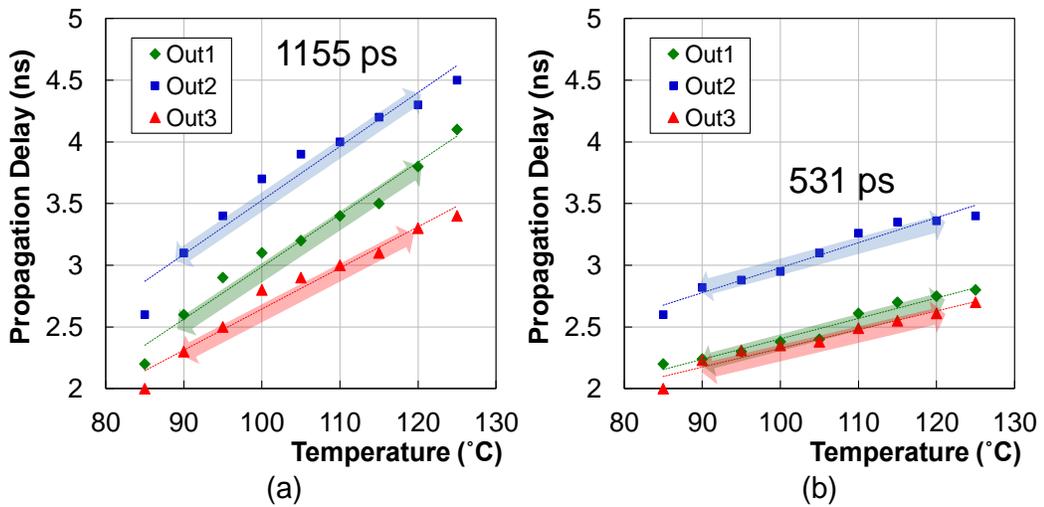
Figure 68 Propagation delay with temperature using the controllable delay method (c) without compensation and (d) with compensation [70].

A more practical test case was applied for temperature varying conditions in real time, as shown in Figure 69. This was accomplished by using the preset values of temperature and time steps in the control unit. Time and temperature during the propagation delay measurement were swept for real time compensation capability. As shown in Figure 69, the deviation of propagation delay is reduced with these methods, which demonstrates resilience to temperature. The measurement results using the FPGA-based test vehicle are correlated with skew mitigation from the proposed methods.
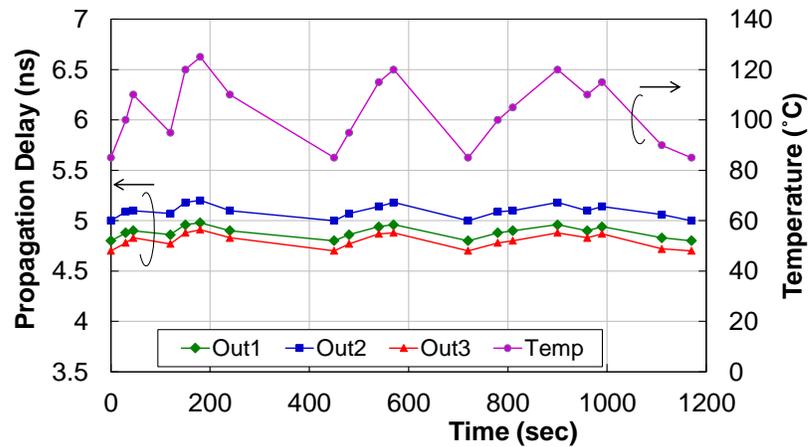
(a)
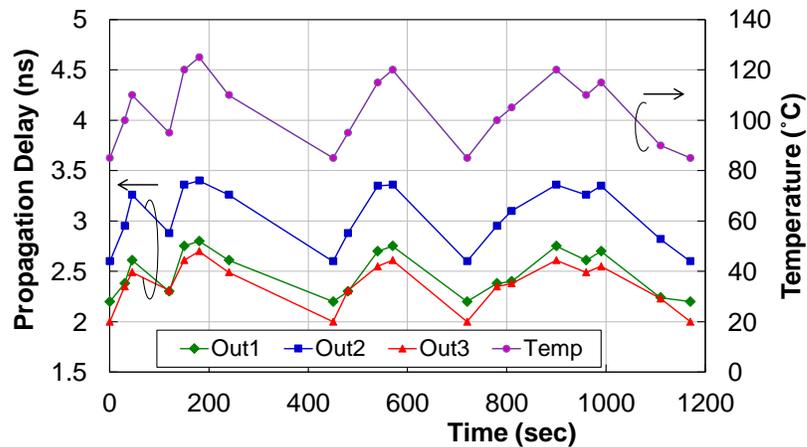
(b)

Figure 69   **Real time compensation test with (a) the adaptive voltage method and (b) the controllable delay method [70].**

4.5.1 Custom IC-based Test Vehicle

Fabricated IC was used for evaluating the thermal effects on the clock distribution network. On-chip temperature gradients were measured using temperature generating and monitoring blocks. We induced variable current to each heater with resistor networks built on test board by varying resistors and input voltage. Diode based temperature sensors were used for temperature measurements. To measure the local temperature, we use temperature monitoring circuits using diodes. Figure 70 (a) and (b) show the measured I-V profile of temperature monitoring circuits and the measured I-V curves for different temperatures, respectively. Differences in current is a measure of local temperature.

(a)   (b)
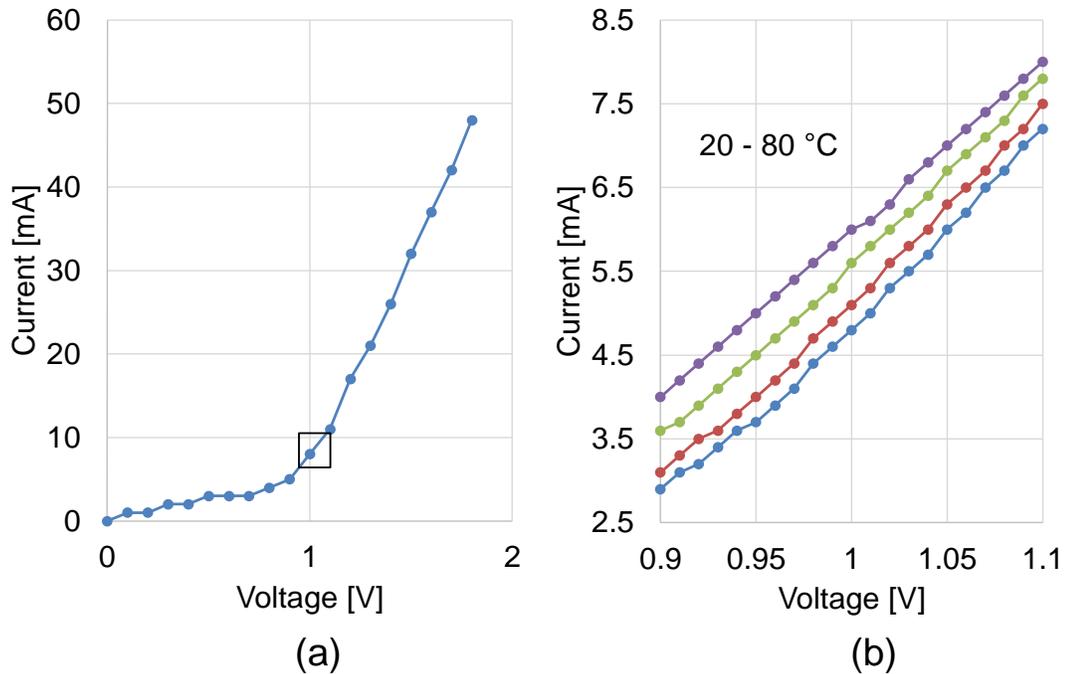
Figure 70   **(a) Measured I-V profile of temperature monitoring circuits and (b) temperature variations [81].**

The power consumed by each heater calculated using a voltage source and resistor divider resulting in a power map as shown in Figure 71 (a). The electrical-thermal solver was used to compute the temperature distribution on the die for the test vehicle. Since the typical heat transfer coefficient natural convection is around 4-5 W/(m$^2\cdot$K), a heat transfer coefficient of 4.0 W/(m$^2\cdot$K) was used as the convection boundary condition for analysis, which accounts for any radiation effects as well. Figure 71 (b) and (c) show the simulated and measured temperature distribution for the power map used in Figure 71 (a).

Measurement results confirmed that the simulated temperature gradients can be generated in the die. These results are reasonably well correlated with electro-thermal simulation results. From Figure 72, it can be seen that the correlation is good for minimum and maximum temperatures for the three power maps (7%-13% and 1%-3% error in simulations) while the error is larger for the temperature gradients due to the smaller values involved (19%-46%). Nevertheless, these correlations provide a reasonable degree of confidence in the simulated temperatures, since there is some inaccuracy in the position of the heaters and monitoring circuits due to the 4x4 grid used for the chip, as opposed to a much finer non-uniform grid used in the simulations.
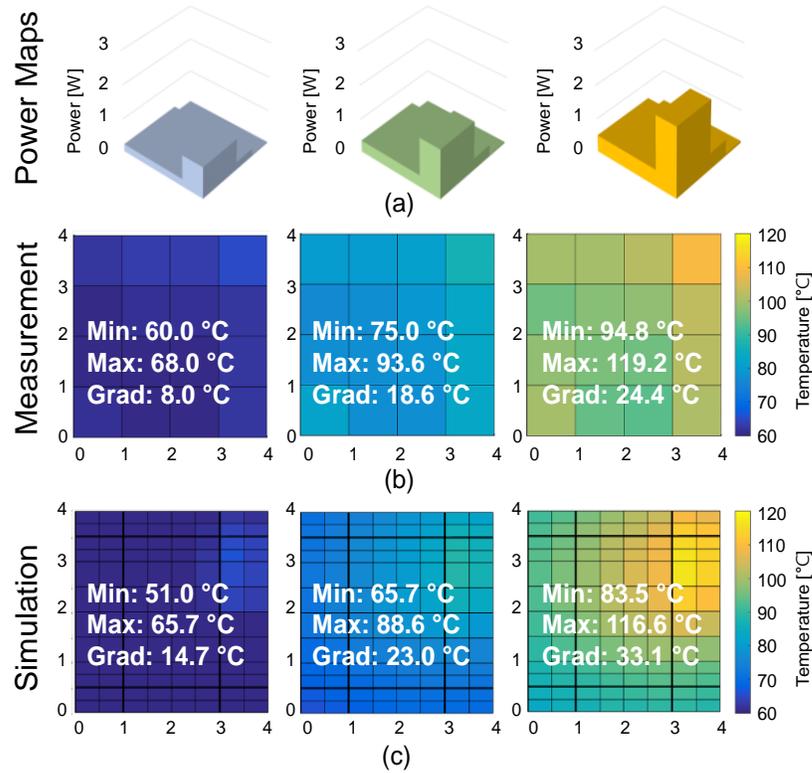
Figure 71 **(a) Power maps used for simulation and measurement (b) measured temperature profiles, and (c) simulated temperature profiles [81].**

The temperature gradient has an effect on the integrity of high speed signals. Power delivery network (PDN) characteristics also vary with temperature. Figure 72 shows scattering parameters of PDN for the test vehicle. Temperature characteristics of PDN are shown in Figure 72 where the temperature changes the s-parameters by 1-2dB above 1.6GHz. Some portions of fabricated chip did not show the expected performance.
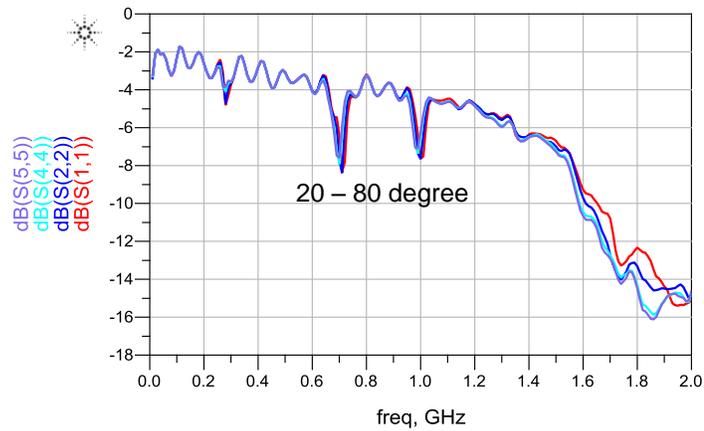
Figure 72   Measured scattering parameters of PDN with temperature variations.

These recent measurements showed the effects of thermal gradients on 3-D IC with performance degradations caused by high temperature gradients on delay and power ground impedance. The measured temperature gradients and thermo-electrical simulations were well correlated. The fabrication was intended to present simple but efficient compensation methods to reduce the delay degradation and demonstrated the implementation of the proposed methods. The methods have different capabilities for skew compensation with power and area overhead. The effectiveness and efficiency of the methods can be implemented with optimal combination for clock network design.

4.6 Summary

In this chapter, validation using two types of test vehicles was shown. FPGA-based and custom IC-based test vehicles were constructed to validate the electro-thermal simulation and proposed mitigation methods that included temperature dependency on delay and PDN impedance. Measurement results using the FPGA-based test vehicle and custom IC- based test vehicle showed good correlation of the delay mitigation, electrical-thermal simulation and other temperature effects.

Measurement results depict the proposed concepts to mitigate the effect of temperature on 3-D IC with performance degradations caused by high temperature gradients on delay and power ground noise. This chapter also presented detailed implementation of the proposed methods and design results using an FPGA and a recent CMOS process. The proposed concepts were successfully validated and correlated with the test vehicle measurements. To validate the performed electrical-thermal simulation, a fabricated chip using 180 nm process was utilized to confirm the accuracy of the solver. The measurement results also was used to determine thermal effect on PDN impedance.

CHAPTER 5 SYSTEM OPTIMIZATION FOR 3-D ICS

5.1 Introduction

Electrical-thermal analysis and correlation results in previous chapters validate the proposed co-design and simulation method. However, 3-D integrated systems have a large number of input control variables, as shown in Figure 73. Moreover, the multi-scale structure of 3-D systems increase compute time for analysis. In addition, trade-offs between computational cost and problem scale exist including nonlinearities and variations, as illustrated in Figure 73. Increase in operation speed and circuit density also results in much narrower system margins which require a large number of system parameters to be controlled. With smaller margins and more variation in the parameters, the effect of system parameters on the response need to be analyzed. This can be a very difficult and time consuming exercise.
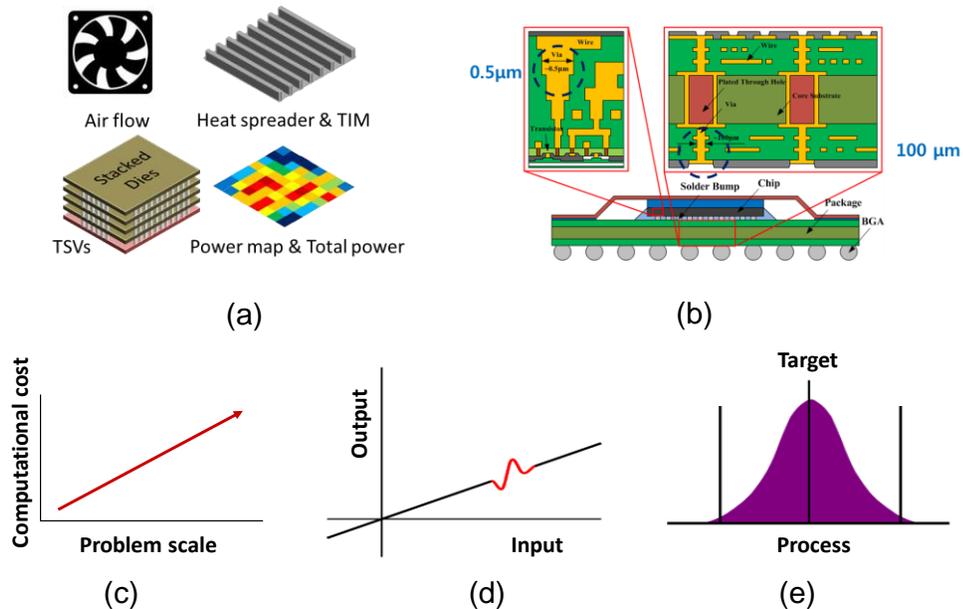
Figure 73 **Challenges of electrical-thermal simulations; (a) large number of input variables, (b) multi-scale structure (> 10^2), (c) computational cost, (d) non-linearity, and (e) process variation.**

To optimize a 3-D system design, combinations of the input system parameters increase the number of electrical-thermal simulations required thereby increasing computational resources. In addition, mesh density can significantly increase computational time, as shown in Figure 74. To optimize a large number of control/input parameters, several previous studies have proposed statistical methods, such as worst case and Monte Carlo analyses [72]. Because of the simulation cases and calculation overhead required by the methods, other studies have proposed approaches that reduce the number of simulations using design of experiments (DOE) [73]. Recently, machine-learning (ML) based methods have become popular and some of these methods have been applied to electromagnetic problems [74] and high speed interconnect systems [75]. These techniques have enabled machines to learn from the training sequences by accumulating data sets through automated learning algorithms.
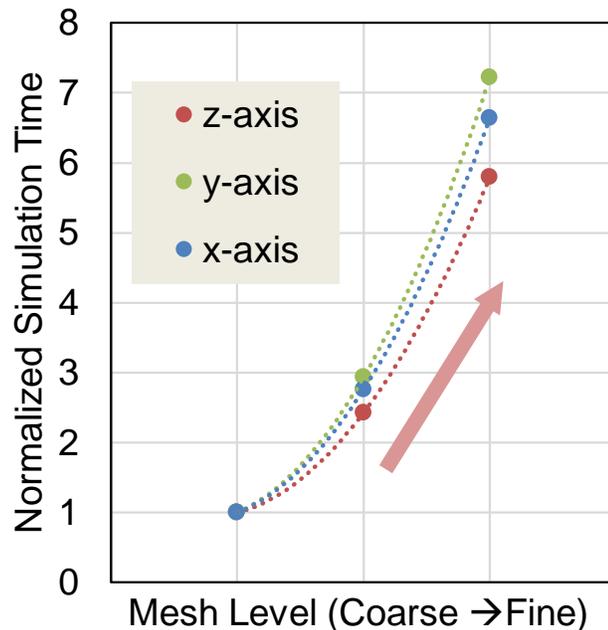
Figure 74   **Increase in simulation time from a coarse mesh level to a find mesh level.**

With the challenges related to electrical-thermal simulation for 3-D systems, a new approach is required for design optimization. Machine-learning (ML) can be a candidate for reducing computational time. Even though machine-learning approaches originated from computational learning theory in artificial intelligence, these methods are being widely used in various fields today that include computer vision, communication network, etc.

In this paper, we optimize electrical- thermal parameters of a 3-D system using ML algorithms by constructing models of the input-output characteristics based on error control, using data points that minimize the number of data sets. This method can be used to solve non-convex optimization problems as well. To minimize the number of training data sets required, we chose Bayesian optimization [76]. We developed a simulation flow with control parameters that includes physical dimensions, material properties, power maps, cooling technologies, and the material properties of thermal interface materials (TIM) that are used for thermal management.
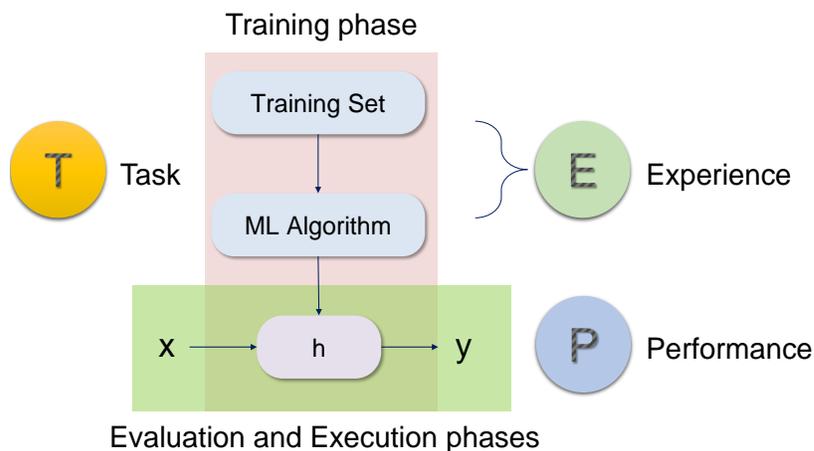
Figure 75　　Concept of machine learning consists of training and evaluation/execution phases [81].

As illustrated in Figure 75, machine-learning has three components, task, experience, and performance which consists of two phases, training and evaluation/execution. "Task" and "performance" represent training and target respectively, while "experience" is used to improve the target performance [77].
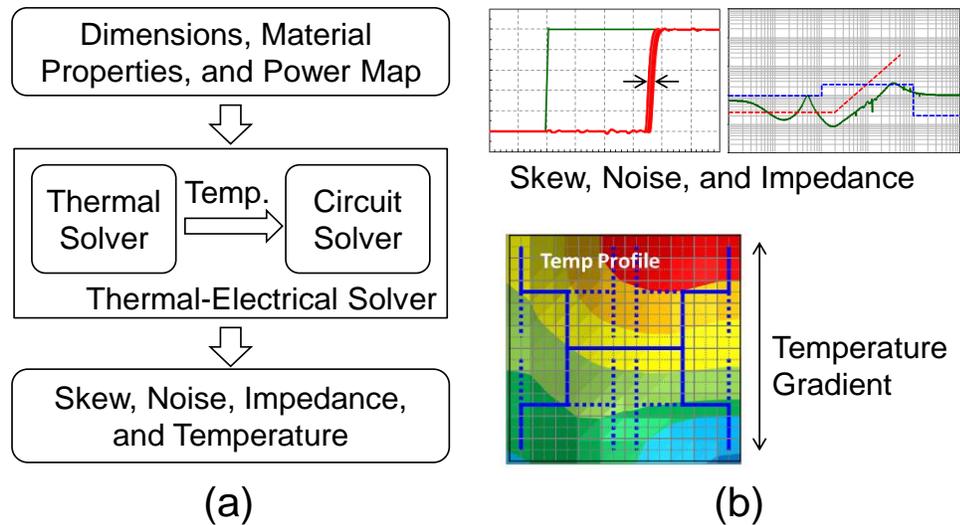
Figure 76   **(a) Electrical-thermal analysis for 3-D systems. (b) Simulated skew, noise, PDN impedance, and temperature gradient [37].**

Figure 76 shows the flow used for electrical-thermal analysis for 3-D systems that has been used in this dissertation (in the earlier chapters) to compute temperature gradient and clock skew. The objective is to optimize the electrical and thermal performance such as clock skew, maximum temperature, and temperature gradients using control parameters. The available control parameters can be chosen based on sensitivity analysis, as shown in Figure 77. The sensitivity analysis is based on sweeping each parameter while maintaining the others at their nominal value. The analysis results provide insight into which parameter is more dominant for achieving target performance. From Figure 76, the control parameters lead to both linear and non-linear relationship with the target parameters such as skew, temperature gradient and maximum temperature.
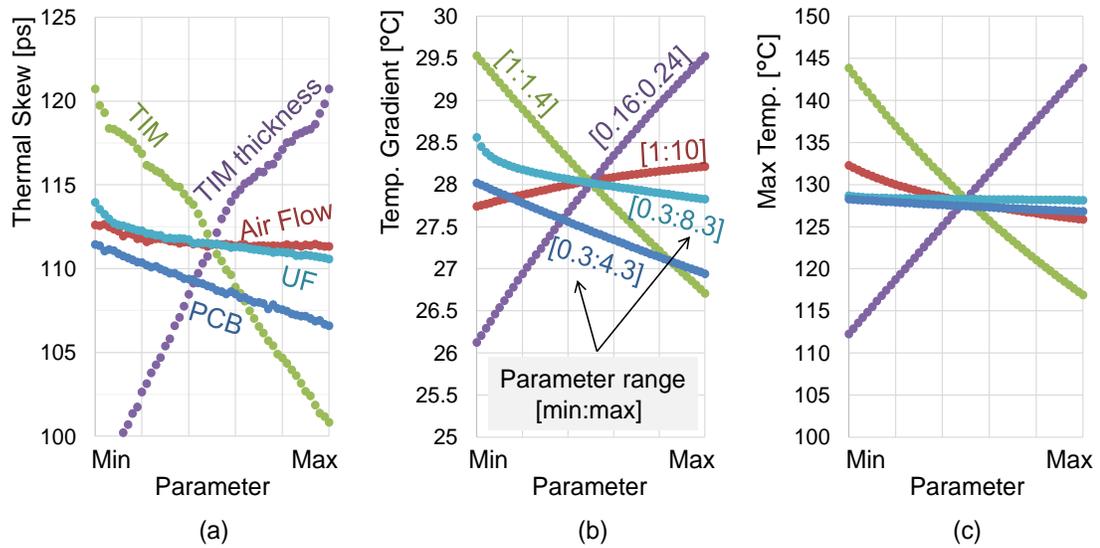
Figure 77   **Sensitivity of parameters with the available range of each parameter: (a) thermal skew, (b) temperature gradient, and maximum temperature.**

5.2 Machine-Learning (ML) Algorithms for Optimization

Machine-learning methods, for example, decision trees, neural networks, and Bayesian optimization have been widely studied in the computer science community. Decision trees use a tree as a predictive model with mapping. Neural networks are based on the structure of biological neural networks. Bayesian networks use a probabilistic graphical model. Table 15 compares a few major machine learning algorithms. In this dissertation our focus is on the use of Bayesian based ML algorithms for 3-D system optimization.

Table 15   **Comparison of classification-based machine-learning algorithms**

Algorithm	Properties	Performance	Application
Decision Trees	Easy, expressive, non-parametric	Splitting criteria, overfitting, interactions	Discrete value, noisy & missing data
Neural Networks	Hard, finite time for linear data	Sensitive to parameter, overfitting	Real/discrete, errors in training data
Bayesian Networks	Simple, efficient, and stable	Fast training and data analysis	Modeling with error approximation

5.2.1 Bayesian Optimization (BO)

Bayesian optimization originated from a well-known equation in probability theory and statistics, namely Bayes' theorem. Bayes' theorem [78] is described by the following equation:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{18}$$

where 'A' and 'B' are events and 'P (A)' and 'P (B)' are the probabilities of 'A' and 'B' without regard to each other. 'P (A|B)' and 'P (B|A) are the conditional probability of observing event 'A' given that 'B' is true and observing event 'B' given that 'A' is true, respectively.

$$P(h|D) = \frac{P(D|h)\,P(h)}{P(D)} \tag{19}$$

We can expand Equation (18) to machine learning, especially for the hypothesis space (h) and training data (D). P (h|D) is the probability of hypothesis 'h' given data 'D' and is called the posterior. 'P (D)' and 'P (h)' are probabilities of observing 'D' and 'h'

independent of data, respectively. They are referred to as the prior over data 'D' and hypothesis 'h', respectively. 'P (D|h)' is the probability of observing data 'D' given a hypothesis 'h' and is referred to as the likelihood.

$$\mathbf{P}(\boldsymbol{h}|\boldsymbol{D}) \propto \frac{\mathbf{P}(\boldsymbol{D}|\boldsymbol{h})\, \mathbf{P}(\boldsymbol{h})}{\mathbf{P}(\boldsymbol{D})} \tag{20}$$

Equation (19) interprets Bayes' rule regarding possibilities of multiple events, before (prior to) and after (posterior to) event. From Equation (20), the proportionality symbol indicates that if 'h' varies but keeping 'D' fixed, the left hand side is equal to a constant times the right hand side. In words, posterior is proportional to prior times likelihood: determined by the Bayes factor [78]. This Bayesian method reduces the number of datasets required for meeting target output performance.

In Bayesian statistics, we model our uncertainty with a prior probability distribution. In other words, we estimate the distribution and use this information to decide the point evaluated next, which is a key point of BO that differentiates it from other methods.

For Gaussian process priors, the model uses a joint Gaussian with the whole set of points. In this dissertation, the function f is defined as a GP prior with mean function 'm' and covariance function $'k'$. Based on prior observation points, the prior function $f(\mathbf{x}_{1:N})$ is defined as a Gaussian process (GP) given by:

$$\mathbf{f}(\mathbf{x}_{\mathbf{1}:N}) = \boldsymbol{N}(\boldsymbol{\mu}(\mathbf{x}_{1:N}), \boldsymbol{k}) \tag{20}$$

where $\mathbf{x}_{1:N}$ represent the N prior observation point, $\boldsymbol{\mu}(\mathbf{x}_{1:N})$ is the mean vector and k (also called the kernel) is the covariance matrix given by [85]:

$$\mu(x_{1:N}) = [\mu(x_1)\,\mu(x_2)\cdots\mu(x_N)]^T \tag{21}$$

$$\mathbf{k}(x_{1:N}) = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix} \tag{22}$$

where the covariance is defined by:

$$k(x_N, x_{N+1}) = \exp\left(-\frac{1}{2}\|x_{N+1} - x_N\|^2\right) \tag{23}$$

To predict $f(\mathbf{x}_{N+1})$ at the next data point, we consider the joint distribution over f of the old data points and new data point, as shown in Equation (25). The optimization problem now relates to maximizing (or minimizing) $f(\mathbf{x})$ subject to $\mathbf{x}$ where $f(\mathbf{x}_{N+1})$ can be a non-convex black-box deterministic function defined by:

$$\begin{bmatrix} f(x_{1:N}) \\ f(x_{N+1}) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu(x_{1:N}) \\ \mu(x_{N+1}) \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ k^T & k(x_{N+1}, x_{N+1}) \end{bmatrix}\right) \tag{25}$$

where $\mathbf{K}$ is the kernel matrix and $\mathbf{k}$ is the kernel function given by Equations (23) and (24). From [85], the mean and variance of $f(\mathbf{x}_{N+1})$ can be computed as:

$$\mu(x_{N+1}) = \mathbf{k}^T \mathbf{K}^{-1} \mathbf{f}_{1:N} \tag{24}$$

$$\sigma^2(x_{N+1}) = k(x_{N+1}, x_{N+1}) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \tag{25}$$

This approach provides a posterior distribution of the unknown function. We can choose the next value of the function representing the targeted values by either maximizing or minimizing an acquisition function (explained later). With this approach, Bayesian optimization can show faster convergence as compared to conventional optimization methods, such as for example random optimization [79], as illustrated in Figure 78.
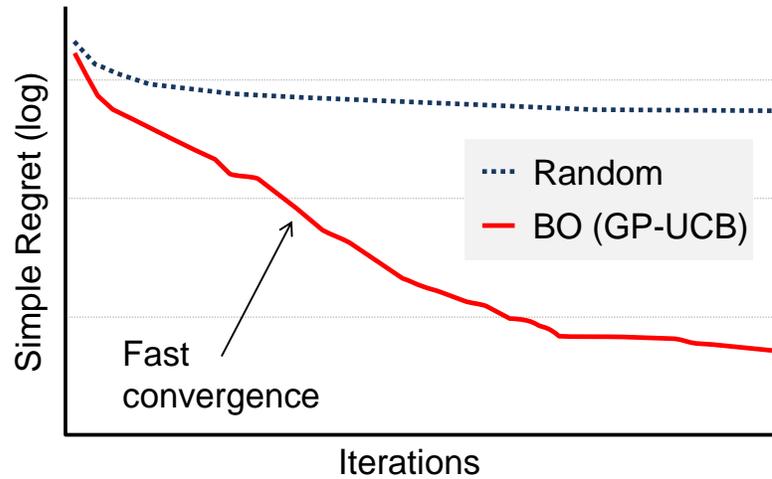
Figure 78   **Performance of Bayesian optimization algorithm compared with random optimization [79].**

The typical flow of Bayesian optimization (BO) using Gaussian process (GP) [82] is as follows:

   i.   Choose initial points of $\mathbf{x}$ and evaluate $\mathbf{f}(\mathbf{x})$ including error (wrt the target value desired).

   ii. While the stopping criterion is not met, calculate Bayesian posterior distribution on '$\mathbf{f}$' from the points observed.

   iii. Using the prior observation points and acquisition function determine the point to evaluate next.

   iv. Stop if the criterion is met, and report the point with the best value.

This approach is based on the Infinite-Metric GP optimization (IMGPO) algorithm presented in [83].

Bayesian optimization predicts posterior using prior, observation, and acquisition function (shown as Lowest Confidence Bound or LCB in the table), as shown in Figure 79, where the maximum of the acquisition function is used to determine the next evaluation point using the mean and variance of the posterior. A more detailed flow is discussed in the next section.
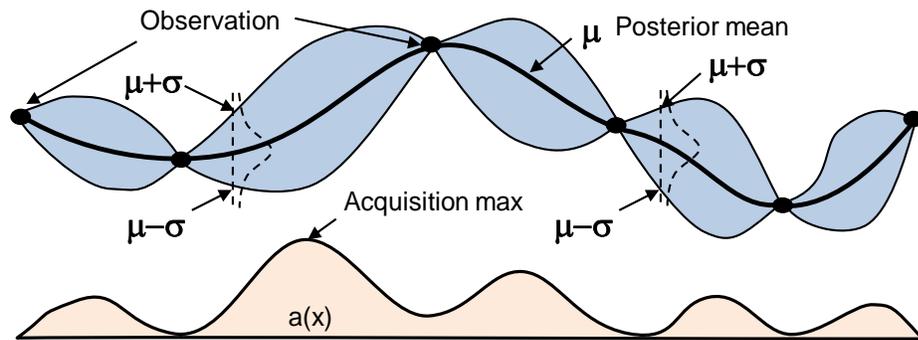
Figure 79   **Bayesian optimization with prior, observation, acquisition function, and posterior [80].**

5.2.2 Flow of Electrical-Thermal Simulation including Bayesian Optimization
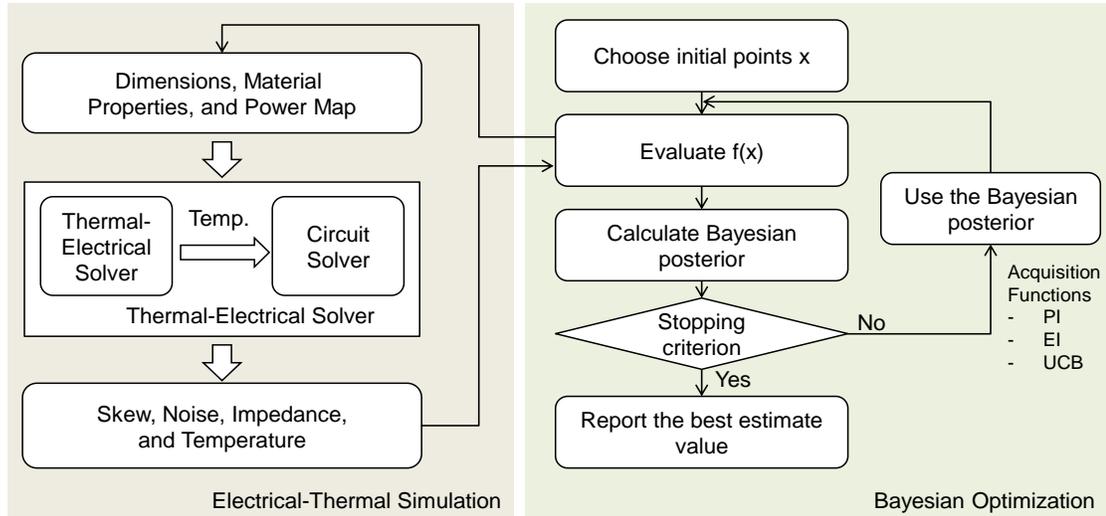
Figure 80   Flow of Bayesian optimization [81].

Figure 80 shows the full flow for optimization where electrical-thermal simulations are combined with Bayesian optimization. In the flow chart, acquisition functions are used to choose the posterior. Following three acquisition functions are widely used in GP based optimization, namely [76]:

      i.   Probability of improvement (PI).

     ii.   Expected improvement (EI).

   iii.   GP upper/lower confidence bound (UCB/UCB).

The goals of the first two strategies are to maximize the probability of improvement and the expected improvement of the current value, respectively. The third strategy is targeted towards exploiting upper or lower confidence bounds with high probability using acquisition functions that minimize regret [76]. For this optimization, a confidence bound-

96

based acquisition function (also called Lowed Confident Bound or LCB) is used, as shown in Equation (28):

$$\mathbf{x}_{N+1} = \arg\min\left[\boldsymbol{\mu}(\mathbf{x}_i) - \boldsymbol{\kappa}\boldsymbol{\sigma}(\mathbf{x}_i)\right]$$

(28)

where $\kappa \geq 0$ and $\kappa = \sqrt{2\log\pi^2 x^2 / 12\nu}$ (where ν equals 0.05), and $\mu(\mathbf{x}_i)$ and $\sigma(\mathbf{x}_i)$ are determined from Equation (26) and (27) for each input parameter.

Recent studies on Bayesian optimization with GP have presented faster convergence without requiring extensive sampling [83] and have tried to connect PI and UCB/LCB approaches [84]. We selected this algorithm for our optimization. This approach provides a posterior distribution of the unknown function. We can choose the next value of the function to search and move closer towards the targeted goal. Figure 81 shows the procedure for calculating the next value from mean, variance, and maximum of the acquisition function [85]. The maximum of the acquisition function is found using posterior mean and variance calculated using Gaussian distribution. The point of maximum acquisition function becomes the next observation point. In Figure 82, the actual function is also shown as "truth".
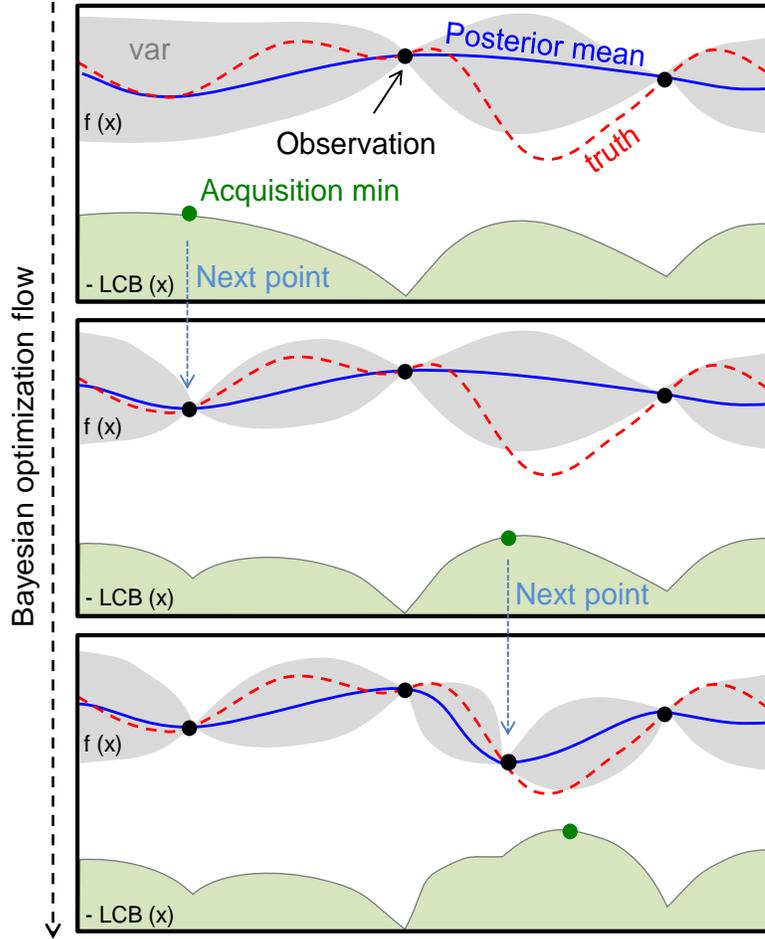
Figure 81   **Bayesian optimization flow with prior, observation, acquisition function, and posterior [85].**

The Bayesian Optimization algorithm described earlier for one-dimensional problems can be extended to multi-dimensional problems as well. Here, we assume that the M random input variables are independent of each other and Equation (21) can be cast in the form:

$$
\begin{aligned}
f\big(X_{1,1:N}, X_{2,1:N}, \ldots\ldots, X_{M,1:N}\big) \\
= N\big(\mu_1(X_{1,1:N}), k_1\big), N\big(\mu_2(X_{2,1:N}), k_2\big), \ldots N\big(\mu_M(X_{M,1:N}), k_M\big)
\end{aligned}
\tag{29}
$$

where N prior observation points are used and $\mu_{i,i=1,M}$ and $k_{i,i=1,M}$ are the mean and co-variances for the corresponding functions. While determining the next point for

$X_{i,i=1,M}$ acquisition functions are defined and chosen as in Equation (28). As an example, Figure 82 shows the distribution of the function with two random variables X_1 and X_2, posterior mean and variance, and the acquisition function defined using the lower confidence bound (LCB). The minimum of the acquisition function is chosen, as shown using a triangle marker, which becomes the next observation point for Bayesian optimization. In Figure 83, X_1 represents air flow as heat transfer coefficient in W/(m²·K) and X_2 represents thermal conductivity of TIM material in W/(m·K) in the 3-D System.
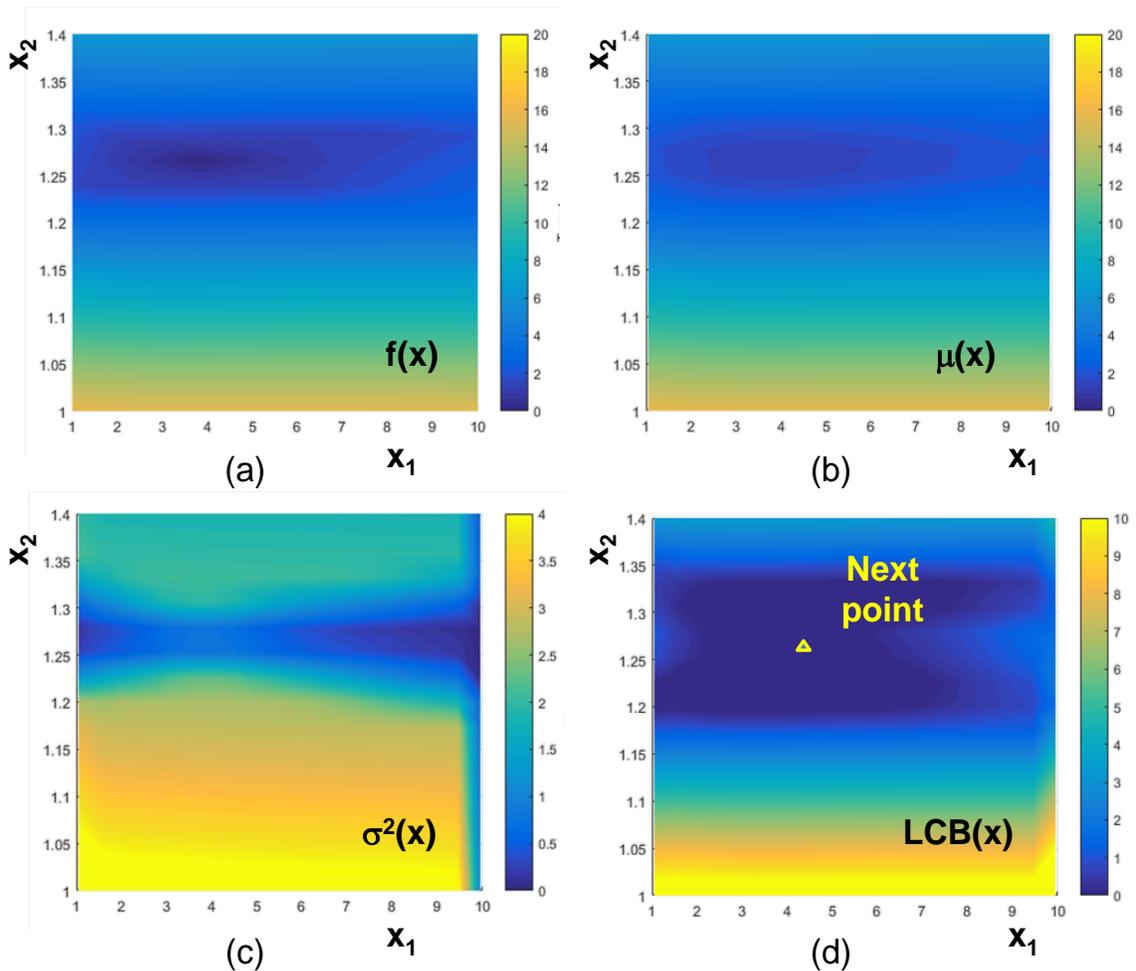
Figure 82   Distribution plots of (a) Function, (b) posterior mean, (c) posterior variance, and (d) LCB acquisition function in 2-D optimization of a 3-D system [81].

5.3 Application of ML Algorithms

The objective of electrical-thermal simulation is to estimate the global skew by considering temperature gradients in a 3-D system. Simulated temperature profiles superimposed on to temperature-sensitive electrical properties and recent technology-based buffer and interconnect models were used, as described in Chapter 2. This analysis procedure also includes power delivery models to estimate power ground noise.

To optimize electrical-thermal performance, a large number of control parameters need to be considered. Physical dimensions and material properties are some of the basic parameters required for electrical-thermal analysis. For thermal design, thermal related parameters such as power map, total power, fan speed, micro-fluidic rate, TIM material property, TSV numbers, etc need to be optimized. The goal of the optimization is to tune these parameters so that appropriate measures such as temperature, clock skew, PDN noise and impedance are reached, as shown in Figure 83.
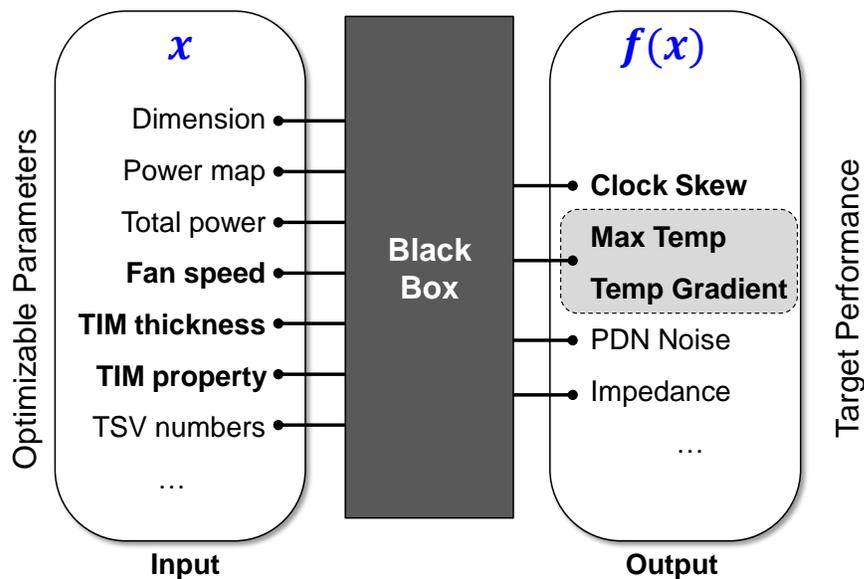
Figure 83   **Input and output parameters for 3-D system design [81].**

Table 16 shows key input variables for optimization of thermal performance which also affects electrical performance. For design optimization, a data range for the following parameters were specified: heat transfer coefficient of air flow (K_A) ∈ [1:50] W/(m²·K), TIM material property (K_{TIM}) ∈ [1.0:1.4] W/(m·K), TIM thickness (t_{TIM}) ∈ [0.16:0.24], and thermal conductivity of PCB material (K_{PCB}) ∈ [0.3:4.3] W/(m·K). These are shown in Table 16. Output parameters for optimization are maximum temperature and temperature gradient in degree Celsius and thermal skew in ps.

Table 16   Input variables for optimization of thermal and electrical performance.

Parameter	Unit	Min	Nominal	Max
Heat transfer coefficient	W/(m²·K)	1	5	50
TIM material	W/(m·K)	1.0	1.2	1.4
TIM thickness	mm	0.16	0.20	0.24
Underfill material	W/(m·K)	0.3	0.3	8.3
PCB material	W/(m·K)	0.3	0.3	4.3

5.4 Analysis

5.4.1 One-Dimensional Optimization

In this chapter, one-dimensional optimization with other parameters at their nominal value was used a starting point for applying BO for these types of problems. With initial optimization of parameters in the total design space as shown in [86], optimization results show an intuitive direction towards extreme values of input parameters leading to highest air flow, highest TIM material thermal conductivity, and lowest total power.

However, our objective is not to obtain the lowest skew but to achieve a target value of 110 ps. Due to fast convergence of the algorithm to our design target value, the optimized input parameter using minimum number of iterations can be obtained, as shown in Figure 84 and Figure 85. The machine learning algorithm was able to achieve the target parameters by starting with an initial guess and not requiring a sweep of the parameters, as is typically done while creating response surfaces. Optimization results show fast convergence to optimal values, as shown in Figure 84 and Figure 85. With initial optimization of parameters in the design space, optimization results provide optimal values of input parameters, as shown in Table 17. Multi-dimensional optimization is described in the next section.
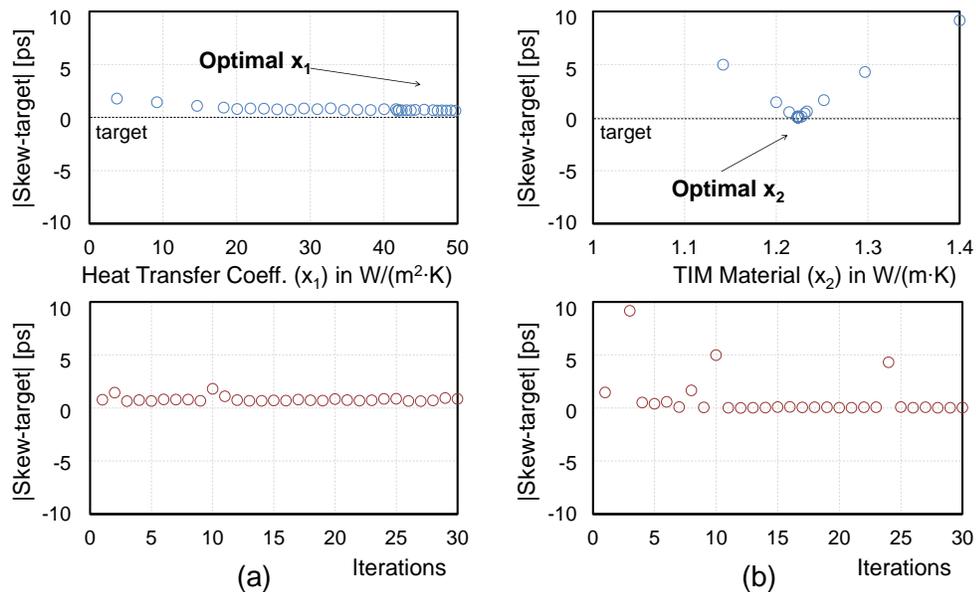
Figure 84 Optimization of each input parameter for clock skew of 110ps showing convergence. (a) Heat transfer coefficient and (b) thermal conductivity of TIM [86].
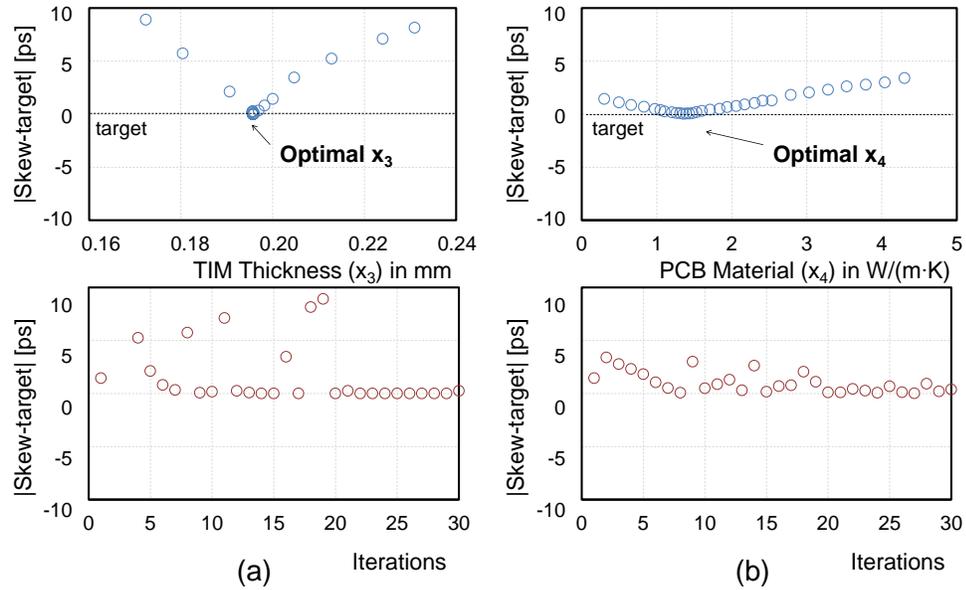
Figure 85   **Optimization of each input parameter for clock skew of 110ps showing convergence. (a) Thickness of TIM and (b) thermal conductivity of PCB [86].**

Table 17   **One-dimensional optimization results.**

	Air flow (X₁)	TIM material (X₂)	TIM thickness (X₃)	PCB material (X₄)
Optimizer	50	1.22	1.96	1.36
Input domain	[1:50]	[1:1.4]	[0.16:0.24]	[0.3:4.3]
Sensitivity	weak	strong	strong	intermediate

5.4.2 *Multi-Dimensional Optimization*

Preliminary analysis focused on one dimensional optimization, as described in the previous section. The number of parameters can be expanded for more practical optimization scenarios. Thermal characteristics such as maximum temperature show more linear dependency with material parameters and dimensions. Figure 86 shows the optimization result with three parameters, heat transfer coefficient of air flow (X_1), TIM material (X_2), and TIM thickness (X_3), where Figure 86(a) shows the convergence towards

the optimum value and Figure 86(b) shows the temperature distribution before and after optimization.
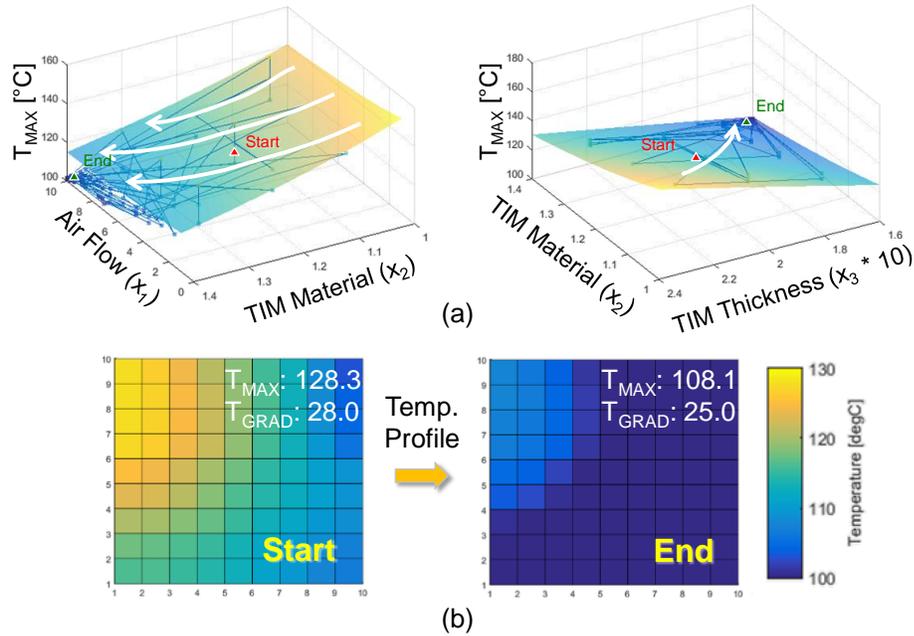
Figure 86 **(a) Optimized parameters for lowering the maximum temperature and (b) optimized temperature [81].**

In Bayesian optimization, appropriate selection of hyper-parameters can improve optimization performance. Covariance function which specifies the covariance between pairs of random variables is one of these hyper-parameters. For the optimization in this dissertation, we used the squared exponential for the covariance based on [85], as shown in Equation (21). However, a linear covariance function can also be explored which is faster because the computing cost reduces from $O(N^3)$ to $O(N)$. Therefore, optimization performance of the algorithms can vary with the selection of the covariance function.

The parameters used show mainly linear relationship with thermal performance (ex: maximum temperature). However, a combination of multiple target objects can lead to non-linear behavior. Multiple objectives, such as maximum temperature, temperature gradient, etc. can be combined using weights to define a target function, as shown in Equation (24),

where BO can be used to minimize such a cost function. Values of weights based on the relative importance of each objective.

$$f(T_{MAX}, T_{GRAD}, \dots) = \sum_{i=1}^{N} w_i * y_i \qquad \text{(24)}$$

In equation (24), w_i and y_i are weighting coefficients and selected outputs, respectively.

In this dissertation, the target function is defined using the equation:

$$f(T_{MAX}, T_{GRAD})$$
$$= w_{TMAX} * y_{TMAX} + w_{TGRAD} * y_{TGRAD} \qquad \text{(26)}$$

where w_{TMAX} and w_{TGRAD} are the weighting coefficients for maximum temperature and temperature gradient, respectively; y_{TMAX} and y_{TGRAD} are output values of maximum temperature and temperature gradient, respectively.

Two objectives, maximum temperature and temperature gradient, with air flow as the input parameter shows an optimal point in between the minimum and maximum values. According to Figure 87, faster the air flow speed, the lower is the maximum temperature (Figure 87 (a)), but larger is the temperature gradient (Figure 87 (b)). Combining these objectives can provide an optima in the design space as shown in Figure 87. In Figure 87, weights of $w_{TMAX} = 0.34$ and $w_{TGRAD} = 4.5$, were used to define the target function.
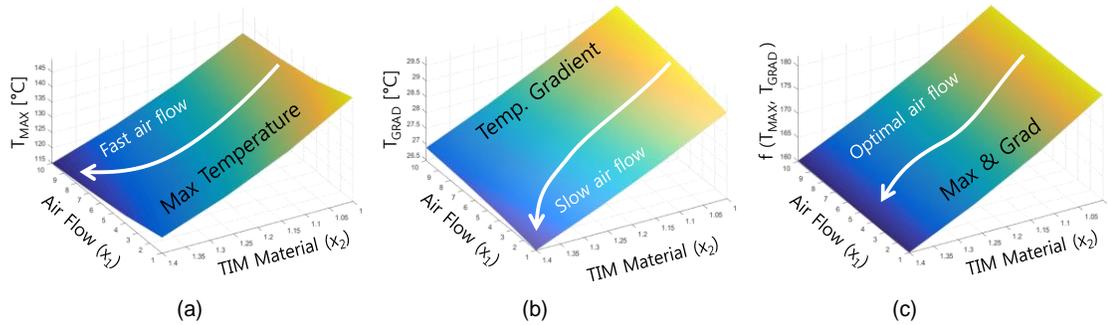
Figure 87 **Non-linear behavior applied to Bayesian optimization using multiple objectives: (a) maximum temperature, (b) temperature gradient, and (c) both maximum and gradient.**

By combining the maximum temperature and temperature gradient using weights as in Equation (23), an optimum point can be determined. In the Bayesian optimization approach, target values can be provided. Some of the electrical-thermal performances are straight-forward from an optimization perspective. For example, better material properties must provide better performance. However, better material properties usually require more resources and higher cost from design and manufacturing perspective. Seeking the best of parameters generates optimal value at the extreme points, but optimization with target performance tends towards a global optima in the design space. Figure 88 illustrates non-linear behavior with two input variables, namely TIM material and thickness, maximum temperature as the target value.
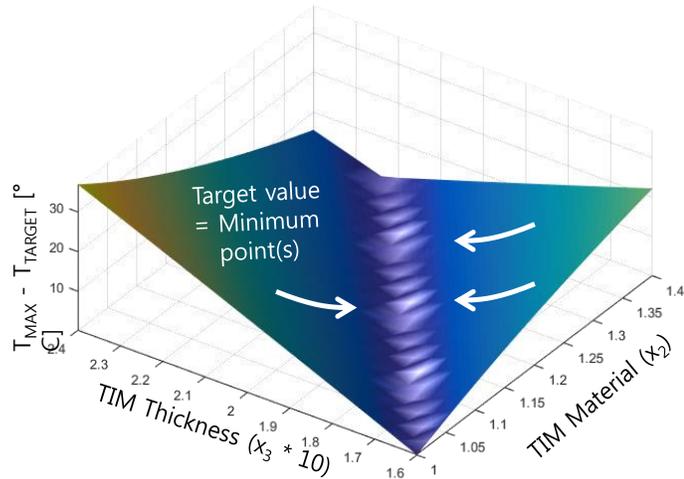
Figure 88   Non-linear trend showing minima with a target value.

To verify the efficiency of the optimization procedure, a case study was performed with various power maps shown in Chapter 2 with more number of input variables (N = 5) and iteration of 200. Figure 89 and Figure 90 show the results with power map II and power map III, respectively. The target values used for T_{MAX} and T_{GRAD} were 110°C and 10°C for power map II and 110°C and 8°C for power map III, respectively. Optimization results show convergence to the target value in Figure 89 and Figure 90. The temperature distribution before and after optimization are also shown in Figure 89 and Figure 90. Before optimization, power map II and power map III resulted in a clock skew of 51.8 ps and 39.2 ps, respectively. After optimization, power map II and power map III resulted in a clock skew of 44.2 ps and 33.0 ps, respectively. Optimization results are shown in Table 15.

Table 18   Optimization results with various power maps.

Power map	T_{MAX} [°C]		T_{GRAD} [°C]		Skew [ps]	
	Before	After (%)	Before	After (%)	Before	After (%)
I	127.4	115.3 (-9.5%)	27.4	25.0 (-8.8%)	108.9	93.7 (-14.0%)
II	120.6	110.1 (-8.7%)	12.7	10.8 (-15.0%)	51.8	44.2 (-14.7%)
III	118.8	109.9 (-7.5%)	10.0	8.8 (-12.0%)	39.2	33.0 (-15.8%)

To verify the efficiency of the optimization procedure, a case study was performed with various power maps shown in Chapter 2. Figure 89 and Figure 90 show the results with power map II and power map III, respectively. Because of different power maps, targets of the maximum temperature and temperature gradient are varied. Target values of T_{MAX} and T_{GRAD} used were 110°C and 12.5°C for power map II and 110°C and 9.5°C for power map III, respectively, where the weights used for maximum temperature and temperature gradient were 0.34 and 4.5, respectively,. Optimization results show convergence to the optimal value and output performance close to the target. After optimization, power map II and power map III resulted in a clock skew of 52.8 ps and 35.8 ps, respectively.
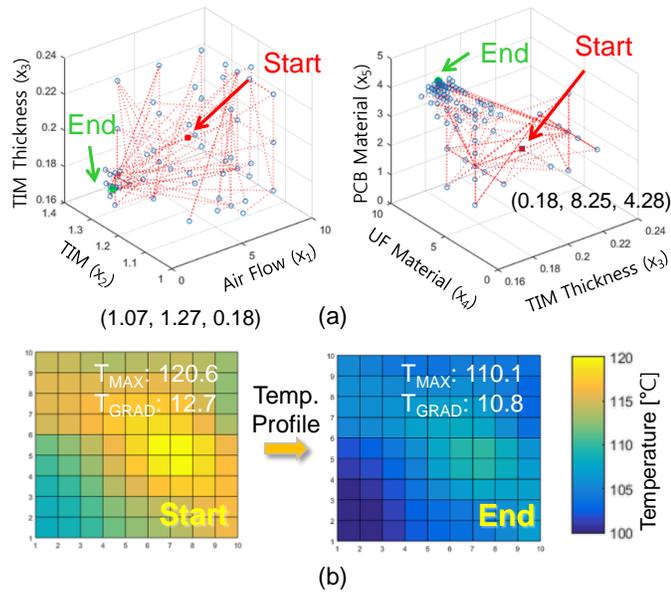
Figure 89   Optimization with power map II; (a) Iterations shown as a function of 3 parameters only and (b) temperature distribution [81].
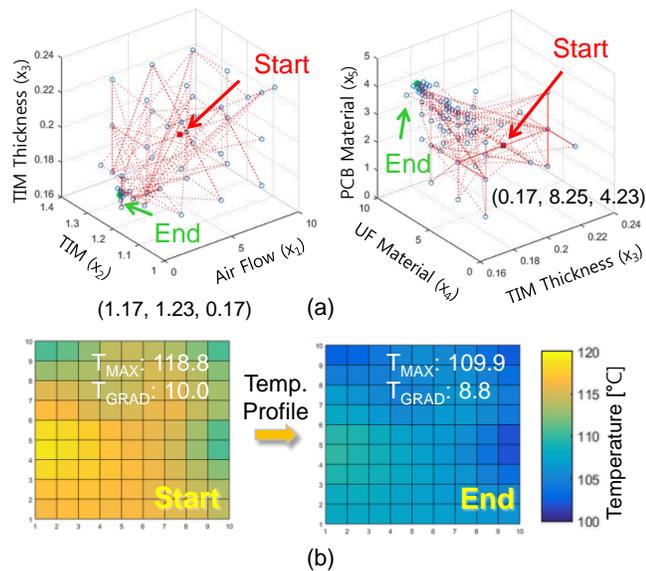
Figure 90   Optimization with power map III; (a) Iterations shown as a function of 3 parameters only and (b) temperature distribution [81].

To compare the optimization performance with existing methods and algorithms, the number of iterations and optimized values were compared. Figure 91 compares the optimization results, for temperature gradient and the resulting skew for power map I and five parameters, using Bayesian optimization (BO) and *'pattern search'* algorithm

(available in MATLAB) after 100 iterations where BO produced temperature gradient and thermal skew of 24.5 °C and 96.2 ps vs 23.8 °C and 88.0 ps using global search algorithm, as illustrated optimization results by using an existing global optimization algorithms and Bayesian optimization with the same number of iterations.
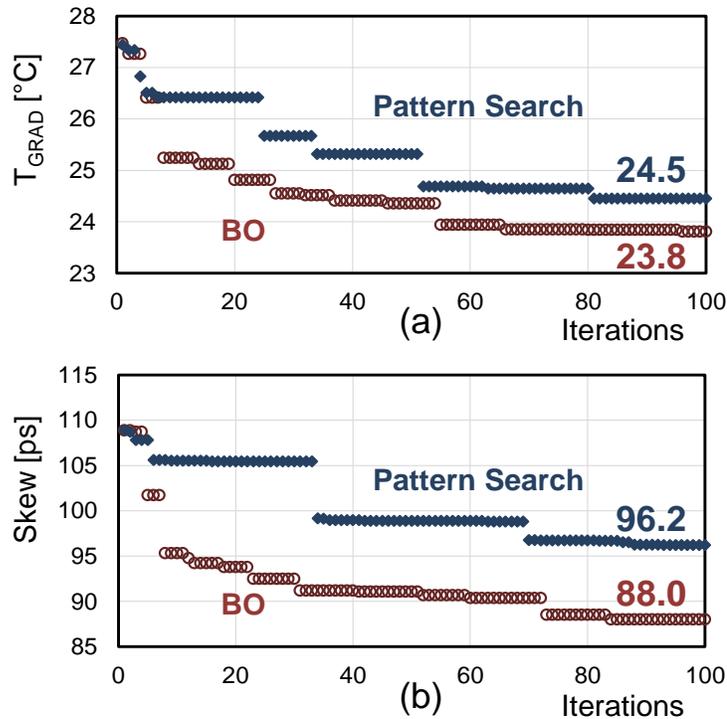
Figure 91   **Comparison of convergence between BO and Pattern Search [81].**

5.5 Summary

This chapter presented machine-learning, namely use of Bayesian optimization (BO) algorithm, to optimize electrical-thermal performance of 3-D integrated circuits and systems. Optimization results showed that this approach is applicable for electrical-thermal simulations. We found that the method can be applied to system-level electrical-thermal simulations, (which often require long simulation time and a large number of simulation cases), is accurate and requires lower computational cost as compared to conventional

design optimization methods. This approach shows the capability of handling a large number of input parameters with fast convergence and flexibility. The optimization approach using machine learning methods can become useful when system complexity increases along with many parameters that need to be optimized simultaneously, especially for 3-D applications. Since many BO algorithms have been presented in the open literature, the efficiency of the optimization described in this dissertation can be increased further.

CHAPTER 6   CONCLUSIONS AND FUTURE WORK

6.1 Summary and Conclusion

This dissertation discussed some challenges in 3-D system analysis and design. As discussed in Chapter 1, one challenge relates to electrical-thermal analysis that include interactions. Inclusion of these interactions for electrical-thermal modeling, thermal induced clock skew, and analysis of 3-D PDN were discussed in Chapter 2. In Chapter 3, mitigation methods for reducing thermal skew were proposed. The proposed methods include 1) adaptive bias, 2) controllable delay, and 3) variable strength that show design flexibility, efficient performance, and tolerable overhead. The proposed approaches including the thermal solver were verified using two test vehicles in Chapter 4. These test vehicles include extensive design, measurement, and correlation. Chapter 5 discussed application of machine learning methods for 3-D system optimization. Chapter 5 showed that the computational resources for electrical-thermal simulations can be reduced with ML-based methods, when control parameters need to be tuned.

The output of this dissertation lead to the following conclusions:

1. The electrical-thermal analysis which includes Joule Heating captured the temperature gradient and its impact on delay, PDN impedance, and power/ground noise. The results show that the temperature gradient in a 3-D system is critical because the temperature affects and degrades clock skew. In addition, a new delay modeling formula was developed. Power delivery network analysis showed that PDN impedance and noise are also affected by temperature, and therefore thermal-aware simulation and design are required.

2. To mitigate thermal effects in PDN design for a 3-D system, a design approach using optimization of on-chip decoupling capacitors is required. Thermal-aware PDN simulation, including a case study with 3-D memory for design that included temperature effects on decoupling was presented.

3. Mitigating thermally induced clock skew is important and to reduce thermal effects on delay three methods were proposed: 1) adaptive voltage, 2) controllable delay, and 3) variable strength. The proposed methods have the ability of active compensation and showed better performance, design flexibility, and efficiency with minimal overhead as compared to available methods.

4. Verification of the electrical-thermal solver and skew mitigation techniques are important. To verify the proposed concepts and simulation tools, two test vehicles were designed and fabricated. The fabricated board with FPGA and custom IC further confirmed the findings.

5. A method based on ML with Bayesian optimization using Gaussian process showed improved optimization results as compared to other techniques, when applied to 3-D systems.

6.2 Future Work

This dissertation covers electrical-thermal modeling, simulation, analysis, solution, verification, and optimization in the context of 3-D ICs and systems. Further work is necessary in the following area:

1. With regard to modeling, the solver calculated steady-state thermal and electrical characteristics, so time-transient analysis with various power and operating conditions

are necessary. Though such a solver has been developed [13], the results need to be validated.

2. Compensation methods proposed in this dissertation are shown at a conceptual level. Though verified, they need to be validated on full chip implementations to show practicality.

3. This dissertation has shown that optimization using machine-learning is useful for 3-D ICs. However, more work is necessary to optimize the method with appropriate acquisition functions and optimal setting of hyper-parameters.

6.3 Contributions

The objective of this dissertation was to develop a co-design method for design of 3-D ICs and systems. With these objectives, the contributions of this dissertation are as follows:

1. Electrical-thermal co-simulation method for 3-D integrated ICs and systems

A temperature-aware simulation flow for signal, power, and thermal integrity was developed. The presented method allows modeling and simulation for 3-D ICs and systems that tend to have higher thermal density and temperature gradients. In this dissertation, the electrical-thermal solver developed using the finite volume method [13] was extended to include temperature and Joule heating effects with various system parameter variations. In addition, the full set of electrical-thermal modeling for 3-D systems enables co-simulations for various applications.

2. **Quantification and relationship of temperature gradient, thermal delay, clock skew, PDN impedance, and noise in 3-D integration**

The proposed electrical-thermal analysis showed the amount of temperature gradient and impact on thermal delay, PDN impedance, and power/ground noise. The results show that temperature gradient in 3-D systems is crucial because the temperature affects and degrades clock skew. In addition, a new delay modeling formula was developed. Using analysis on a power delivery network, it was shown that PDN impedance and noise are also affected by temperature, so thermal-aware simulation and design are required.

3. **PDN design approach for mitigating thermal effects**

To mitigate thermal effects in PDN design for 3-D system, a design approach using optimization of on-chip decoupling capacitors in a die was presented. Thermal-aware simulation showed that PDN with increased resistance parasitics of ODCs need to be analyzed at operating frequency range. A case study using high-density 3-D memory showed practicality of the approach.

4. **Delay compensation methods for mitigating thermal effects**

Three methods to reduce thermal effects on delay were proposed. These methods are; 1) adaptive voltage, 2) controllable delay, and 3) variable strength. The proposed methods have comparable ability of active compensation and showed performance improvement as compared to previous methods. In addition, though these methods originated from existing approaches, they showed better performance, design flexibility, and efficiency with minimal overhead.

5. **Hardware correlation**

As a verification procedure, two types of test vehicles were designed and fabricated to validate the proposed method and thermal solver. The fabricated FPGA-based and custom IC-based test vehicles showed that compensation techniques developed word and provide good results. In addition, the thermal measurement results correlated well with thermal simulation.

6. **Optimization of 3-D system using machine-learning**

Design optimization using a new approach, machine-learning, was presented to more effectively perform system optimization for 3-D systems. The proposed approach using Bayesian optimization with Gaussian process, showed better optimization performance and efficiency compared to other existing techniques. The analysis showed that the approach is applicable for the optimization of electrical-thermal performances arising in 3-D integrated systems.

6.4 Publications

The following journal and conference papers have been submitted, accepted, and published with this dissertation research.

6.4.1 Journals

▪ **S. J. Park**, N. Natu, M. Swaminathan, "Analysis, design, and prototyping of temperature resilient clock distribution network for 3-D ICs," *IEEE Trans. on*

Components, Packaging and Manufacturing Technology (CPMT), pp. 1669-1678, Oct. 2015.

- **S. J. Park**, M. Swaminathan, "Application of Machine-Learning for Optimization of 3-D Integrated Circuits and Systems," *Submitted to IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, 2016.

6.4.2 Conferences

- **S. J. Park**, H. Yu, M. Swaminathan, "Design Optimization Using Machine-Learning for Electrical-Thermal Performance of 3-D Systems," *To be submitted to IEEE Electrical Design of Advanced Packaging & Systems (EDAPS) Symposium*, 2016.

- **S. J. Park**, H. Yu, M. Swaminathan, "Preliminary Application of Machine-Learning Techniques for Thermal-Electrical Parameter Optimization in 3-D IC," *Accepted to IEEE Signal and Power Integrity Conference (SIPI 2016),* July 2016.

- **S. J. Park**, M. Swaminathan, "Temperature-aware power distribution network designs for 3D ICs and systems," *IEEE Electronic Components and Technology Conf. (ECTC)*, pp. 732-737, May 2015.

- **S. J. Park**, N. Natu, M. Swaminathan, "Design and early validation (using FPGA) of temperature resilient clock distribution networks for 3D ICs," *Proc. IEEE 23rd Conf. of EPEPS*, pp. 127-130, Oct. 2014.

- **S. J. Park**, N. Natu, M. Swaminathan, B. Lee, S. M. Lee, W. H. Ryu, K. S. Kim, "Timing analysis for thermally robust clock distribution network design for 3D ICs," *Proc. of IEEE 22nd Conf. of EPEPS*, pp. 69-72, Oct. 2013.

- S. K. Kim, S. Telikepalli, **S. J. Park**, M. Swaminathan, "Implementation of power transmission lines for field programmable gate arrays for managing signal and power integrity," *International Symposium on EMC*, pp. 322-327, Aug. 2013.

- S. Telikepalli, S. K. Kim, **S. J. Park**, M. Swaminathan, Youkeun Han, "Managing signal and power integrity using power transmission lines and alternative signaling schemes," *IEEE Fourth Latin American Symposium on Circuits and Systems (LASCAS),* pp. 1-4, March 2013.

- **S. J. Park**, J. Y. Choi, M. Swaminathan, "Simultaneous switching noise analysis of reference voltage rails for pseudo differential interfaces," *IEEE 21st Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, pp. 47-50, Oct. 2012.

REFERENCES

[1]   G. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, no. 8, pp. 114-117, 1965.

[2]   M. Waldrop, "The chips are down for Moore's law," *Nature*, vol. 530, no. 7589, pp. 144-147, 2016.

[3]   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*, vol. 89, no. 5, pp. 602-633, 2001.

[4]   M. Swaminathan and K. J. Han, *Design and Modeling for 3D ICs and Interposers*. WORLD SCIENTIFIC, 2013.

[5]   G. Van der Plas et. al., "Verifying electrical/thermal/thermo-mechanical behavior of a 3D stack - Challenges and solutions," *IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1-4, Sept. 2010.

[6]   N. Aghaee, Z. Peng, and P. Eles, "An efficient temperature-gradient based burn-in technique for 3D stacked ICs," *Design Automation & Test in Europe Conference &. Exhibition (DATE)*, pp. 1-4, Mar. 2014.

[7]   S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," *Proceedings of Design Automation Conference*, vol. 64, pp. 338-342, June 2003.

[8]   S. J. Park, N. Natu, M. Swaminathan, B. Lee, S. M. Lee, W. H. Ryu, and K. S. Kim, "Timing analysis for thermally robust clock distribution network design for 3D ICs," *IEEE Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, pp. 69-72, Oct. 2013.

[9] I. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits", *IEEE Transactions on Circuits Systems I*, vol. 48, no. 7, pp. 876-884, July 2001.

[10] J. C. Ku and Y. Ismail, "On the scaling of temperature-dependent effects," *IEEE Transactions on CAD of Integrated Circuits and Systems*, vol. 26, no. 10, pp.1882-1888, Oct. 2007.

[11] E. G. Friedman, "Clock distribution networks in synchronous digital integrated circuits," *Proceeding of the IEEE*, vol. 89, no. 5, pp. 665-692, May 2001.

[12] K. Bharath et al., "Signal and power integrity co-simulation for multi-layered system on package module," *IEEE International Symposium on Electromagnetic Compatibility*, pp. 1-6, July 2007.

[13] J. Xie and M. Swaminathan, "Electrical-thermal co-simulation of 3D integrated systems with micro-fluidic cooling and Joule heating effects," *IEEE Transactions Components, Packaging and Manufacturing Technology (CPMT)*, vol. 1, no. 2, pp. 234-246, Feb. 2011.

[14] D. Sekar et. al., "A 3D-IC technology with integrated microchannel cooling," *IEEE International Interconnect Technology Conference (IITC)*, pp. 13-15, June 2008.

[15] L. Shang, L.-S. Peh, and N. K. Jha, "Dynamic voltage scaling with links for power optimization of interconnection networks," *International Symposium on High-Performance Computer Architecture (HPCA)*, pp. 91-102, Feb. 2003.

[16] M. Cho, S. Ahmed, and D. Z. Pan, "TACO: Temperature aware clock-tree optimization," *IEEE/ACM International Conference on Computer-Aided Design*

(ICCAD), pp. 582–587, Nov. 2005.

[17] K. Shakeri and J. Meindl, "Temperature variable supply voltage for power reduction," *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pp. 64-67, Apr. 2002.

[18] A. Calimera, E. Macii, and R. I. Bahar, "Temperature-Insensitive Synthesis Using Multi-Vt Libraries," *ACM/IEEE Great Lakes Symposium on VLSI*, pp. 5-10, May 2008.

[19] J. T. Kao, M. Miyazaki, and A. P. Chandrakasan, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, Nov. 2002.

[20] G. Ono, M. Miyazaki, K. Watanabe, and T. Kawahara, "An LSI system with locked in temperature insensitive state achieved by using body bias technique," *IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 632–635, May 2005.

[21] D. Wolpert and P. Ampadu, "Exploiting programmable temperature compensation devices to manage temperature-induced delay uncertainty," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 4, pp. 735-748, Apr. 2012.

[22] A. Chakraborty and K. Duraisami, "Dynamic thermal clock skew compensation using tunable delay buffers," *International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 162-167, June 2006.

[23] T. Ragheb, A. Ricketts, M. Mondal, S. Kirolos, G. M. Links, V. Narayanan, and Y. Massoud, "Design of thermally robust clock trees using dynamically adaptive

clock buffers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 2, pp. 374-383, June 2008.

[24] C. P. C. Park, J. P. John, K. Klein, J. Teplik, J. Caravella, J. Whitfield, K. Papworth, and S. C. S. Cheng, "Reversal of temperature dependence of integrated circuits operating at very low voltages," *International Electron Devices Meeting*, pp. 71-74, Dec. 1995.

[25] R. Kumar and V. Kursun, "Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 10, pp. 1078-1082, Oct. 2006.

[26] A. Bellaouar, A. Fridi, M. I. Elmasry, and K. Itoh, "Supply voltage scaling for temperature insensitive CMOS circuit operation," *IEEE Transaction on Circuits and Systems II: Analog Digital Signal Processing*, vol. 45, no. 3, pp. 415-417, Mar. 1998.

[27] M. D. Meehan and J. Purviance, *Yield and Reliability in Microwave Circuit and System Design. Boston*, MA: Artech, 1993.

[28] G. Taguchi and S. Konishi, *Orthogonal Arrays and Linear Graphs*. Dearborn, MI: American Supplier Inst. Press, 1987.

[29] Global optimization tool box user's guide in MATLAB, http://www.mathworks.com/help/pdf_doc/gads/gads_tb.pdf.

[30] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design. Norwood*, MA: Artech House, 2000.

[31] A. B. Kahng, M. Luo, S. Nath, "SI for free: machine learning of interconnect coupling delay and transition effects," *ACM/IEEE International Workshop on*

System Level Interconnect Prediction (SLIP), pp. 1-8, June 2015.

[32] W. T. Beyene, "Application of artificial neural networks to statistical analysis and nonlinear modeling of high-speed interconnect system," *IEEE Transactions Computer-Aided Design of ICs and Systems*, vol. 26, no. 1, pp. 166-176, Jan. 2007.

[33] N. Ambasana, G. Anand, B. Mutnury, and D. Gope, "Eye height/width prediction from S-parameters using learning-based models," *IEEE Transactions on Component, Packaging, and Manufacturing Technology*, vol. 6, no. 6, pp. 873-885, June 2016.

[34] S. J. Park, J. Y. Choi, and M Swaminathan, "Simultaneous switching noise analysis of reference voltage rails for pseudo differential interfaces," *IEEE Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, Oct. 2012.

[35] D. Kim, J.Kim, J. Cho, J. S. Pak, J. Kim, H. Lee, J. Lee, and K. Park, "Distributed multi TSV 3D clock distribution network in TSV-based 3D IC," *IEEE Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, pp. 87-90, Oct. 2011.

[36] D. Kim, J. Pak, H. Lee, J. Lee, K. Park, and J. Kim, "Vertical tree 3-dimensional TSV clock distribution network in 3D IC," *IEEE Electronic Components and Technology Conference (ECTC)*, pp. 1945-1950, June 2012.

[37] S. J. Park, N. Natu, and M. Swaminathan, "Analysis, design, and prototyping of temperature resilient clock distribution network for 3-D ICs," *IEEE Transactions on Component, Packaging, and Manufacturing Technology*, vol. 5,

no. 11, pp. 1669–1678, Oct. 2015.

[38] "45nm NCSU FreePDKTM," 2012. [Online]. Available: http://www.si2.org.

[39] "PTM (Predictive Technology Model)," 2012. [Online]. Available: http://ptm.asu.edu.

[40] ITRS, "ITRS Roadmap - Interconnect," 2011.

[41] J. Kim, J. S. Pak, J. Cho, E. Song, J. Cho, H. Kim, T. Song, J. Lee, H. Lee, H. Park, S. Yang, M.-S. Suh, K.-Y. Hyun, J. Kim, "High-frequency scalable electrical model and analysis of a through silicon via (TSV)," *IEEE Transactions on Component, Packaging, and Manufacturing Technology*, vol. 1, no. 2, pp. 181-195, Feb. 2011.

[42] F. W. Grover, *Inductance Calculations: Working Formulas and Tables*, Dover Publications, pp.31-35, 1946.

[43] Y. Liang and Y. Li, "Closed-form expressions for the resistance and the inductance of different profiles of through-silicon vias," *IEEE Electron Device Letters*, vol. 32, no. 3, pp.393-395, Feb. 2011.

[44] V. S. Pandit, W. H. Ryu, K. Pushparaj, S. Ramanujam, and F. Fattouh "Simulation and characterization of GHz on-chip power delivery network (PDN)," in DesignCon 2008, 2008.

[45] J. S. Pak, J. Kim, J. Cho, K. Kim, T. Song, S. Ahn, J. Lee, H. Lee, K. Park, and J. 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," *IEEE Transactions on Component, Packaging, and Manufacturing Technology*, vol. 1, no. 2, pp. 208-219, Feb. 2011.

[46]   J. L. Knighten et. al., "PDN design strategies : I. Ceramic SMT decoupling capacitors – What values should I choose?" IEEE EMC Society Newsletter, no. 207, pp. 46-53, 2005.

[47]   J. M. Rabaey, *Digital Integrated Circuits*, Prentice-Hall, 2003.

[48]   N. Spennagallo, L. Codecasa, D. D'Amore, and P. Maffezzoni, "Lumped electro-thermal model of on-chip interconnects," International Workshop on Thermal Investigation of ICs and Systems (THERMINIC), pp. 220-224, 2006.

[49]   M. Swaminathan, "Signal integrity," *IEEE Electromagnetic Compatibility Magazine*, vol. 2, no. 3, pp. 60-68, 2013.

[50]   R. Saleh and K. Arabi, "Novel decoupling capacitor designs for sub-90nm CMOS technology," *International Symposium on Quality Electronic Design*, pp. 266-271, Mar. 2006.

[51]   M. Popovich, A. V. Mezhiba, and E. G. Friedman, *Power Distribution Networks with On-Chip Decoupling Capacitors*, Boston, Springer, 2008.

[52]   S. J. Park and M. Swaminathan, "Temperature-aware power distribution network designs for 3D ICs and Systems*," IEEE Electronic Components and Technology Conference (ECTC)*, pp. 732-737, May 2015.

[53]   H. H. Chen, J. S. Neely, M. F. Wang, and G. Co, "On-chip decoupling capacitor optimization for noise and leakage reduction," *Symposium on Integrated Circuits and Systems Design (SBCCI),* pp. 251-255, Sept. 2003.

[54]   Y. Chen, H. Li, K. Roy, C.-K. Koh, "Gated decap: Gate leakage control of on-chip decoupling capacitors in scaled technologies," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 17, no. 12, pp. 1749-1752, Dec. 2009.

[55]   J. Gu, H. Eom, and C. H. Kim, "On-chip supply noise regulation using a low-power digital switched decoupling capacitor circuit," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 6, pp. 1765-1775, June 2009.

[56]   R. Harjani and C. H. Kim, "Design and Implementation of Active Decoupling Capacitor Circuits for Power Supply Regulation in Digital ICs," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 17, no. 2, pp. 292-301, Feb. 2009.

[57]   P. Zurcher et. al., "Integration of thin film MIM capacitors and resistors into copper metallization based RF-CMOS and Bi-CMOS technologies," *International Electron Devices Meeting*, pp. 153-156, Dec. 2000.

[58]   C. H. Ng, C.-S. Ho, S.-F. S. Chu, and S.-C. Sun, "MIM Capacitor Integration for Mixed-Signal/RF Applications," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, pp. 1399-1409, July 2005.

[59]   B. Dang et. al., "Three-Dimensional Chip Stack With Integrated Decoupling Capacitors and Thru-Si Via Interconnects," *IEEE Electron Device Letter*, vol. 31, no. 12, pp. 1461-1463, Dec. 2010.

[60]   R. Saleh and K. Arabi, "Layout of decoupling Capacitors in IP Blocks for 90-nm CMOS," *IEEE Transaction on Very Large Scale Integration Systems*, vol. 16, no. 11, pp. 1581-1588, Nov. 2008.

[61]   D. Stringfellow and J. Pedicone, "Decoupling capacitance estimation, implementation, and verification : A practical approach for deep submicron SoCs," Synopsys. Synopsys User Group, 2007.

[62]   P. Larsson, "Parasitic resistance in an MOS transistor used as on-chip

decoupling capacitance," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 4, pp. 574-576, Apr. 1997.

[63] J. Rius and M. Meijer, "A high-frequency nonquasi-static analytical model including gate leakage effects for on-chip decoupling capacitors," *IEEE Transactions on Advance Packaging*, vol. 29, no. 1, pp. 88-97, Feb. 2006.

[64] J. Rius and M. Meijer, "Analysis of the influence of substrate on the performance of on-chip MOS decoupling capacitors," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 2, pp. 484-494, Feb. 2009.

[65] T. Charania, A. Opal, and M. Sachdev, "Analysis and design of on-chip decoupling capacitors," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 21, no. 4, pp. 648-658, Apr. 2013.

[66] J. Kim, W. Lee, Y. Shim, S. Shim, K. Kim, J. S. Pak, and J. Kim, "Chip-package hierarchical power distribution network modeling and analysis based on a segmentation method," *IEEE Transactions on Advanced Packaging*, vol. 33, no. 3, pp. 647-659, Aug. 2010.

[67] U. Kang et. al., "8Gb 3D DDR3 DRAM using through-silicon-via technology," *IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, pp. 130–131, 2009.

[68] D. U. Lee et. al., "A 1.2V 8Gb 8-channel 128GB/s high-bandwidth memory (HBM) stacked DRAM with effective microbump I/O test methods using 29nm process and TSV," *IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 432–433, 2014.

[69] S. K. Kim, S. Telikepalli, S. J. Park, M. Swaminathan, D. Keezer, and Y. Han,

"Implementation of power transmission lines to field programmable gate array ICs for managing signal and power integrity," *IEEE International Symposium on Electromagnetic Compatibility*, pp. 322–327, 2013.

[70]　S. J. Park, N. Natu, and M. Swaminathan, "Design and early validation (using FPGA) of temperature resilient clock distribution networks for 3D ICs," *IEEE Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS),* pp. 127-130, Oct. 2014.

[71]　N. U. Natu, "Design and prototyping of temperature resilient clock distribution networks design," Master degree thesis, 2014.

[72]　M. D. Meehan and J. Purviance, *Yield and Reliability in Microwave Circuit and System Design*, Boston, MA: Artech, 1993.

[73]　G. Taguchi and S. Konishi, *Orthogonal Arrays and Linear Graphs*. Dearborn, MI: American Supplier Inst. Press, 1987.

[74]　Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Norwood, MA: Artech House, 2000.

[75]　W. T. Beyene, "Application of artificial neural networks to statistical analysis and nonlinear modeling of high-speed interconnect system," *IEEE Transactions on Computer-Aided Design of ICs and Systems*, vol. 26, no. 1, pp. 166-176, Jan. 2007.

[76]　J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," Proceedings of Advances in Neural Information Processing Systems (NIPS), pp. 2951-2959, 2012.

[77]　Tom Mitchell, *Machine Learning*. McGraw Hill, 1997.

[78] A. Gelman et al., *Bayesian Data Analysis*, CRC Press, 2008.

[79] Emile Contal, Cédric Malherbe, and Nicolas Vayatis, "Optimization for Gaussian processes via chaining," NIPS Workshop on Bayesian Optimization, 2015.

[80] R. P. Adams, "A tutorial on Bayesian optimization for machine learning," NCAP, 2014.

[81] S. J. Park, M. Swaminathan, "Application of machine-learning for optimization of 3-D integrated circuits and systems," *submitted to IEEE Transactions on Very Large Scale Integration (VLSI) Systems (in-publication),* 2016.

[82] W. B. Powell and P. I. Frazier, "Optimal learning" INFORMS Tutorials in Operations Research, 2012.

[83] K. Kawaguchi, L. P. Kaelbling, and T. Lozano-P´erez, "Bayesian optimization with exponential convergence," *Proceedings of Adv. in Neural Information Processing Systems (NIPS)*, 2015.

[84] Z. Wang, B. Zhou, and S. Jegelka, "Optimization as estimation with Gaussian processes in bandit settings," *International Conf. on Artificial and Statistics (AISTATS)*, 2016.

[85] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," arXiv:1012.2599, 2010.

[86] S. J. Park, H. Yu, M. Swaminathan, "Preliminary application of machine-learning techniques for thermal-electrical parameter optimization in 3-D IC," *Accepted from IEEE Signal and Power Integrity Conference (SIPI 2016),* 2016.

VITA

Sung Joo Park received B.S. and M.S. degrees in electronic engineering from Sogang University in Seoul, Korea, in 1998 and 2000, respectively. He is pursuing a Ph.D. degree at the Georgia Institute of Technology in Atlanta, Georgia. His research focus is on high-speed interfaces and signal, power, and thermal integrity in designs of 3-D ICs and systems.

Since 2000, he has been a senior engineer in the DRAM Development and Technology Department at Samsung Electronics Co., Ltd. where he worked on the design and development of memory systems. He has 14 issued US patents on memory modules and systems and has several patents pending. Mr. Park is a senior member of IEEE. He is a member of the JEDEC Solid State Technology Association where he has developed industry standards since 2006 and had served as (vice) chair of a committee and a task group. He received the JEDEC Chairman's Award in 2010, the Samsung Fellowship in 2011, and the best student paper award at the IEEE EMC symposium in 2013.