

國立交通大學

電信工程學系

碩士論文

考慮製程變異及擁有對功率模型具高度相  
容性的電熱模擬器

An Electro-Thermal Simulator Considering  
Process Variations with High Compatibility of  
Power Model

研究生：張懷中

指導教授：李育民 教授

中華民國九十八年六月

考慮製程變異及擁有對功率模型具高度相容性的電熱模擬器  
An Electro-Thermal Simulator Considering Process Variations with High  
Compatibility of Power Model

研究生：張懷中

Student : Huai-Chung Chang

指導教授：李育民

Advisor : Yu-Min Lee

國立交通大學

電信工程學系

碩士論文



Submitted to Department of Communication Engineering  
College of Electrical and Computer Engineering  
National Chiao Tung University  
in partial Fulfillment of the Requirements  
for the Degree of  
Master  
in  
Communication Engineering

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月


# 考慮製程變異及擁有對功率模型具高度相容性的電熱模擬器

學生：張懷中

指導教授：李育民 博士

國立交通大學電信工程學系碩士班

## 摘 要



本篇論文中提出一個統計型的電熱模擬器，此模擬器考慮了漏電流、晶片與晶片間的製程變異和一個晶片內具有空間相關的製程變異。利用卡洛展開(Karhunen-Loève expansion)，吾人可以將一個具空間相關的製程變異參數，轉換成一組不具相關性的隨機變數做表示，接著在此不具空間相關性的隨機變數，以及代表晶片與晶片間製程變異的隨機變數所共同組成的隨機空間中，使用史摩亞克稀疏網格方法(Smolyak sparse grid method)在此隨機空間中去取樣以求解統計型熱傳方程式。接著透過電熱耦合演算法，可以在每一個取樣點得到一個晶片上的熱分佈。這些計算所得到的熱分佈，會被用來內插在一個晶片上的統計熱分佈，而一個統計上的熱分佈結果可以透過機率的運算所萃取出來。

本篇論文提出的統計型電熱模擬器的準確度，吾人利用蒙地卡羅分析(Monte Carlo analysis)做為比較，而此分析器的效率，是透過蒙地卡羅分析達到同一分析精確度的執行時間做為比較基準。根據實驗結果，本統計型電熱模擬器可以達到比蒙地卡羅分析快一個數量級的速度，且其結果在一個晶片上的溫度期望值最大的誤差在0.36%之內，溫度標準差的誤差小於1.88%。除此之外，本篇論文的電熱模擬器具有對不同功率模型的高度相容性，這個特性對於快速演進的科技是非常重要的。

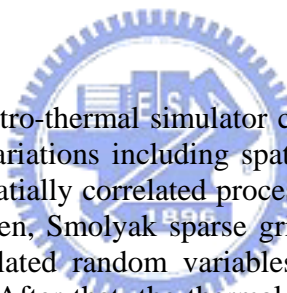
# An Electro-Thermal Simulator Considering Process Variations with High Compatibility of Power Model

Student: Huai-Chung Chang

Advisor: Dr. Yu-Min Lee

Department of Communication Engineering  
National Chiao Tung University

## ABSTRACT



In this paper, a statistical electro-thermal simulator considering leakage power, inter-die process variations, and intra-die process variations including spatial correlation is developed. With applying Karhunen-Loève expansion, the spatially correlated process parameters can be transformed to a set of uncorrelated random variables. Then, Smolyak sparse grid method is applied to sample the random space expanded by these uncorrelated random variables and inter-die random variables to tackle stochastic heat transfer equations. After that, the thermal profile at each sampling point is built by a constructed electro-thermal coupling algorithm. These calculated thermal profiles are integrated to interpolate the stochastic temperature profile over a chip. Finally, the statistical temperature profile can be extracted.

The accuracy and efficiency of the presented statistical electro-thermal simulator are demonstrated by comparing with the Monte Carlo analysis. Experimental results indicate that the developed simulator is orders of magnitude faster than that of the Monte Carlo analysis under the same accuracy level. The maximum error is less than 0.36% and 1.88% in mean and standard deviation of temperature profiles, respectively. The proposed simulator is also highly compatible with different power models and spatial correlation functions. This characteristic is important in such fast innovative technology.

# 誌 謝

本篇論文得以順利完成，首先感謝我的指導教授李育民博士，老師在研究還有課業上給予的指導，是我在碩士生涯中最大的收穫。對於未來的路，我不會徬徨，因為在老師的指導下，我學到一個研究生該有的學習態度，這是我學習生涯中得到最大的寶藏。

研究的路上，感謝在我碩士階段，時常找我運動、程式給予指導的博士班學長柏毅，更感謝一直和我切磋砥礪的博士班學長培育，你放浪不羈的形象與個性爽直的言談，卻是我研究生涯最鮮明的扉頁。還有感謝一群曾與我共同在實驗室打拼的夥伴，學長至鴻、國富、志康、炳勳、哲宇、佳鴻、焯基、庚達、斯安、同窗宗祐、學弟麒文、志昇、學妹書含、巧翎、亭蓉，你們給予的關心和幫助，豐富了我碩士的生活。

我將最誠摯的感謝，獻予支持我的家人、女友琇琪及其父母親，因為有你們的相伴與關懷，我才可以在人生的路口，找到自己的路，論文中的字句彷彿在你們面前跳一支感謝的舞，用這些年的光陰譜成，謝謝你們。

同時我將這份喜悅與快樂獻給所有關心我的人，並希望閱讀這份論文的讀者，能給予敝人指教，謝謝。

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Overview of Our Statistical Electro-Thermal Simulator . . . . .	4
1.3	Our Contributions . . . . .	4
1.4	Organization of the Thesis . . . . .	5
<b>2</b>	<b>Preliminaries and Problem Formulation</b>	<b>6</b>
2.1	The Importance of Electro-Thermal Coupling in Deterministic and Statistical Thermal Simulations . . . . .	6
2.2	Statistically Cell-based Leakage Current Modeling . . . . .	8
2.2.1	Presented Leakage Current Models vs. Previous Works . . . . .	9
2.3	Smolyak Sparse Grid Formula . . . . .	11
2.4	Problem Formulation . . . . .	15
<b>3</b>	<b>Statistical Electro-Thermal Framework</b>	<b>17</b>
3.1	Statistical Electro-Thermal Flow . . . . .	17
3.2	Parameter Modeling . . . . .	19
3.3	Smolyak Sparse Grid Interpolation Based Simulation . . . . .	20
3.4	Complexity Analysis . . . . .	22
<b>4</b>	<b>Experimental Results</b>	<b>25</b>
4.1	Accuracy and Efficiency . . . . .	25
4.2	Without vs. With Including the Effect of Electro-Thermal Coupling . . . . .	30
<b>5</b>	<b>Application–Thermal Yield</b>	<b>32</b>
5.1	Thermal Yield of Circuit . . . . .	32
5.2	Statistical Thermal Yield Analysis Problem . . . . .	33
<b>6</b>	<b>Conclusions</b>	<b>35</b>

# List of Figures

1.1	Temperature dependency and process variations of subthreshold leakage current in one NAND gate. . . . .	2
1.2	Temperature dependency and process variations of gate tunneling leakage current in one NAND gate. . . . .	2
1.3	Leakage current and frequency variations [1]. . . . .	3
2.1	Total power consumption of an NAND gate at different operating temperatures. This cell is assumed to be surrounded with a thermal isolation system, and the power is only dissipated through the package. . . . .	7
2.2	Clenshaw-Curtis sampling points of Smolyak formula and full tensor product of a two-dimensional parameter space ( $d=2$ ). (a) Smolyak sparse grids with maximum level $q=3$ . (b) Full tensor product of $q=3$ . (c) Smolyak sparse grids with maximum level $q=5$ . (d) Full tensor product of $q=5$ . . . . .	13
2.3	Compact thermal model of physical design. . . . .	15
3.1	Sparse grid based statistical electro-thermal simulation flowchart. . . . .	17
3.2	Electro-thermal-coupling algorithm . . . . .	21
3.3	Leakage-power-updating algorithm . . . . .	22
3.4	Simulating algorithm of the proposed statistical electro-thermal simulator. . . . .	23
4.1	(a) The floorplan of test chip. (b) The geometry setting of test chip. . . . .	26
4.2	The temperature profile at the top surface of the die. (a) The mean temperature distribution without considering electro-thermal coupling. (b) The mean temperature distribution with considering electro-thermal coupling. . . . .	28
4.3	The temperature profile at the top surface of the die. (a) The spatial standard deviations without considering electro-thermal coupling. (b) The spatial standard deviations with considering electro-thermal coupling. . . . .	28
4.4	Distribution of the temperature using Monte Carlo (MC) simulation, with and without electro-thermal coupling, and the proposed method at the location of the hottest mean temperature. (a) Probability density function (PDF). (b) Cumulative distribution function (CDF). . . . .	29
4.5	PDFs and CDFs of the total leakage power using MC simulation with and without considering electro-thermal coupling. . . . .	31
5.1	PDFs of the temperature at two locations of the chip for indicating which one is more critical on the chip. . . . .	33
5.2	Probability of exceeding the reference temperature $\overline{\Phi}_{\log}(T_{ref}) = 90^{\circ}C$ from statistical thermal simulator (a) with considering electro-thermal coupling. (b) without considering electro-thermal coupling. . . . .	34

# List of Tables

2.1	Error comparison of $I_{sub}$ and $I_{gate}$ with HSPICE simulation results for an NAND gate. . . . .	10
2.2	The number of sampling points use Smolyak formula and full tensor product formula in d-dimensional sampling space with $q=3$ . . . . .	14
4.1	Accuracy and efficiency compared with the Monte Carlo method. . . . .	27





# Chapter 1

## Introduction

### 1.1 Motivation

As technology is scaling down continuously and power density is rapidly increasing, power dissipation and thermal management have become important issues of VLSI design. Furthermore, temperature and thermal gradients have significant influence on IC performance, reliability, and the cost of cooling and package system. Because the leakage power has become the major contributor of total power in the modern technology, it is necessary to estimate and model the leakage power accurately and efficiently. However, leakage power is exponentially correlated with process parameters and temperature shown in Fig. 1.1 and Fig. 1.2, so that process variations and thermal impacts need to be taken into concern cautiously. The authors in [1] indicated that 30% intra-die process variations can lead to 20 times of leakage power causing the drastic fluctuations of temperature distributions as shown in Fig. 1.3.

Moreover, because of the lithography and chemical mechanical polishing defects, physical parameters are varied with spatial positions which the closer gates may have more likelihood to have similar physical characteristics. Without considering spatial correlations of intra-die process variations, the standard deviation of temperature distribution can be 3 to 4 times lower than the results with considering spatial correlations [2].

Using the deterministic thermal analysis to obtain one deterministic temperature-dependent leakage power simulator has been proposed in [3,4]. However, as considering process variations, all analysis problems need to be transformed to random process problems, and a statistical simulator is needed. In power analysis, several works have successfully quantified the process variations into leakage power [5–7]. Nevertheless, none of them consider the electro-thermal

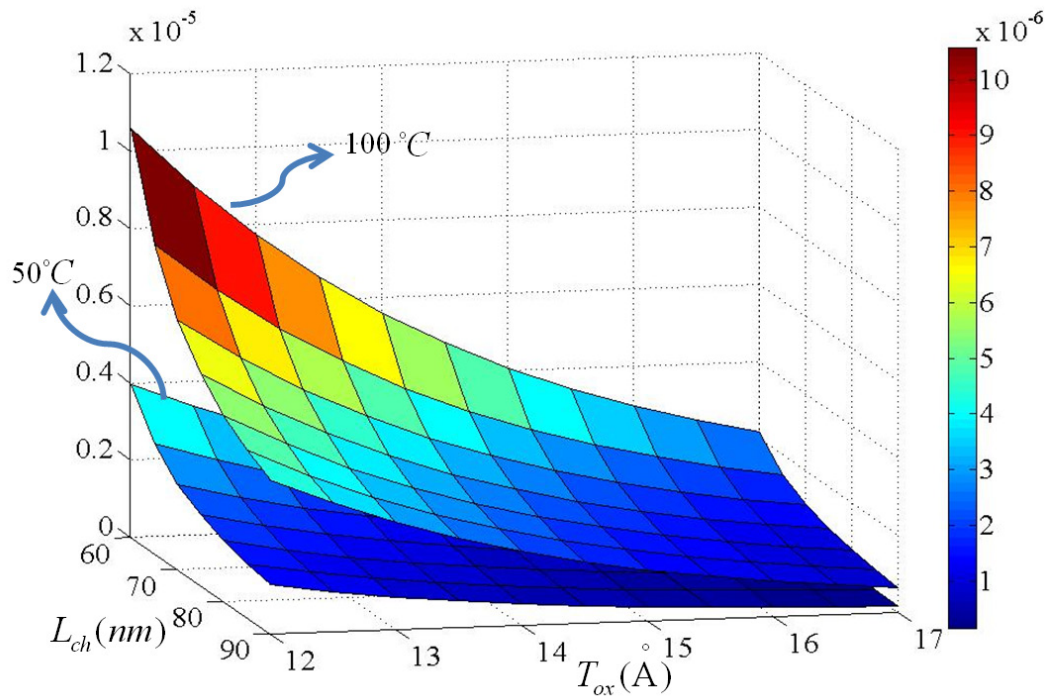


Fig. 1.1: Temperature dependency and process variations of subthreshold leakage current in one NAND gate.

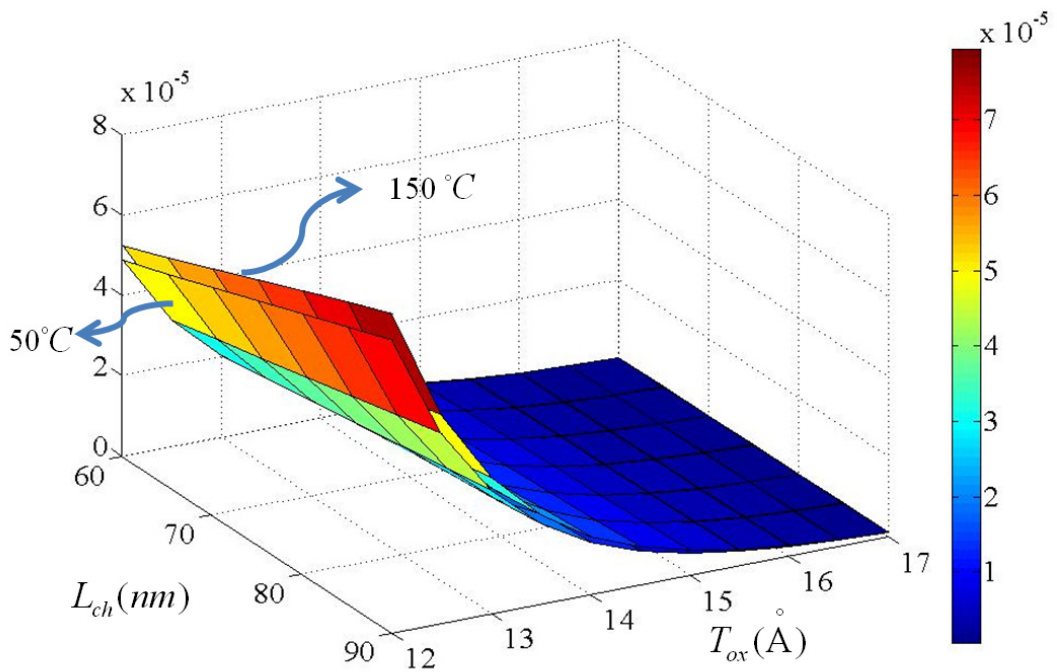


Fig. 1.2: Temperature dependency and process variations of gate tunneling leakage current in one NAND gate.

feedback in statistical power analysis.

In thermal analysis, existing statistical thermal simulators [2, 8] considering process variations and spatial correlations have some limitations in their methodologies. Authors in [2] didn't take the electro-thermal coupling into account. An architectural-level simulator proposed in [8] needs to fit the power model for each grid every time as the design changes, and this limits its usage after the floorplanning stage. Moreover, both two have limitations in the forms of power models. The power projection algorithm in [2] has the limitation of power model form. Because of using the log-normal assumptions in each analysis step of [8], there is a limitation of power model form. Because the scaling down technology will lead more complicated power model forms to enhance the accuracy, it is urgent to develop a statistical thermal simulator which has the high capability of adopting different and complicated power model forms for any technology generations.

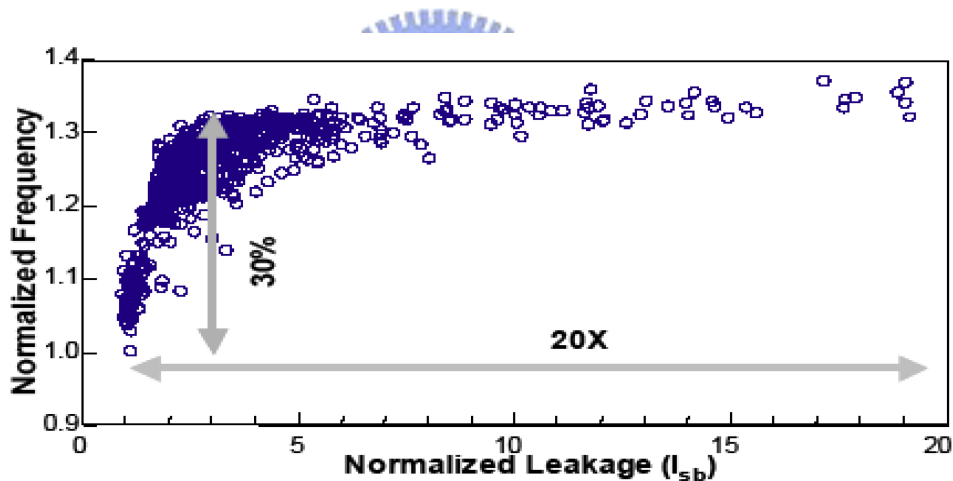


Fig. 1.3: Leakage current and frequency variations [1].

Monte Carlo method is the most popular method to obtain statistical solution of a statistical problem. Besides, it can be implemented to solve statistical thermal problem using any power model forms, because each sampling knot can make the statistical thermal problem become a deterministic thermal problem which is related to the power value rather than power model forms. Although the concept and implementation of Monte Carlo method are straightforward, its convergence rate is very slow in a large number of random variables. An alternative way to efficiently obtain statistical solution of a statistical problem is the statistical collocation method.

By applying sparse grids in the high level statistical collocation method can dramatically reduce the calculating complexity comparing with that of Monte Carlo method and maintain the advantage of applying Monte Carlo method in statistical thermal problem.

## 1.2 Overview of Our Statistical Electro-Thermal Simulator

In this work, we develop a statistical electro-thermal simulator that considers the effects of spatial correlation under intra-die process variations and inter-die variations. Because the sparse grid collocation technique, a Monte-Carlo-like method, is utilized, the proposed simulator can handle any power model forms and spatial covariance functions. Hence, an extremely accurate statistical cell-based leakage power model form is developed, so the proposed simulator can provide more accurate results than the architectural-level simulator. Moreover, as the developed electro-thermal simulator is used for thermal-driven floorplan/placement problems, it can be rapidly adopted without reconstructing the power model since we used a cell-based power model rather than a grid-based power model [8].

Firstly, the Karhunen-Loève (KL) expansion is utilized to transform the spatially fluctuating physical process parameters to a set of uncorrelated random variables. Then, the Smolyak sparse grid method [9] is applied to sample the random space expanded by these uncorrelated random variables added with random variables of inter-die variations. Given the initial temperature profile of a full-chip, for each sampling point, the power profile over a chip can be obtained by the proposed power model forms of cells. After using an existent deterministic thermal simulator to update the temperature profile, the power profile over a chip is also updated. The above temperature-power updating procedure is repeatedly until it is convergent. Finally, those calculated thermal profiles under all sampling points are utilized to interpolate the stochastic temperature profile over a chip, and the statistical temperature profile can be extracted.

## 1.3 Our Contributions

Our major contributions are

1. To the authors' best knowledge, this work is the first gate-level statistical electro-thermal simulator including the effect of intra-die variations with spatial correlations and inter-die

variations. This simulator also shows the high compatibility to handle any complicated power model forms and spatial correlation functions.

2. The developed statistical electro-thermal simulator can accurately and efficiently provide the mean temperature distribution profile and the spatial standard deviation profile of temperature distribution. The circuit designers can utilize the above information to take effectively strategies for fighting against thermal failures with considering process variations. Experimental results reveal that ignoring electro-thermal coupling in statistical thermal simulations can mislead circuit designers to an unreliable design direction.
3. A thermal yield analysis problem is formulated. By using statistical thermal profile from statistical thermal simulators, the thermal yield of circuit can be obtained. This information is useful for designers to avoid the thermal runaway and predict yield of the chip.

## **1.4 Organization of the Thesis**

The rest of the thesis is organized as follows. In chapter 2, the importance of electro-thermal coupling and background are illustrated. Moreover, the problem of statistical thermal simulation is formulated. Then, the statistical electro-thermal framework is presented in chapter 3. After that, the experimental results are given in chapter 4, and an application of thermal yield is investigated in chapter 5. Finally, this work is concluded in chapter 6.

# Chapter 2

## Preliminaries and Problem Formulation

In this chapter, the importance of electro-thermal coupling in both deterministic and statistical thermal simulator is illustrated in section 2.1. Then, in section 2.2, a survey of statistical leakage current models is introduced and novel leakage current models are presented in subsection 2.2.1. The background of Smolyak sparse grid formula is investigated in section 2.3. The end of this chapter is problem formulation.

### 2.1 The Importance of Electro-Thermal Coupling in Deterministic and Statistical Thermal Simulations

A simple schematic example shown in Fig. 2.1 is used to highlight the importance of electro-thermal coupling and the impact of process variations. Given a single NAND gate surrounded with a thermal isolation system and the only power dissipation path is through the package, its power consumption with/without considering process variation is shown in Fig. 2.1. Although the temperature of a cell depends on its neighbor cells in a real chip, this schema still works for indicating the importance of electro-thermal coupling in statistical and deterministic thermal simulations.

Given an initial temperature, the power consumption of an NAND gate can be calculated. Based on the zeroth law of thermodynamics [10], to achieve the equilibrium of generating power and power dissipated by package, the surplus power that cannot be dissipated by package must be transformed to heat and stored in this system. Hence, the system temperature is increased. On the contrary, as the capacity of power dissipated by package is larger than the produced power, the system temperature decreases. Because the leakage power is highly dependent on

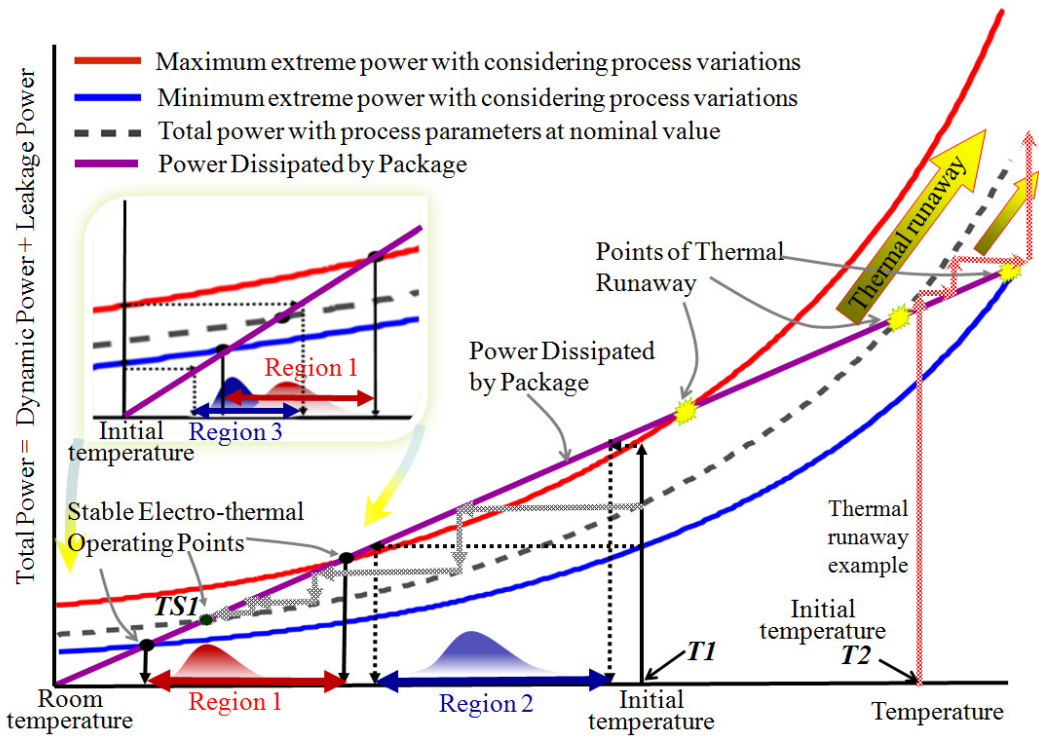


Fig. 2.1: Total power consumption of an NAND gate at different operating temperatures. This cell is assumed to be surrounded with a thermal isolation system, and the power is only dissipated through the package.

temperature, the total power needs to be adjusted with the updated temperature, and this procedure is called electro-thermal coupling. The above procedure is recursively performed until the system reaches the equilibrium of power production and dissipation, and the temperature is converged. After that, the stable operating temperature of this cell is gotten. If the system cannot reach the thermal equilibrium, the system is thermal runaway and is under high risk of system melted down. For example, in Fig. 2.1, the dash line indicates the power consumption of an NAND gate operating at different temperatures with process parameters being nominal values. The straight line passing through the room temperature indicates the maximum power that can be dissipated by the package at each operating temperature. Given an initial temperature  $T1$ , the stable operating temperature is  $TS1$  after performing the electro-thermal coupling. On the other hand, if the initial temperature is  $T2$ , it will cause the thermal runaway.

However, with considering process variations, the equilibrium temperature can not be represented as a deterministic form. For example, in Fig. 2.1, the top curve is the maximum extreme power consumption of an NAND gate operating at different temperatures with considering process variations, and the bottom curve is the minimum extreme power consumption of an

NAND gate operating at different temperatures with considering process variations. As shown in Fig. 2.1, given an initial temperature  $T_I$ , the equilibrium temperature distribution falls into Region 1 *with* considering the electro-thermal coupling. However, the final temperature distribution falls into Region 2 *without* considering electro-thermal coupling. Given a different initial temperature such as the room temperature shown in the sub-plot of Fig. 2.1, the final temperature distribution falls into Region 3 *without* considering electro-thermal coupling. However, the equilibrium temperature distribution still falls into Region 1 *with* considering electro-thermal coupling.

The uncertainty of final temperature confidential region and the drastic error between Region 2/Region 3 and Region 1 show that it is necessary to consider electro-thermal coupling while performing statistical thermal simulation. Similarly, statistical power analysis should also take electro-thermal coupling into account.

## 2.2 Statistically Cell-based Leakage Current Modeling

When the oxide thickness of a device is reduced, the probability of electrons tunneling through oxide thickness is getting higher. This results in the gate tunneling leakage current which is related of oxide thickness  $T_{ox}$  and gate area referring to channel length  $L_{ch}$ . Because the number of electrons tunneling through the barrier which is influencing the tunneling probability is dependent on temperature [11], we also take temperature  $T$  into our leakage current model. As the device turns into "off" state ( $V_{gs} < V_{th}$ ), the minority carriers diffusing through the channel induce the current flowing from the drain to the source of a transistor. This is known as subthreshold leakage current.

Many compact leakage current models have been developed in [2–6, 8, 12]. However, none of leakage power models proposed in [2–6] took both temperature and process variation effects into account, their accuracy degrades as the technology scales down. For the authors' best knowledge, only [8, 12] proposed the leakage current models considering both effects. Nevertheless, the leakage current model in [12] was based on 90nm technology. Hence, as the technology advances, its accuracy is deteriorated. The authors in [8] developed a grid-based leakage power model in the architectural level. Each fitted form was used to coarsely approxi-



mate the total leakage current in each grid, and this limits its use after the floorplanning stage. Moreover, the grid-based leakage power model will be transformed into one nonlinear curve fitting problem as obtaining the coefficients of its model. Authors decomposed the nonlinear problem into several linear problems to acquire the coefficients, but this method cannot guarantee the solutions located into the global optimal region.

The leakage current of each cell depends on input patterns and is highly correlated with process parameters and operating temperatures. Hence, we apply different input patterns via varying physical process parameters and operating temperatures for each cell by using HSPICE and the design kit from industry to generate the fitting data. Then, using the least square fitting method, the coefficients of different average leakage current models such as the average subthreshold leakage ( $I_{sub}$ ) and the average gate tunneling leakage ( $I_{gate}$ ) can be obtained.

Since  $I_{sub}$  is the off-state leakage mechanism, and  $I_{gate}$  occurs in both on and off states of transistor [13], the leakage power of a cell can be represented as

$$P_{Leak} = V_{dd} \times (I_{gate} + (1 - Sw) I_{sub}), \quad (2.1)$$

where

$$I_{gate} = a_0 \cdot \exp(f_{gate}(T_{ox}, L_{ch}, T)), \quad (2.2)$$

$$I_{sub} = b_0 \cdot \exp(f_{sub}(T_{ox}, L_{ch}, T)). \quad (2.3)$$

Here,  $a_0$  and  $b_0$  are fitting constants,  $L_{ch}$  and  $T_{ox}$  are the channel length and oxide thickness, respectively.  $T$  is the operating temperature which may be updated every thermal loop,  $Sw$  is the switching activity,  $V_{dd}$  is the supply voltage, and  $f_{gate}$  and  $f_{sub}$  are specific fitting forms.

### 2.2.1 Presented Leakage Current Models vs. Previous Works

In this subsection, a novel cell-based leakage power model considering the process variations and temperature dependence is presented. Then, the comparison with latest works is shown by experimental results presented in Table 2.1.

Owing to the property of Smolyak sparse grid collocation method, any leakage current forms can be adopted in the proposed electro-thermal simulator. The presented leakage current forms

Table 2.1: Error comparison of  $I_{sub}$  and  $I_{gate}$  with HSPICE simulation results for an NAND gate.

$f_{gate}$		Max. Error	Avg. Error	Error > 3%
Without temperature	$T_{ox}, T_{ox}^2, L_{ch}, L_{ch}^2$ [5]	6.48%	2.70%	4.37%
With temperature	$L_{ch}, T, T_{ox}$	3.20%	0.97%	0.35%
	$\dagger L_{ch}, T, T_{ox}, T_{ox}^2$	1.55%	0.29%	0.00%
$f_{sub}$		Max. Error	Avg. Error	Error > 3%
Without temperature	$L_{ch}, L_{ch}^2, T_{ox}^{-1}, T_{ox}^2$ [5]	347.32%	70.65%	98.27%
	$L_{ch}, L_{ch}^2, T_{ox}^{-1}, T_{ox}, T_{ox}^2, T_{ox}/L_{ch}, L_{ch}/T_{ox}, T_{ox} \times L_{ch}$ [6]	314.13%	70.52%	100.00%
With temperature	$L_{ch}, T, T_{ox}$ [12]	32.23%	8.73%	76.62%
	$(L, T_{ox}, T)$ are fully expanded to 2nd order $\implies$ $L_{ch}, L_{ch}^2, T_{ox}, T_{ox}^2, T, T^2, L_{ch} \times T_{ox}, L_{ch} \times T, T_{ox} \times T$	10.31%	1.53%	8.47%
	$\dagger (L, T_{ox}, T)$ are fully expanded to 3rd order $\implies$ $L, L^2, T_{ox}, T_{ox}^2, T, T^2, L \times T_{ox}, L \times T, T_{ox} \times T, L^3, T_{ox}^3, T^3,$ $L^2 \times T_{ox}, L^2 \times T, T_{ox}^2 \times L, T_{ox}^2 \times T, T^2 \times T_{ox}, T^2 \times L$	1.31%	0.19%	0.00%

$\dagger$  The adoptive forms of  $f_{gate}$  and  $f_{sub}$  in this paper.

are based on equations (2.2) and (2.3) of

$$f_{gate}(T_{ox}, L_{ch}, T) = (a_1 \cdot L_{ch} + a_2 \cdot T + a_3 \cdot T_{ox} + a_4 \cdot T_{ox}^2),$$

$$f_{sub}(T_{ox}, L_{ch}, T) = (b_1 \cdot L_{ch} + b_2 \cdot T_{ox} + b_3 \cdot T + b_4 \cdot L_{ch} \cdot T_{ox} + b_5 \cdot T \cdot T_{ox} + b_6 \cdot L_{ch} \cdot T + b_7 \cdot L_{ch}^2 + b_8 \cdot T_{ox}^2 + b_9 \cdot T^2 + b_{10} \cdot L_{ch} \cdot T_{ox}^2 + b_{11} \cdot L_{ch} \cdot T^2 + b_{12} \cdot T \cdot T_{ox}^2 + b_{13} \cdot T \cdot L_{ch}^2 + b_{14} \cdot T_{ox} \cdot L_{ch}^2 + b_{15} \cdot T_{ox} \cdot T^2 + b_{16} \cdot T_{ox} \cdot T \cdot L_{ch} + b_{17} \cdot L_{ch}^3 + b_{18} \cdot T_{ox}^3 + b_{19} \cdot T^3),$$

where  $a_i$ 's and  $b_i$ 's are fitting constants. These forms gain the maximum error within 1.55%, and the average error within 0.5% for all cells built in leakage power cell library for this work.

Different fitting forms of equations (2.2) and (2.3) with an NAND gate under 65nm technology are shown in Table 2.1. As shown in Table 2.1, different components in equations (2.2) and (2.3) can lead to different errors compared with the simulation results from HSPICE. We do not compare the power form of [8] here, because the models compared in Table 2.1 are cell-based models and modeling the different combination of leakage current individually for the higher accuracy rather than a grid based total leakage model in [8]. These drastic errors in [5, 6, 12] are because of the ignorance of either temperature or developing technology. Compared with other forms [5, 6, 12], the adoptive forms gain the high accuracy which the maximum error is within 1.31% and 1.55% in subthreshold and gate tunneling leakage current, respectively. This table also shows that it is necessary to take temperature into leakage current model,

and it is importance to having the advantage of handling any power models in power or thermal simulator.

## 2.3 Smolyak Sparse Grid Formula

The idea of interpolation method is to construct a polynomial by using several known values of a desired function to approximate the desired function. The one-dimensional and level  $i_1$ <sup>1</sup> approximation applied to the function  $T$  is denoted as  $Q^{i_1}(T)$ . Here, the interpolation method based on Lagrange polynomials is briefly recalled. Assume that we want to approximate a one-dimension function  $T(\xi) : [-1, 1]^{d=1} \rightarrow \mathbb{R}$  by using a set of sampling points  $\{\xi_1^{i_1}, \dots, \xi_{m_{i_1}}^{i_1}\} \subset [-1, 1]$  of the variable  $\xi$ .  $m_{i_1}$  is the needed number of sampling points of the variable  $\xi$  for interpolating. Then the interpolated function by using the Lagrange interpolation can be written as

$$Q^{i_1}(T)(\xi) = \sum_{j=1}^{m_{i_1}} T(\xi_j^{i_1}) a_j^{i_1}(\xi) \quad (2.4)$$

where  $i_1 \in N$  and it denotes the highest level of the interpolating polynomial in the 1st-direction,  $a_j^{i_1} \in C([-1, 1])$  are the Lagrange polynomial of degree  $i$ ,  $a_j^{i_1}(\xi) = \prod_{\substack{k=1 \\ k \neq j}}^{m_{i_1}} \frac{(\xi - \xi_k^{i_1})}{(\xi_j^{i_1} - \xi_k^{i_1})}$ .

For the multivariate case, we would like to approximate a  $d$ -dimensional function  $T$ . Conventionally, the full tensor product interpolation formula  $Q_d(T) = (Q^{i_1} \otimes \dots \otimes Q^{i_j} \otimes \dots \otimes Q^{i_d})(T)$  can be used to approximate it by full grid collocation. Here,  $\otimes$  is the tensor product operator, and  $i_j$  is the highest level of the interpolating polynomial in the  $j$ th-direction. For example,  $(a\xi_1 + b\xi_1^2) \otimes (c\xi_2 + d\xi_2^2)$  is equal to  $(ac\xi_1\xi_2 + ad\xi_1\xi_2^2 + bc\xi_1^2\xi_2 + bd\xi_1^2\xi_2^2)$  where  $a, b, c,$  and  $d$  are the coefficients. The full tensor product formula needs  $\prod_{j=1}^d m_{i_j}$  counts of total sampling points. Here,  $m_{i_j}$  is the number of sampling points in the  $j$ th-direction. Using Lagrange polynomial for interpolating as an example here, the full tensor product interpolation formula is

$$(Q^{i_1} \otimes \dots \otimes Q^{i_d})(T) = \sum_{j_1=1}^{m_{i_1}} \dots \sum_{j_d=1}^{m_{i_d}} T(\xi_{j_1}^{i_1}, \dots, \xi_{j_d}^{i_d}) \cdot (a_{j_1}^{i_1} \otimes \dots \otimes a_{j_d}^{i_d}), \quad (2.5)$$

However, using the full tensor product to approximate a multivariate function is inefficient especially as the dimension increases. Smolyak [9] proposed a sparse grid stochastic collocation

<sup>1</sup>In this work, the number of sampling points,  $m_i$ , in level  $i$  is defined as  $m_1 = 1$  and  $m_i = 2^{i-1} + 1$  for  $i > 1$ , because the chosen sampling points are nested.

method to reduce the number of sampling points from full grid collocation, and this method was investigated by [14]. With  $Q^0 = 0$  and  $i \in N_+$ , the authors in [14] denoted  $|\mathbf{i}| = i_1 + \dots + i_d$  and defined the difference between two interpolating polynomials of level  $i$  and  $i - 1$  as

$$\Delta^i = Q^i - Q^{i-1}. \quad (2.6)$$

Then the Smolyak formula can be given as

$$A(q, d)(T) = \sum_{q-d+1 \leq |\mathbf{i}| \leq q} (\Delta^{i_1} \otimes \dots \otimes \Delta^{i_d})(T). \quad (2.7)$$

Equivalently, formula (2.7) can be written as [14]

$$A(q, d)(T) = \sum_{q-d+1 \leq |\mathbf{i}| \leq q} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} (Q^{i_1} \otimes \dots \otimes Q^{i_d})(T). \quad (2.8)$$

where  $A(q, d)(T)$  is the approximated polynomial,  $q$  denotes the level of desired solution, and  $d$  is the dimension of functional space.

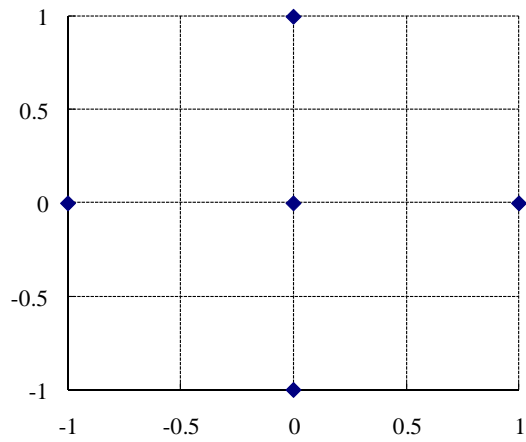
For a function  $u \in C^r$ , the error of interpolating on a Smolyak sparse grid is guaranteed to satisfy  $O(m^{-r}(\log(m))^{(d-1)(r-1)})$ , where  $m$  is the total number of sampling points [15].

According to formulas (2.7) and (2.8), we only need to know the function values on the sparse grid rather than the full grid [16]. The set of sparse sampling points in (2.7) is derived as

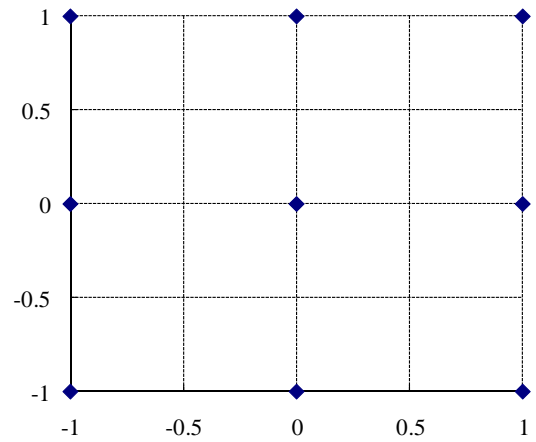
$$H(q, d) = \bigcup_{q-d+1 \leq |\mathbf{i}| \leq q} (\vartheta^{i_1} \times \dots \times \vartheta^{i_j} \times \dots \times \vartheta^{i_d}), \quad (2.9)$$

where  $\vartheta^{i_j}$  denotes the vector of sampling points in the  $j$ th-direction. The number of points from Smolyak sparse grid formula increases as  $O\left(\frac{d^{q-d}}{(q-d)!}\right)$  which is less than that from full grid collocation.

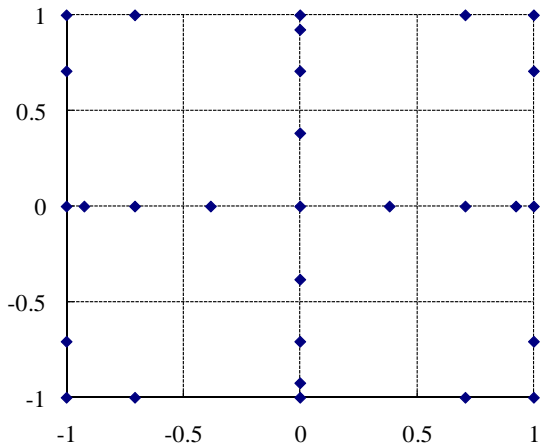
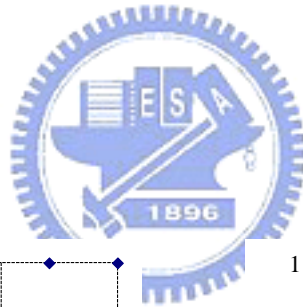
A simple example is presented for clearer specifying Smolyak sparse grid interpolation. With the dimension  $d=2$  and the Smolyak sparse grid formula of  $q=d+1$  using the sampling value in one random variable of  $\{a, b, c\}$  in (2.8) and according to the condition  $q - d + 1 \leq |\mathbf{i}| \leq q$ , we can obtain  $|\mathbf{i}| = 2 \Rightarrow i_1 = 1, i_2 = 1$  and  $|\mathbf{i}| = 3 \Rightarrow i_1 = 1, i_2 = 2$  or  $i_1 = 2, i_2 = 1$ , where  $\vartheta^1 = \{a\}$ ,  $\vartheta^2 = \{a, b, c\}$ . The sampling points of the Smolyak sparse grid can be obtained by



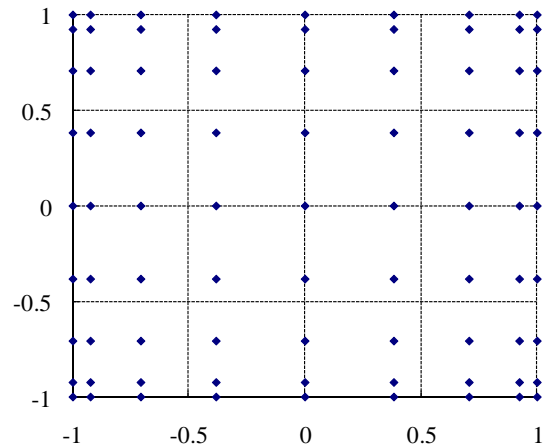
(a)



(b)



(c)



(d)

Fig. 2.2: Clenshaw-Curtis sampling points of Smolyak formula and full tensor product of a two-dimensional parameter space ( $d=2$ ). (a) Smolyak sparse grids with maximum level  $q=3$ . (b) Full tensor product of  $q=3$ . (c) Smolyak sparse grids with maximum level  $q=5$ . (d) Full tensor product of  $q=5$ .

Table 2.2: The number of sampling points use Smolyak formula and full tensor product formula in d-dimensional sampling space with  $q=3$ .

$d=N_\xi$	Smolyak Formula	Full Tensor Product
1	3	3
2	5	9
3	7	27
$\vdots$	$\vdots$	$\vdots$
$d$	$2 \cdot d + 1$	$3^d$

the derivation of (2.9) where

$$\begin{aligned}
 H(3, 2) &= (\vartheta^1 \times \vartheta^1) \cup (\vartheta^1 \times \vartheta^2) \cup (\vartheta^2 \times \vartheta^1) \\
 &= \{(a, a)\} \cup \{(a, a), (a, b), (a, c)\} \cup \\
 &\quad \{(a, a), (b, a), (c, a)\}
 \end{aligned} \tag{2.10}$$

$$= \{(a, a), (a, b), (a, c), (b, a), (c, a)\} \tag{2.11}$$

Based on the pristine formulation of Smolyak sparse grid collection method, we should perform the polynomial interpolation for each set of cross product in (2.10). Since the knots in (2.10) are nested, we can execute one polynomial interpolation for the union of collected knots in (2.11) rather than performing polynomial interpolations in (2.10) to improve the efficiency [16].

We take one example in Fig. 2.2 using Clenshaw-Curtis abscissas for the construction of Smolyak formula and compare it with full tensor product interpolation formula to show the reduction of sampling points when applying Smolyak formula. The sampling points using Smolyak formula for the 2-dimensional example is in Fig. 2.2(a) and Fig. 2.2(c) with  $q=3$  and  $q=4$ , respectively. The full tensor grids is shown in Fig. 2.2(b) and Fig. 2.2(d). The counts of sampling points is reduced when using Smolyak formula, and the improvement of counts is clearer in a high dimension sampling space.

In our case, we need to use high dimensional sampling space which will show the drastic reduction of sampling points if using Smolyak sparse grid formula. The comparison of the

number of sampling points using Smolyak formula and full tensor product formula of  $q=3$  is shown in Table 2.2. In Table 2.2, the number of points derived from Smolyak sparse grid formula is linearly dependent on the dimension; nevertheless, it is exponentially dependent on the dimension by using full tensor product to interpolate.

## 2.4 Problem Formulation

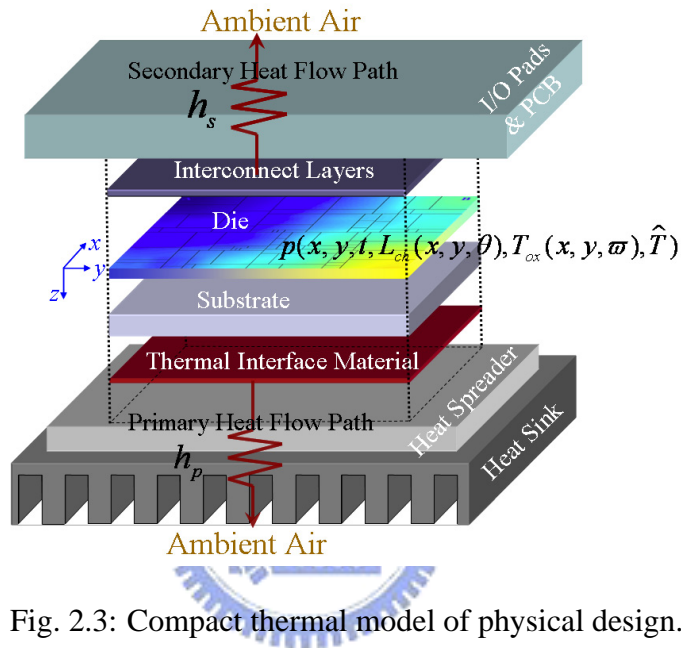


Fig. 2.3: Compact thermal model of physical design.

The compact thermal model of a chip consisting of three portions for physical design stage [17, 18] can be represented by Fig. 2.3. The primary heat dissipation path is composed of thermal interface material, heat spreader, and heat sink. The secondary heat dissipation path involves interconnect layers, I/O pads, and the print circuit board. The functional blocks on the die are modeled as many power generating sources attached to the thin layer close to the top surface of the die with the thickness being equal to the junction depth of device [19]. The main heat sources are coming from the dynamic and leakage power consumed by devices. Because the dynamic power is insensitive to process variations, it can be treated to be deterministic. However, the leakage power is strongly dependent on process parameters such as channel length and oxide thickness. As considering process variations, these parameters need to be viewed as random processes [5]. Moreover, leakage is also highly sensitive to the temperature; hence, the thermal coupling needs to be taken into account for modeling the statistical leakage power.

By combining the compact thermal model and the statistical power consumption considering thermal coupling, the steady state temperature distribution  $\widehat{T}(\mathbf{r}, \theta, \varpi)$  of die is determined by the following statistical steady-state heat transfer equation.

$$\nabla \cdot \left( \kappa(\mathbf{r}, \widehat{T}) \nabla \widehat{T}(\mathbf{r}, \theta, \varpi) \right) = -p \left( \mathbf{r}, L_{ch}(x, y, \theta), T_{ox}(x, y, \varpi), \widehat{T} \right), \quad (2.12)$$

subject to the following boundary condition

$$\kappa(\mathbf{r}_{b_s}, \widehat{T}) \frac{\partial \widehat{T}(\mathbf{r}_{b_s}, \theta, \varpi)}{\partial n_{b_s}} + h_{b_s} \widehat{T}(\mathbf{r}_{b_s}, \theta, \varpi) = f_{b_s}(\mathbf{r}_{b_s}). \quad (2.13)$$

Here,  $\nabla$  is the diverge operator, and  $\kappa(\mathbf{r}, \widehat{T})$  is the thermal conductivity (W/m $\cdot$ °C) of die. The  $p(\mathbf{r}, L_{ch}(x, y, \theta), T_{ox}(x, y, \varpi), \widehat{T})$  is the random process of power density profile which consists of dynamic power density profile  $p_d(\mathbf{r})$ , subthreshold leakage power density profile  $p_{sub}(\mathbf{r}, L_{ch}(x, y, \theta), T_{ox}(x, y, \varpi), \widehat{T})$ , and gate tunneling leakage power density profile  $p_{gate}(\mathbf{r}, L_{ch}(x, y, \theta), T_{ox}(x, y, \varpi), \widehat{T})$ . The  $\mathbf{r} = (x, y, z) \in D$ ,  $D = (0, L_x) \times (0, L_y) \times (-L_z, 0)$  is the domain of die,  $L_x$  and  $L_y$  are lateral sizes of die, and  $L_z$  is the thickness of die. The  $\theta$  and  $\varpi$  are sampling values of manufacturing outcomes  $\Omega_{L_{ch}}$  and  $\Omega_{T_{ox}}$  for the channel length and oxide thickness, respectively. The  $L_{ch}(x, y, \theta)$  and  $T_{ox}(x, y, \varpi)$  are the random processes of the device channel length and the oxide thickness, respectively. The  $b_s$  is any specific boundary surface of the die, and  $\mathbf{r}_{b_s}$  is the position located on  $b_s$ . The  $h_{b_s}$  is the heat-transfer coefficient on  $b_s$ ,  $f_{b_s}(\mathbf{r}_{b_s})$  is the heat flux function on  $b_s$ , and  $\partial/\partial n_{b_s}$  is the differentiation along the outward direction normal to  $b_s$ .

Since the major part of device current passes through the channel, the power density distribution has its value only when  $\mathbf{r} \in (0, L_x) \times (0, L_y) \times (-j_d, 0)$ . Here,  $j_d$  is the junction depth of device [19].

With the statistical steady-state heat transfer equations (2.12)–(2.13), our goal is to evaluate the mean and variance profiles of steady-state full-chip temperature distribution considering spatially correlated intra-die process variations, inter-die process variations, and electro-thermal coupling.



# Chapter 3

## Statistical Electro-Thermal Framework

### 3.1 Statistical Electro-Thermal Flow

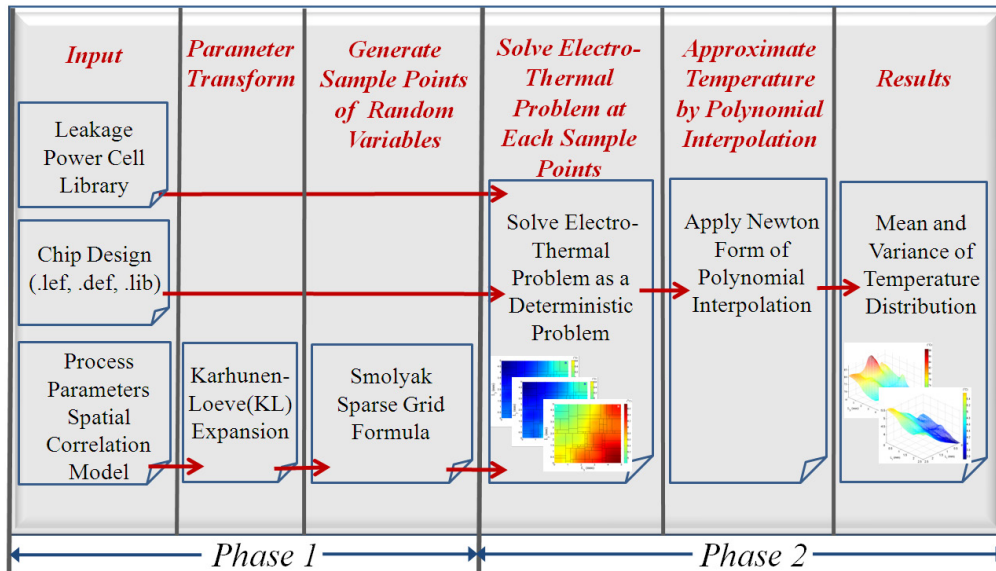


Fig. 3.1: Sparse grid based statistical electro-thermal simulation flowchart.

The flowchart of the developed sparse grid based statistical electro-thermal simulation is shown in Fig. 3.1, and it consists of two phases. Each operation in *Phase 1* is only related with technology node rather than design pattern, and the operations of *Phase 2* are design dependent.

In the beginning of *Phase 1*, to take the temperature effect into account for the leakage power cell library, accurate forms of statistical cell-based subthreshold and gate leakage current models are developed and detailed in section 2.2. Then, given a spatial covariance function of physical parameters such as the channel length and the oxide thickness, the KL expansion is employed to decompose the correlated physical parameters into a set of uncorrelated random variables which are introduced in section 3.2. After that, the Smolyak sparse grid formula

in [9] is applied to generate a set of sampling points of the random space expanded by these uncorrelated random variables and inter-die random variables.

In *Phase 2*, a proposed Smolyak sparse grid based statistical electro-thermal simulation method is used to construct an interpolation formula of the stochastic temperature profile over the chip. For each sampling point on the space of random variables, an electro-thermal coupling algorithm shown in Fig. 3.2 is used to get the deterministic thermal profile of the chip. Then, all thermal profiles are integrated to build an interpolation representation of stochastic temperature profile over the chip. Finally, the statistical temperature profile can be extracted. The detail is presented in section 3.3.

Since each operation in *Phase 1* is irrelevant to design pattern, they only need to be pre-performed once while applying the proposed statistical electro-thermal simulator to the optimal thermal-aware design procedure. Therefore, the proposed statistical electro-thermal simulator has the high compatibility for the power model and the function of spatial correlation model.

For example, as the technology is advanced and different leakage current model forms are required to maintain the accuracy, only the leakage power models in *Phase 1* need to be reconstructed and the rest procedures are unchanged. However, the related works, [2] and [8], are limited by specific power model forms and can not maintain the accuracy at different technology node.

The advantages of the proposed sparse grid based statistical electro-thermal simulator are summarized as follows.

1. Any spatial covariance functions can be adopted.
2. Any complex leakage current models especially taking thermal effect into account can be dealt with in the simulator. Therefore, leakage current models can be very complex to reach very high accuracy without reserving cares of simulating complexity in *Phase 2*.
3. It can readily apply the parallel programming to improve efficiency because the generating procedure of thermal profile at each sampling knot is uncorrelated, and the simulation results at all sampling knots only need to be collected in the end.

## 3.2 Parameter Modeling

Generally, process variations of one physical parameter  $P$  can be classified into intra-die  $\Delta P^{intra}$  and inter-die  $\Delta P^{inter}$  variations which both can be modeled as Gaussian random variables [5].

The physical parameter  $P \in \{T_{ox}, L_{ch}\}$  with its expected value  $\bar{P}$  at position  $\mathbf{r}$  can be written as

$$T_{ox}(\mathbf{r}, \varpi) = \bar{T}_{ox}(\mathbf{r}) + \Delta T_{ox}^{intra}(\mathbf{r}, \varpi_i) + \Delta T_{ox}^{inter}(\mathbf{r}, \varpi_j), \quad (3.1)$$

$$L_{ch}(\mathbf{r}, \theta) = \bar{L}_{ch}(\mathbf{r}) + \Delta L_{ch}^{intra}(\mathbf{r}, \theta_i) + \Delta L_{ch}^{inter}(\mathbf{r}, \theta_j). \quad (3.2)$$

Here,  $\varpi_i$  and  $\varpi_j$  are subsets of  $\varpi$ . The  $\theta_i$  and  $\theta_j$  are subsets of  $\theta$ .

According to [5],  $T_{ox}(x, y, \varpi)$  is spatially uncorrelated. Because the spatial correlation of  $\Delta L_{ch}^{intra}(\mathbf{r}, \theta_i)$  may have different decreasing rates in  $x$ - and  $y$ -directions, the spatial covariance function proposed in [20] is adopted for  $\Delta L_{ch}^{intra}(\mathbf{r}, \theta_i)$ . Given  $\sigma$  as the standard deviation of target random process, and correlation lengths  $\eta_x$  and  $\eta_y$  in  $x$ -direction and  $y$ -direction, respectively, the spatial covariance function between two random variables at points  $\mathbf{r}_1$  and  $\mathbf{r}_2$  is

$$C(\mathbf{r}_1, \mathbf{r}_2) = \sigma^2 \exp\left(-\frac{|x_1 - x_2|}{\eta_x}\right) \exp\left(-\frac{|y_1 - y_2|}{\eta_y}\right). \quad (3.3)$$

*Remark:* Although we choose this specific spatial covariance function (3.3) in this work, any valid spatial covariance functions can be adopted in the proposed electro-thermal simulation flow.

With applying KL expansions,  $\Delta L_{ch}^{intra}(\mathbf{r}, \theta_i)$  based on function (3.3) can be approximated as

$$\Delta L_{ch}^{intra}(\mathbf{r}, \theta_i) \approx \sum_{m=1}^{N_{L_{ch}}} \sqrt{\chi_m} q_m(\mathbf{r}) \zeta_m(\theta_i). \quad (3.4)$$

Here,  $\chi_m$ 's are eigenvalues of  $C(\mathbf{r}_1, \mathbf{r}_2)$ ,  $q_m$ 's are related eigenvectors, and  $N_{L_{ch}}$  is the expansion length.  $\{\zeta_m(\theta_i)\}$  is the set of uncorrelated standard normal random variables. According to the property of KL expansion, the expanded random variables are Gaussian random variables if the target random process is Gaussian [21]. The closed form of eigen-pairs  $(\chi_m, q_m(x, y))$  can be derived by [21]. In this paper,  $\zeta = \{\zeta_m\}$  and  $\varsigma = \{\varsigma_n\}$  which are sets of random variables to represent  $L_{ch}$  and  $T_{ox}$ , respectively. We drop  $\theta$  and  $\varpi$  for the sake of notation simplicity.

### 3.3 Smolyak Sparse Grid Interpolation Based Simulation

Given the placement/floorplan of circuit and technology files, the leakage power models developed in section 2.2 are built, and the Karhunen-Loève expansion is used to transform the spatially correlated process parameters to a set of uncorrelated random variables. Then, the expanded random variable set of inter-die and intra-die variations for  $L_{ch}$  and  $T_{ox}$  is represented as  $\{\xi_1, \dots, \xi_d\}$  which is the union of  $\zeta$  and  $\varsigma$ . For simplicity, we use  $\tilde{\xi} = (\xi_1, \dots, \xi_d)^T$  to represent these  $d$ -dimensional random variables. Based on the concepts of Smolyak formula in (2.7) and (2.9), we can set  $d$ , the number of random variables, as the dimension of the functional space and  $q$  as the level of the desired solution to acquire sampling points. After that, roots of the Hermite polynomial chaos [22] are chosen as sampling points for achieving the best approximation in the  $q$  level [6] since the temperature profile over the chip,  $T$ , is a function of normal random variables.

Fig. 3.2 shows the algorithm of the electro-thermal procedure afterward. The algorithm is applied to each sampling point until all sampling points are accomplished. With each sampling point over the expanded random space,  $\Delta L_{ch}$  and  $\Delta T_{ox}$  of each specified position on the chip can be obtained. Then, the leakage power profile of the design can be acquired by Leakage-power-updating algorithm shown in Fig. 3.3. Since the temperature profile is built by partitioning the die region into  $P_{row} \times Q_{col} = PQ$  blocks, each block may be across the process variation grids. The process-variation grids section the die region over  $U_{row} \times V_{col} = UV$  grids, and within each grid, the process parameters are viewed as having the same characteristic of variation. Here,  $P, Q, U$ , and  $V$  are the user setting integers for deciding the numbers of blocks or grids meshed over the chip. In Fig. 3.3, the leakage power of one temperature block is continually added by using equation (2.1) until all types of functional gates and all process-variation grids inside the block have been done. Then, the leakage power profile is added with dynamic power profile to obtain the total power profile. After using any existing deterministic thermal simulators <sup>1</sup>, the total power profile can be transformed into temperature profile. Because of considering electro-thermal coupling, the temperature profile needs to be performed iteratively by updating the leakage power until the temperature profile transformed

---

<sup>1</sup>The GIT [18] is used in this work.

---



---

**Algorithm** Electro-thermal-coupling

**Input:** A sampling point  $\tilde{\xi}^i$ , initial temperature  $T^{ini*}$ , dynamic power\*, and switching activity  $Sw^*$ .

**Output:** Stable temperature  $T^*(\tilde{\xi}^i)$

---



---

1 **Begin**

2  $T_{ox}^*$  and  $L_{ch}^*$  can be obtained according to  $\tilde{\xi}^i$

3  $T^* \leftarrow T^{ini*}$ ,  $T^{*'} \leftarrow 0$

4 **While** ( $T^* - T^{*'} \leq$  converging criterion)

5   **do**  $T^{*'} \leftarrow T^*$

6     Leakage power\*  $\leftarrow$  *Leakage-power-updating*

7     Total power\*  $\leftarrow$  Leakage power\* + dynamic power\*

8     Using GIT<sup>†</sup> to transfer total power\* into  $T^*$

9     **if** ( $T^* =$  Infinite) **then** Thermal runaway

10 **return**  $T^*$

11 **End**

---



---

\* denotes the distributed values over a chip.

† one deterministic thermal simulator [18]. Any deterministic thermal simulators can be used here.

Fig. 3.2: Electro-thermal-coupling algorithm

from updating power profile is slight changed.

Finally, Newton's interpolating method [23] is applied for generating an interpolated polynomial to approximate  $T$  with the set of sampling points,  $\{\tilde{\xi}^k\}_{k=1}^m$ . The temperature of multivariate interpolated polynomial form expanded by  $\tilde{\xi}$  can be written as

$$T(\tilde{\xi}) = \hat{a}_1\phi_1(\tilde{\xi}) + \cdots + \hat{a}_n\phi_n(\tilde{\xi}) + \cdots + \hat{a}_m\phi_m(\tilde{\xi}), \quad (3.5)$$

where each  $\phi_n(\tilde{\xi})$  is a function of  $\tilde{\xi}$  in this expanded space, and  $\{\hat{a}_1, \cdots, \hat{a}_m\}$  is the set of the unknown coefficients of Newton's interpolating polynomial [23].

Based on the basic idea of interpolation that the approximated function must match each known data, the interpolated polynomial in (3.5) must satisfy the following equation for each  $\tilde{\xi}^k$ .

$$\hat{a}_1\phi_1(\tilde{\xi}^k) + \cdots + \hat{a}_n\phi_n(\tilde{\xi}^k) + \cdots + \hat{a}_m\phi_m(\tilde{\xi}^k) = T(\tilde{\xi}^k), \quad (3.6)$$

Furthermore, according the property of  $\phi_n(\tilde{\xi})$  which is constructed in a particularly way [23],

---



---

**Algorithm** Leakage-power-updating

**Input:**  $T_{ox}^*$ ,  $T^*$ ,  $L_{ch}^*$ ,  $Sw^*$ , and Leakage Power Cell Library

**Output:** Leakage Power\*

---



---

```

1 Begin
2 For each temperature block  $T(p, q) \in T^*$ 
3   do  $T \leftarrow T(p, q)$ 
4     For each process-variation grid  $T_{ox}(u, v) \in T_{ox}^*$ ,  $L_{ch}(u, v) \in L_{ch}^*$ 
5       do  $T_{ox} \leftarrow T_{ox}(u, v)$ 
6          $L_{ch} \leftarrow L_{ch}(u, v)$ 
7         For each gate type occurring in this process-variation grid (u,v)
8           do  $P_{Leak}$  density  $\leftarrow$  (equation (2.1) with  $(T_{ox}, L_{ch}, T)$ ,  $Sw^*$ )
9              $\times$  gate-area portion of the block's area
10           $P(p, q)$  added with  $P_{Leak}$ 
11 return Leakage Power* is constructed by  $P(p, q)$  for  $p = 1 \rightarrow P$  and  $q = 1 \rightarrow Q$ 
12 End

```

---



---

\* denotes the distributed values over a chip.

Fig. 3.3: Leakage-power-updating algorithm

(3.6) can be written as the following matrix-vector expression for finding each  $\hat{a}_n$ .

$$\begin{bmatrix} \phi_1(\tilde{\xi}^1) & 0 & \cdots & 0 \\ \phi_1(\tilde{\xi}^2) & \phi_2(\tilde{\xi}^2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\tilde{\xi}^m) & \phi_2(\tilde{\xi}^m) & \cdots & \phi_m(\tilde{\xi}^m) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_m \end{bmatrix} = \begin{bmatrix} T(\tilde{\xi}^1) \\ T(\tilde{\xi}^2) \\ \vdots \\ T(\tilde{\xi}^m) \end{bmatrix}, \quad (3.7)$$

Each  $\hat{a}_n$  can be calculated in linear time since the system matrix of (3.7) is a lower triangular matrix. After each  $\hat{a}_n$  has been calculated, the statistical temperature profile can be extracted as

$$E\{T(\tilde{\xi})\} = E\{\hat{a}_1\phi_1(\tilde{\xi}) + \cdots + \hat{a}_m\phi_m(\tilde{\xi})\}, \quad (3.8)$$

$$Var\{T(\tilde{\xi})\} = Var\{\hat{a}_1\phi_1(\tilde{\xi}) + \cdots + \hat{a}_m\phi_m(\tilde{\xi})\}. \quad (3.9)$$

Fig. 3.4 is the simulating algorithm, *SETS*, of the proposed simulator. As discussion in section 3.1, *Phase 2* is the part needed to re-perform when design is changed. *Phase 1* is related to the technology node and unchanged as used the simulator under the same process.

### 3.4 Complexity Analysis

In this section, the complexity of *Phase 2* in Fig. 3.4 is analyzed. The temperature profile over the chip is analyzed into  $PQ$  blocks, where  $P$  and  $Q$  have the same definitions in section 3.3. Equally, the power profile is also approximated by these blocks. According to [18], the complexity of the deterministic thermal solver used in this work is  $O(PQ \log_2 N_x N_y)$ , where  $N_x$

---

---

**Algorithm** Statistical-Electro-Thermal-Simulation (SETS)**Input:** Leakage power cell library, Chip Design, and spatial correlation model**Output:** statistical temperature profile  $E\{T^*\{\tilde{\xi}\}\}$  and  $Var\{T^*\{\tilde{\xi}\}\}$ 

---

---

**Phase 1**

- 1 Parse input files
- 2 Applying Karhunen-Loève expansion to transform the spatially correlated process parameters
- 3 Construct the sampling points by Smolyak sparse formula.

**Phase 2**

- 4 **For** each sampling point  $\tilde{\xi}^i \in \tilde{\xi}$
  - 5   **do** *Electro-Thermal-Coupling*
  - 6       Solve unknown coefficients of Newton form of polynomial interpolation by equation (3.7)
  - 7  $E\{T^*\{\tilde{\xi}\}\}$  and  $Var\{T^*\{\tilde{\xi}\}\}$
- 
- 

\* denotes the distributed values over a chip.

Fig. 3.4: Simulating algorithm of the proposed statistical electro-thermal simulator.

and  $N_y$  are the truncated number of bases in x- and y-direction, respectively, and these are far less than the number of blocks  $PQ$ . Because leakage power is highly correlated by temperature, it is updated by Leakage-power-updating algorithm in Fig. 3.2. In line 4 of Fig. 3.2, because process-variation grids are determined by process rather than the circuit, the grids are usually orders of number less than that of temperature blocks; the temperature blocks are finer than process-variation grids. It also shows that most of temperature blocks have only one process-variation grid inside. Therefore, since there are  $N_{type}$  types, in worst case, of functional gates in each process-variation grid over all temperature blocks, the complexity of updating leakage power for  $PQ$  blocks is  $O(PQN_{type})$ . In general,  $N_{type}$  is determined by the number and the spatial proportion of functional types in the circuit, and it is far less than the number of blocks  $PQ$ , too. To find the worst extreme bound of complexity,  $N_{type}$  in one temperature block of such process-variation grid can be simulated as a cumulative counts of functional types sorted area in an increasing series. It is referred to the maximum  $N_{type}$  is occurred when functional types having smallest area are gathered into one temperature block. For the previous discussion, the computational complexity of one electro-thermal loop from line 5 to line 6 in Fig. 3.2 is  $O(PQ \log_2 N_x N_y) + O(PQN_{type})$ . The iteration of electro-thermal coupling is based on the converging criterion and initial temperature setting. According to our experiment with sampling knots constructed by Monte Carlo method, the average count of iteration loop in Fig. 3.2

is less than 5. The converging criterion of the experiment is set as the temperature value for all blocks are less than 0.5% differing from the value in previous loop, and all the initial temperature values are set as room temperature. We conclude that the computational complexity of electro-thermal coupling algorithm is  $O(rPQ(\log_2 N_x N_y + N_{type}))$ , where  $r$  is the count of average electro-thermal coupling loop.

The simulating algorithm of the proposed statistical electro-thermal simulator is shown in Fig. 3.4. *Phase 2* is the part needed to be recomputed as circuit design changing. In line 6, because the calculation of equation (3.7) is without the computation of matrix inverse and the matrix size is dependent on the number of sampling points  $m$ , the coefficients of it can be obtained in linear time. Since each sampling point needs to enter the electro-thermal coupling algorithm and the statistical temperature profile can be extract in linear time of line 7, the complexity of the proposed simulator is  $O(mrPQ(\log_2 N_x N_y + N_{type}))$ .





# Chapter 4

## Experimental Results

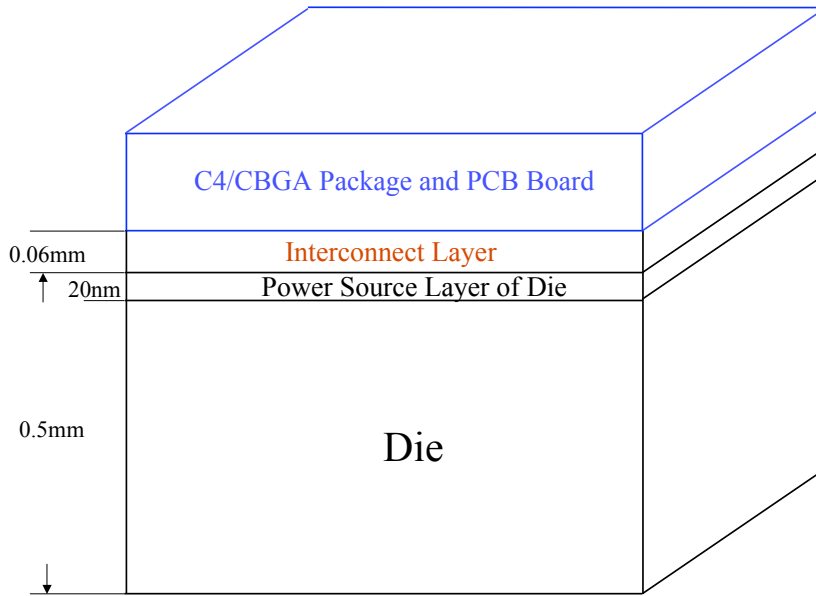
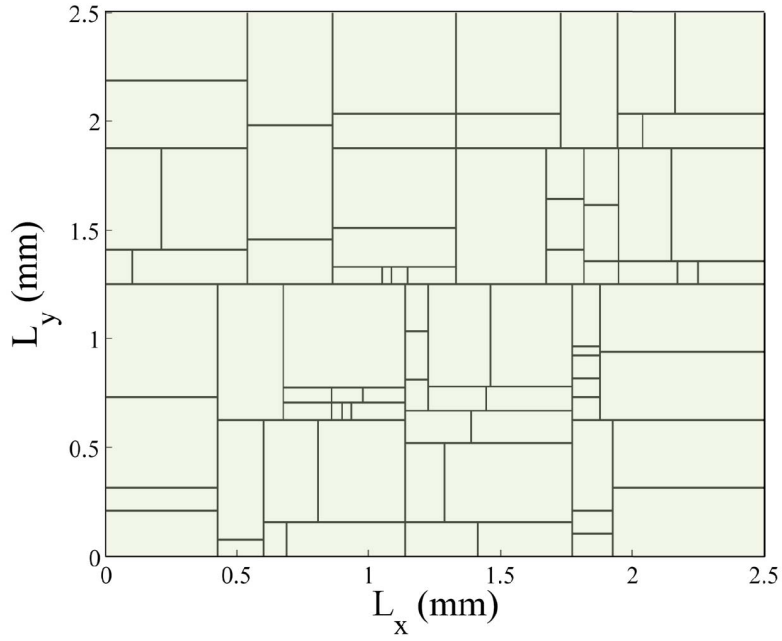
The developed statistical electro-thermal simulator is implemented in C++ language and tested on a Linux system with Intel Xeon 3.0-GHz CPU and 32 GB memory.

The die size is  $2.5\text{ mm} \times 2.5\text{ mm} \times 0.5\text{ mm}$ . The junction depth is  $20\text{ nm}$  which is the nominal value for the  $65\text{ nm}$  technology [19]. The floorplan of test chip which having 1.2 million functional gates is shown as Fig. 4.1(a), and the geometries of chip and package are shown in Fig. 4.1(b). By applying the modeling skill of thermal parameter and iterative 1-D thermal computation scheme [17], the equivalent heat transfer coefficients of the primary and secondary heat flow paths, and thermal conductivity are  $12000\text{ W}/(\text{m}\cdot^\circ\text{C})$ ,  $2017\text{ W}/(\text{m}\cdot^\circ\text{C})$ , and  $148.13\text{ W}/(\text{m}\cdot^\circ\text{C})$ , respectively. The boundary condition of each vertical surface is set to be isothermal [18].

The nominal values of channel length and oxide thickness are  $65\text{ nm}$  and  $1.5\text{ nm}$ , respectively. The  $3\sigma_{L_{ch}}$  and  $3\sigma_{T_{ox}}$  are set to 12% and 5% of nominal values, respectively. Both  $\eta_y/L_y$  and  $\eta_x/L_x$  are set to 0.98 which means the correlation between two devices located half of the chip dimension away in the  $x$ -direction or the  $y$ -direction is 0.6. The temperature profiles is analyzed in  $128 \times 128$  blocks and the process-variation grids is set as  $10 \times 10$  grids. The setting of deterministic thermal simulator with truncated number of basis in  $x$ - and  $y$ -direction are both 32 which can reach higher accuracy than author's recommend in [18].

### 4.1 Accuracy and Efficiency

To verify the simulator, the Monte Carlo (MC) method is also implemented by  $10^5$  samples as reference golden solutions which consider the same issues such as electro-thermal coupling,



(b)

Fig. 4.1: (a) The floorplan of test chip. (b) The geometry setting of test chip.

Table 4.1: Accuracy and efficiency compared with the Monte Carlo method.

Inter-die / Total Variations	Intra-die / Total Variations	Our Proposed Method <sup>†</sup>				Monte Carlo <sup>‡</sup>		Speedup (X)
		max. mean error	max. std. error	runtime (s)		sampling knots	runtime (s) <sup>‡</sup>	
				Phase 1	Phase 2			
40%	60%	0.33%	1.70%	3.23	1.04	6736	326.49	313.93
50%	50%	0.35%	1.88%	3.27	1.04	6465	313.82	301.75
60%	40%	0.36%	1.84%	3.40	1.04	6422	311.47	299.49

<sup>†</sup> Our proposed method is compared with the golden solution constructed by Monte Carlo method using  $10^5$  samples.

<sup>‡</sup> To show the efficiency, Monte Carlo method here is simulated till achieving the same accuracy of standard deviation as our proposed method. The runtime here does not include the time of input parser which is only performed once in Monte Carlo simulation.

spatially intra-die variations, and inter-die variations. The proposed electro-thermal simulator takes 10 random variables to expand process variations and uses Smolyak sparse grid formula with  $q=11$ . Hence, the stochastic thermal profile over the test chip is interpolated by 21 individual sampling points. The results with three different ratios of inter-die variations and intra-die variations to the total variations in a reasonable region are shown in Table 4.1.

Compared with the golden solution, the proposed simulator is extremely accurate and can be finished in seconds for the test chip. For example, in the case of inter-die variations being 50% of total variations, the proposed simulator can achieve the maximum errors of 0.35% and 1.88% in spatial mean and spatial standard deviation of temperature distribution, respectively. The execution time is only 3.27 seconds and 1.04 seconds in *Phase 1* and *Phase 2*, respectively. The similar results can be found in the rest two cases.

Since each operation in *Phase 1* of the proposed simulator is irrelevant to design pattern, they only need to be pre-performed once while applying the proposed simulator to the optimal thermal-aware design procedure. Therefore, to show the efficiency of the proposed simulator, the runtime of *Phase 2* is compared with the execution time through the Monte Carlo simulation fulfilling the same accuracy of standard deviation as ours. Table 4.1 shows that the proposed simulator is orders of magnitude faster than the Monte Carlo analysis under the same accuracy level. the same Since each sampling point is independent, the parallel programming technique can be easily applied to further enhance the speedup.

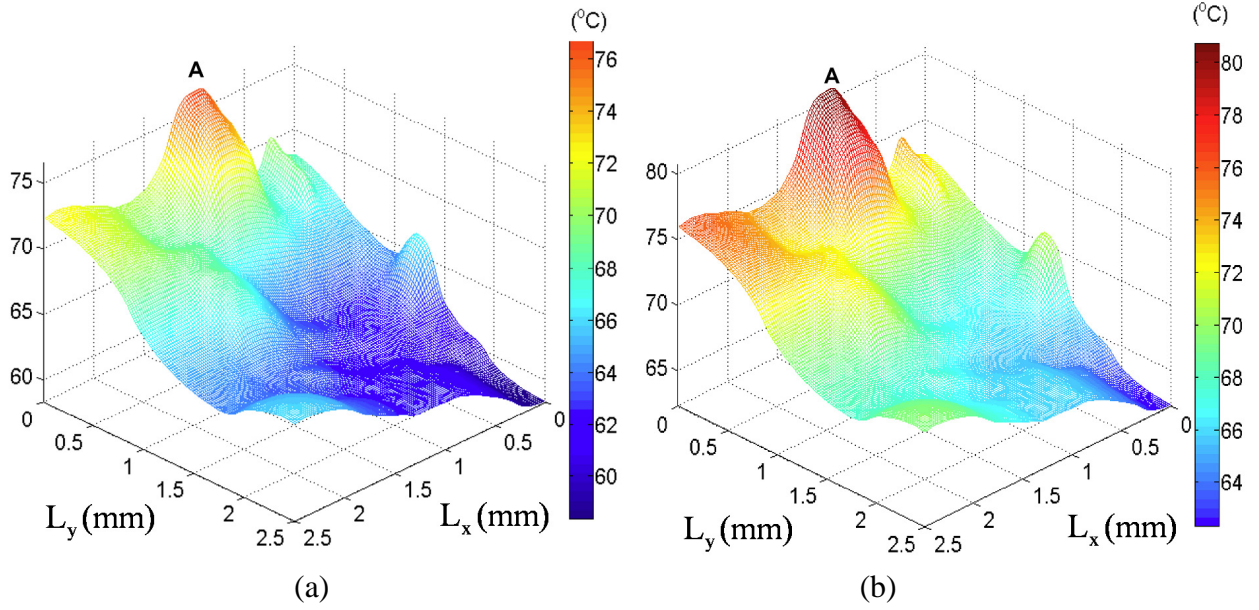


Fig. 4.2: The temperature profile at the top surface of the die. (a) The mean temperature distribution without considering electro-thermal coupling. (b) The mean temperature distribution with considering electro-thermal coupling.

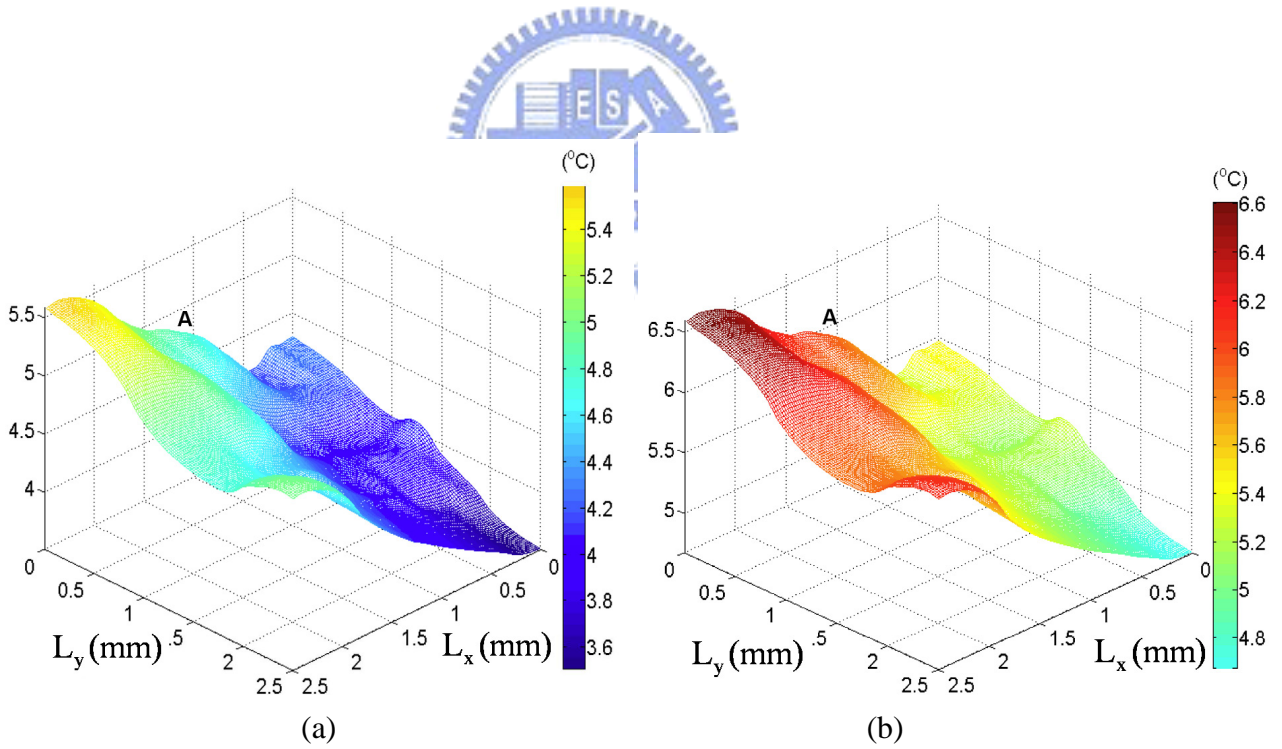
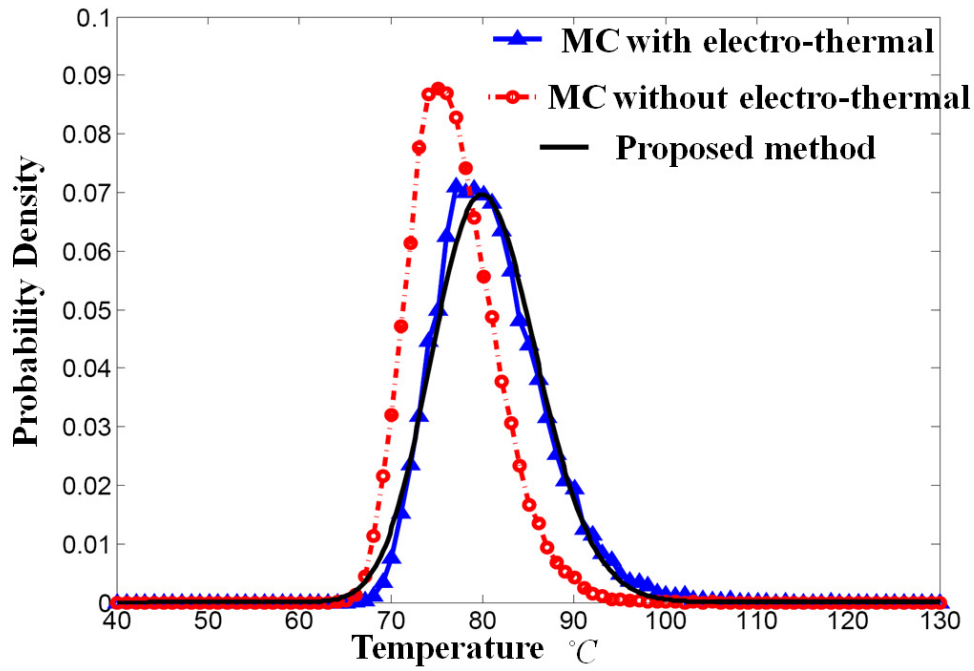
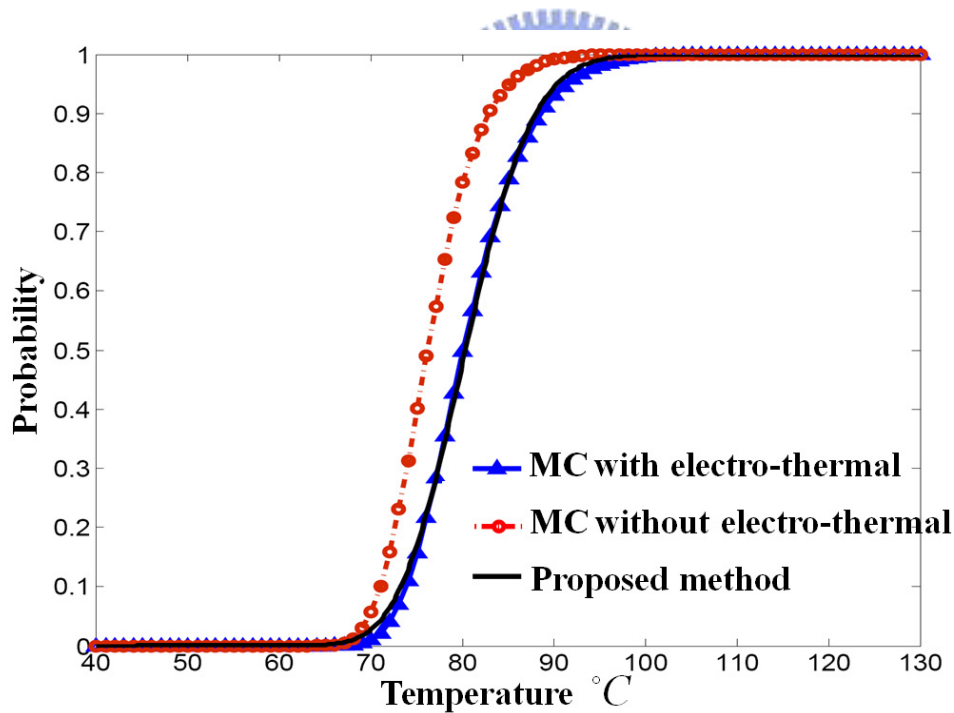


Fig. 4.3: The temperature profile at the top surface of the die. (a) The spatial standard deviations without considering electro-thermal coupling. (b) The spatial standard deviations with considering electro-thermal coupling.



(a)



(b)

Fig. 4.4: Distribution of the temperature using Monte Carlo (MC) simulation, with and without electro-thermal coupling, and the proposed method at the location of the hottest mean temperature. (a) Probability density function (PDF). (b) Cumulative distribution function (CDF).

## 4.2 Without vs. With Including the Effect of Electro-Thermal Coupling

Fig. 4.2 and Fig. 4.3 show the spatial mean and spatial standard deviations of the temperature distribution at the top surface of the test chip, respectively. Fig. 4.2(a) and Fig. 4.3(a) are the results without considering electro-thermal coupling. Fig. 4.2(b) and Fig. 4.3(b) are the results with considering electro-thermal coupling. These two figures reveal the dramatic differences of the spatial mean and spatial standard deviation profiles between the results without considering electro-thermal coupling and the results considering electro-thermal coupling. As we can see, the difference of spatial mean profile can reach 6.54%, and the difference of spatial standard deviation profile is over 25.01%.

According to [8], the temperature profile of each location on the chip can be approximated as a log-normal distribution. The probability density function (PDF) and cumulative distribution function (CDF) of the temperature distribution at an arbitrary location on the chip are plotted in Fig. 4.4(a) and Fig. 4.4(b), respectively. The blue solid line marked in triangles is the result obtained from the Monte Carlo simulation with considering electro-thermal coupling. The red dash line marked in circles is the result acquired from the Monte Carlo simulation without considering electro-thermal coupling. The black solid line is an approximation using log-normal distribution and its mean and variance are obtained by the proposed simulator. Fig. 4.4 shows that the proposed method can provide accurate estimations of PDF and CDF for the thermal profile, and the simulation results without considering electro-thermal coupling are unreliable.

The similar result also happens in the statistical analysis of total leakage power. The PDFs and CDFs of the total leakage power of the test chip by the Monte Carlo simulation are shown in Fig. 4.5. Obviously, the statistical leakage power analysis without electro-thermal coupling is not reliable.

From the above discussion, it shows that the statistical thermal or leakage power analysis method without considering electro-thermal coupling can lead the simulation results into an unreliable region and provide a dubitable confidence interval. To give the correct and reliable analysis results for designers, it is necessary to take electro-thermal coupling into consideration for not only leakage power analysis but also thermal analysis, and the proposed electro-thermal

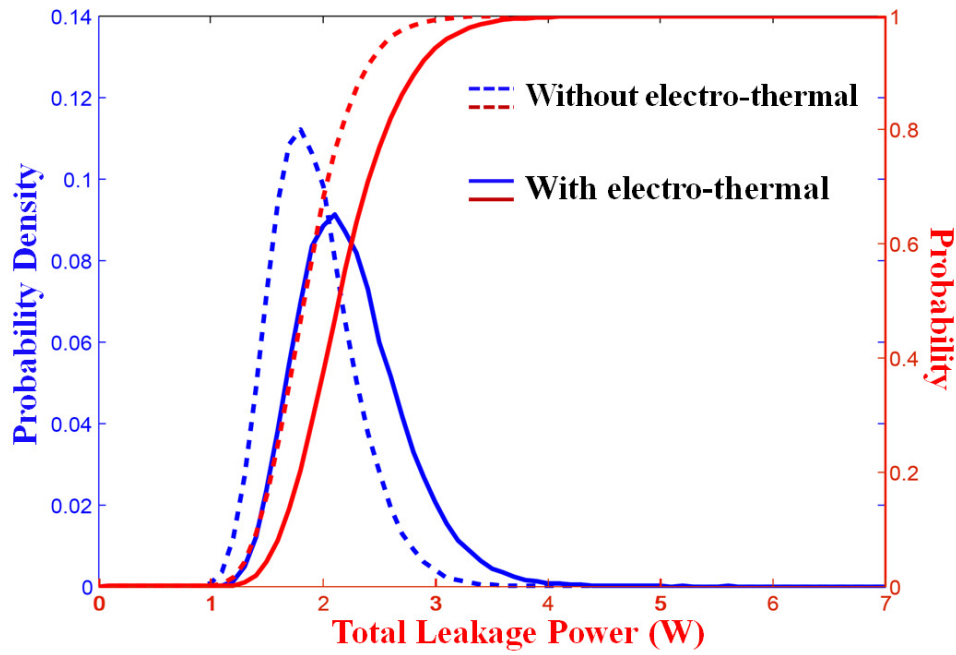


Fig. 4.5: PDFs and CDFs of the total leakage power using MC simulation with and without considering electro-thermal coupling.

simulator can accurately and efficiently achieve these.



# Chapter 5

## Application–Thermal Yield

### 5.1 Thermal Yield of Circuit

Considering process variations, the temperature is approximated as a log-normal random variable at each position over a chip [8], and it is also been verified in Fig. 4.4. Consequently, for a thermal-aware design, our statistical electro-thermal simulator can be applied to provide the thermal yield. The statistical thermal yield can be defined as

$$Yield(T(\xi)) \stackrel{\text{def}}{=} \Pr(T(\xi) < T_{ref}) = \Phi_{\log}(T_{ref}), \quad (5.1)$$

where  $\Phi_{\log}$  denotes the cumulative distribution function of a log-normal random variable, and  $T_{ref}$  is the reference temperature. The probability of exceeding the reference temperature is defined as  $\overline{\Phi_{\log}}(T_{ref}) = 1 - \Phi_{\log}(T_{ref})$ .

In traditionally deterministic thermal analysis, the hottest place is the one that has the highest temperature; that is, the one needs to be reallocated and carefully concerned. Given two PDFs of temperature distribution at two arbitrary locations on a chip as shown in Fig. 5.1, “*R*” will be the more critical position if the conventional worst-case analysis is used for specifying hotspot. However, by comparison, in thermal yield analysis, the place which needs to be well-handled should be the one having the most likelihood of exceeding tolerable temperature. The reason is that the thermal-aware design must first tackle the place with the highest probability of breaking down because it may dominate the full chip reliability. Hence, “*B*” is the more critical position since it has larger  $\overline{\Phi_{\log}^B}(T_{ref})$  which also means that it has a worse thermal yield.



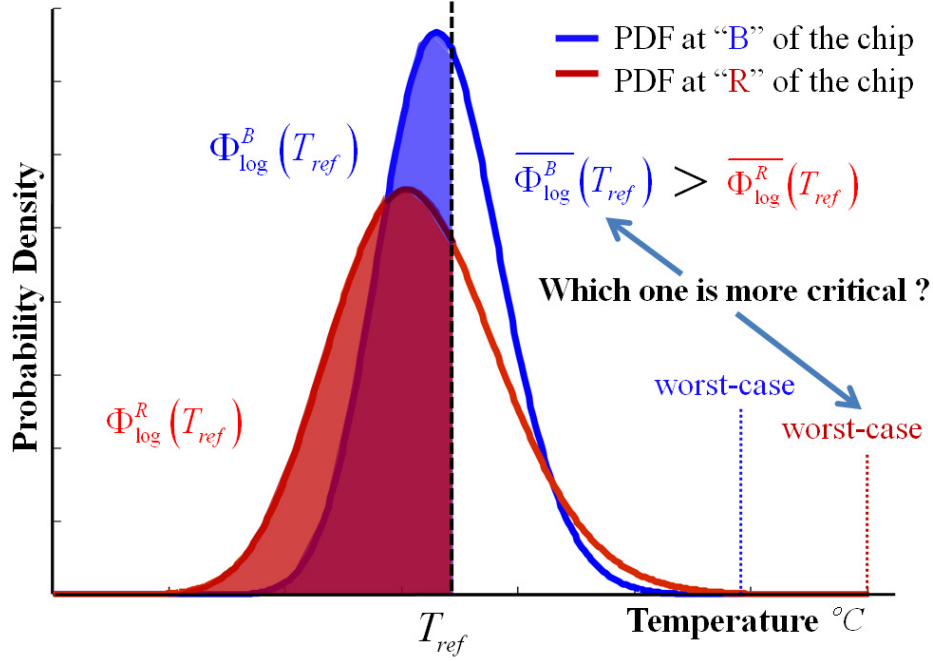


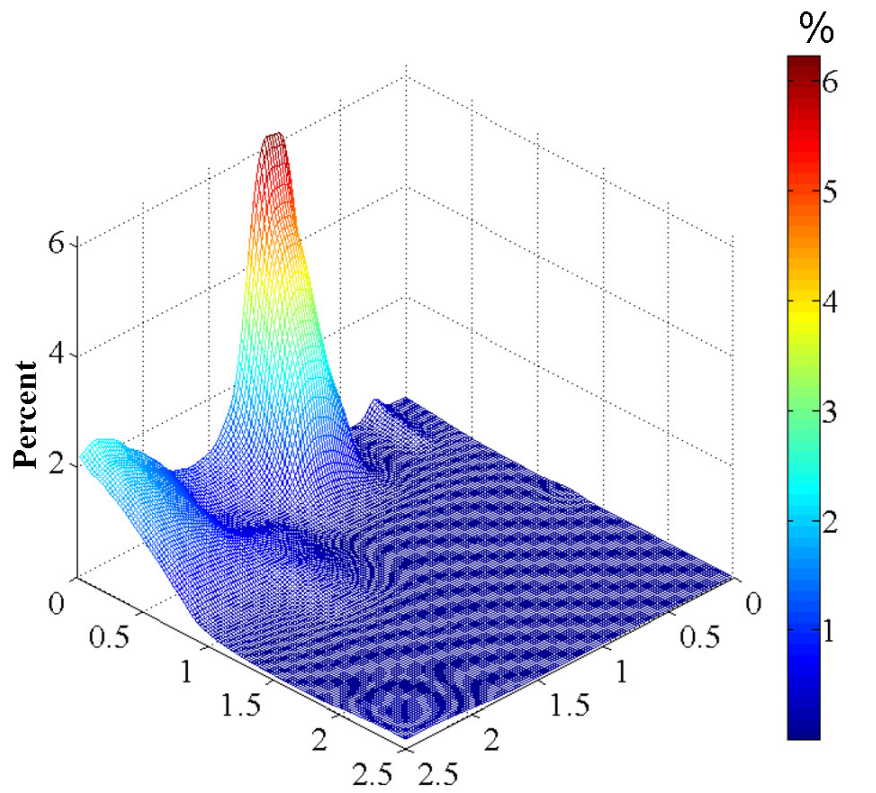
Fig. 5.1: PDFs of the temperature at two locations of the chip for indicating which one is more critical on the chip.

## 5.2 Statistical Thermal Yield Analysis Problem

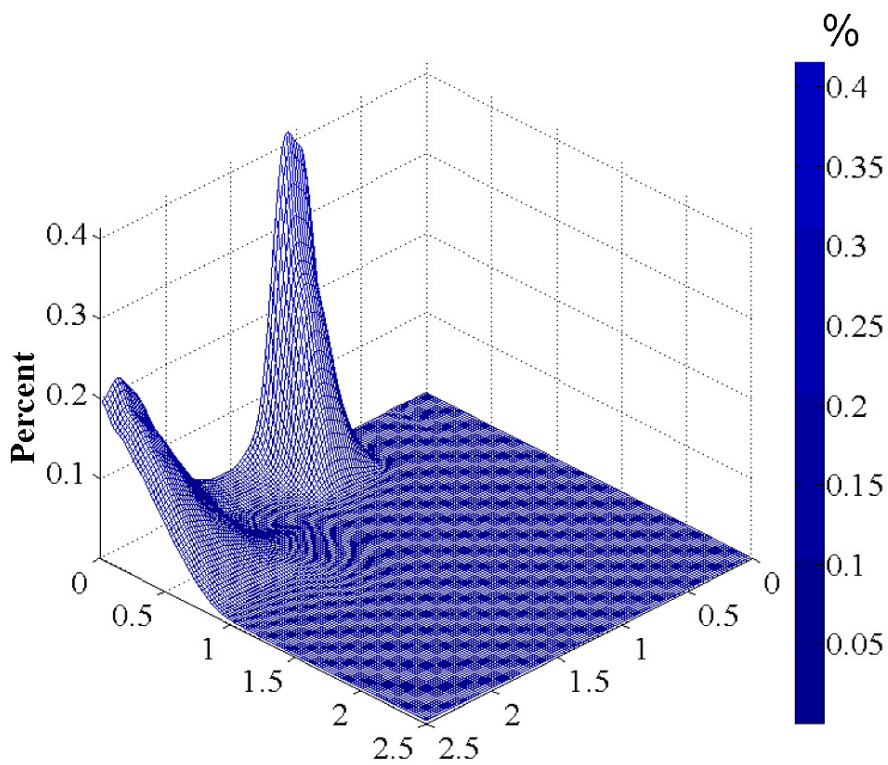
The statistical thermal yield analysis problem for a given circuit is formulated as following:

*For a circuit, given a reference temperature  $T_{ref}$ , analyze the temperature distributed over the chip as considering process variations and get the statistical temperature profiles by dealing with stochastic heat transfer function. Based on the statistical temperature profiles, find the thermal yield,  $\Phi_{\log}(T_{ref})$ .*

To analyze the reliability within the reference temperature as considering process variations, designers can use the simulation results from the proposed simulator and the thermal yield analysis provided in this work.  $\overline{\Phi_{\log}}(T_{ref})$  over our test chip by using the statistical results from the proposed simulator is shown in Fig. 5.2(a) for  $T_{ref}$  being  $90^{\circ}\text{C}$ . The region with the highest probability of exceeding  $T_{ref}$  is the place which needs to be seriously concerned for the chip reliability, because the region has the worst thermal yield. However, by contrast, a thermal yield from one statistical thermal simulator without considering electro-thermal coupling is shown in Fig. 5.2(b). Without considering electro-thermal coupling may lead nearly one order of magnitude lower of thermal yield. To provide designers an correct guideline from thermal yield estimation, it is necessary to take electro-thermal coupling into thermal simulator.



(a)



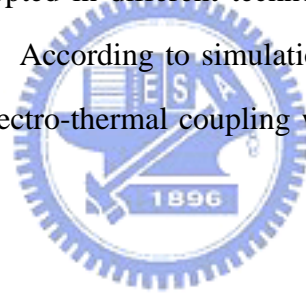
(b)

Fig. 5.2: Probability of exceeding the reference temperature  $\overline{\Phi}_{\log}(T_{ref}) = 90^{\circ}C$  from statistical thermal simulator (a) with considering electro-thermal coupling. (b) without considering electro-thermal coupling.

# Chapter 6

## Conclusions

An efficiently statistical electro-thermal simulator considering inter-die variations and intra-die variations including the spatial correlation has been presented. The proposed simulator can efficiently provide the accurate simulation results and has the advantages of high capability for any complex leakage power models and the spatial correlation function. The statistical electro-thermal framework can be adopted in different technology nodes and assistant designers to correctly predict yield of chip. According to simulation results, we have also indicated that it is not allowable to ignore electro-thermal coupling when considering process variations in statistical thermal simulation.



# Bibliography

- [1] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De. Parameter variations and impact on circuits and microarchitecture. *Proc. Des. Autom. Conf.*, pages 338–342, June 2003.
- [2] P. Y. Huang, J. H. Wu, and Y. M. Lee. Stochastic thermal simulation considering spatial correlated within-die process variations. *Proc. Asia and South Pacific Des. Autom. Conf.*, pages 55–60, June 2009.
- [3] Y. Zhang, D. Parikh, K. Sankaranarayanan, K. Skadron, and M. Stan. Hotleakage: A temperature-aware model of subthreshold and gate leakage for architects. Technical Report CS-2003-05, Univ. of Virginia, May 2003.
- [4] Y. Liu, R. P. Dick, L. Shang, and H. Yang. Accurate temperature-dependent integrate circuit leakage power estimation is easy. *Proc. Des. Auto. Test in Euro. Cof.*, pages 1–6, April 2007.
- [5] H. Chang and S. S. Sapatnekar. Prediction of leakage power under process uncertainties. *ACM Trans. Design Autom. Electron. Syst.*, 12, April 2007.
- [6] R. Shen, N. Mi., and S. Tan. Statistical modeling and analysis of chip-level leakage power by spectral stochastic method. *Proc. Asia and South Pacific Des. Autom. Conf.*, pages 31–36, June 2009.
- [7] K. R. Heloue, N. Azizi, and F. N. Najm. Modeling and estimation of full-chip leakage current considering within-die correlation. *Proc. Des. Autom. Conf.*, pages 93–98, June 2007.

- [8] J. Jaffari and M. Anis. Statistical thermal profile considering process variation: Analysis and applications. *IEEE Trans. Comput.-Aided Des. Integr. Circuit Syst.*, 27:1027–1040, June 2008.
- [9] S. A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, pages 240–243, 1963.
- [10] M. J. Moran and H. N. Shapiro. *Fundamentals of Engineering Thermodynamics*. Wiley, 6 edition, May 2007.
- [11] N. M. Ravindra and J. Zhao. Fowler-nordheim tunneling in thin  $SiO_2$  films. *Smart Mater. Struct.*, 1:197–201, 1992.
- [12] S. A. Yu, P. Y. Huang, and Y. M. Lee. A multiple supply voltage based power reduction method in 3-d ics considering process variations and thermal effects. *Proc. Asia and South Pacific Des. Autom. Conf.*, pages 55–60, June 2009.
- [13] K. Roy, S. Mukhopadhyay, and h. Mahmoodi-Meima. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits. *Proceedings of the IEEE*, pages 305–327, Feb. 2003.
- [14] G. W. Wasilkowski and H. Wozniakowski. Explicit cost bounds of algorithms for multivariate tensor product problems. *Journal of Complexity*, pages 1–56, 1995.
- [15] J. Taylor and F. Hover. High dimensional stochastic simulation and electric ship models. *Digital Object Identifier*, 21-23:402–407, May 2007.
- [16] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, pages 2309–2345, May 2008.
- [17] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. R. Stan. Hotspot: A compact thermal modeling methodology for early-stage vlsi design. *IEEE Trans. Very Large Scale Integr. Syst.*, 14:501–513, May 2006.

- [18] P. Y. Huang and Y. M. Lee. Full-chip thermal analysis for the early design stage via generalized integral transforms. *IEEE Trans. Very Large Scale Integr Syst.*, 17:613–626, May 2009.
- [19] F. Lallement, B. Duriee, A. Grouillet, F. Amaud, B. Tavel, F. Wacquant, P. Stalk, M. Woo, Y. Erokhin, J. Scheuer, L. Gadet, J. Weeman, D. Distaso, and D. Lenoble. Ultra-low cost and high performance 65nm cmos device fabricated with plasma doping. *Symp. VLSI Technol. Dig. Tech. Papers*, pages 178–179, 2004.
- [20] S. Bhardwaj, S. Vrudhula, P. Ghanta, and Y. Cao. Modeling of intra-die process variations for accurate analysis and optimization of nanoscale circuits. *Proc. Des. Autom. Conf.*, pages 791–796, 2006.
- [21] B. Cline, K. Chopra, D. Blaauw, and Y. Cao. Analysis and modeling of cd variation for statistical static timing. *Proc. Int. Conf. on Comput.-Aided Des.*, pages 60–66, 2006.
- [22] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, 2003.
- [23] G. M. Phillips. *Interpolation and Approximation by Polynomials*. Springer, 2003.

