

**ULTRA-FAST TEMPERATURE ESTIMATION
METHODS FOR ARCHITECTURE-LEVEL
THERMAL MODELING**

HAMEEDAH SULTAN



**AMAR NATH AND SHASHI KHOSLA SCHOOL OF
INFORMATION TECHNOLOGY**

INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2021

© **Indian Institute of Technology Delhi (IITD), New Delhi, 2021**

**ULTRA-FAST TEMPERATURE ESTIMATION
METHODS FOR ARCHITECTURE-LEVEL
THERMAL MODELING**

by

HAMEEDAH SULTAN

Amar Nath and Shashi Khosla School of Information Technology

Submitted

**in fulfilment of the requirements of the degree of Doctor of Philosophy
to the**



INDIAN INSTITUTE OF TECHNOLOGY DELHI

OCTOBER 2021

Certificate

This is to certify that the thesis titled **ULTRA-FAST TEMPERATURE ESTIMATION METHODS FOR ARCHITECTURE-LEVEL THERMAL MODELING** being submitted by **Ms. Hameedah Sultan** for the award of **Doctor of Philosophy** in Information Technology is a record of bona fide work carried out by her under my guidance and supervision at the Amar Nath and Shashi Khosla School of Information Technology, Indian Institute of Technology Delhi. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Smruti Ranjan Sarangi

Associate Professor

Department of Computer Science and Engineering

Indian Institute of Technology Delhi

New Delhi - 110016

Acknowledgements

One can never pay in gratitude: one can only pay
'in kind' somewhere else in life.

Anne Morrow Lindbergh

I would like to thank my advisor, Smruti R. Sarangi, for being a constant source of encouragement and guidance throughout this work. His faith in my abilities, combined with his continued leadership, has helped push me to achieve more than what I thought was possible. I want to thank all my student research committee members, Dr. Preeti R. Panda, Dr. Balakrishnan, and Dr. Amit Gupta, for their valuable suggestions and help in my work.

I would like to express my gratitude to my senior, Gayathri Ananthanarayanan, who played a significant role in initiating me into the research and helped familiarize me with the field. I am also thankful to the students I mentored, Shashank, Ankit, Anand, and Anjali, for they provided me with wonderful learning opportunities. I would also like to thank all the helpful members of my lab and friends: Ismi, Lokesh, Rajesh, Rajshekar, Prathmesh, Janib, Neetu, Sakshi, Hadi, Priyanka, Diksha, Omais, Divya, Shubhankar and Shashank.

I would like to extend a special thanks to Shashank for sharing his Icepak microchannels model and SPICE simulations that I used for validating my work.

I am also thankful to the very helpful administrative staff, Ms. Vandana, Ms. Manju, Mr. Rajesh and Mr. Suresh.

I became a mother in graduate school. And it does not take willpower or resilience to navigate graduate school as a mother. It takes more than that – it takes a strong support network. My deepest gratitude goes to the incredible support network I had that helped me get through grad school. If I had to pick which aspect of the support I received was more instrumental in helping me get to the end of this journey, I would fail badly. Each one of them contributed in ways uniquely significant and more than others. My parents-in-law left the peace and quiet of their

retirement life to come and take care of my vibrant and energetic one-year-old daughter so that I could focus my time and energy on my research. How can I express my gratitude to them in words? Then there is my father, who taught me to always strive for excellence. My brother, Abdur Rahman, believed in my abilities before anyone else and urged me to better utilize my potential. My mother, whose dream has been to see me achieve what she did not get the opportunities for and has made endless sacrifices for me. My husband Shadab always had an ear for my rants and kept encouraging me to go on even in the most adverse circumstances. And then there is my little daughter Maryam, who has the supernatural ability to make all my worries disappear through one little smile of hers, and to whom all of my work is dedicated.

Above all, all thanks to Almighty God, who guided me and provided me the opportunity to reach this stage.

Hameedah Sultan

Abstract

As the power density of modern-day chips has increased, the chip temperature, too, has increased steadily. High temperature causes several adverse effects, affecting the chip's performance and reliability. It also increases the leakage power, which further increases the on-chip temperature, resulting in a feedback effect. In order to carry out temperature-aware design optimization, it is often necessary to conduct thousands of temperature simulations at various stages of the design cycle, and thus the speed of simulation without a concomitant loss in accuracy is essential.

State-of-the-art works in thermal estimation have serious limitations in modeling some important aspects of thermal modeling. Additionally, these methods are slow. We overcome the limitations of these works by developing fast Green's function-based analytical methods.

Green's function-based methods have been shown to be faster than the traditional finite difference and finite element methods, but their use has been limited to simple problems in the past. In this thesis, we develop Green's function-based methods for a wide range of complex thermal problems and demonstrate their superior efficiency compared to existing works.

In the first work, we developed a fast and analytical leakage-aware thermal simulator for 3D chips using Green's functions (impulse response of a unit power source). We use several novel algebraic techniques and transforms to arrive at a closed-form expression for the Green's function. Next, we extend this work to compute the thermal profile for 3D chips with microchannels by taking care of the anisotropy in the thermal profile of such chips.

After that, we proceed to model the effects of process variation on the temperature since process variation is increasingly becoming important as the device dimensions are decreasing. We propose an analytical thermal simulator that considers the effects of process variation, temperature-dependent conductivity, and temperature-dependent leakage power altogether.

However, an analytical process variation-aware method cannot be extended to complex geome-

tries. So in our final work, we propose a convolutional neural network architecture for thermal estimation in the presence of variability that leverages the physics of the heat transfer. This approach is capable of modeling modern-day 3D chips with microchannels and incorporates accurate leakage power models.

In summary, by using a combination of rare mathematical transforms, algebraic manipulations, and numerical and machine learning-based methods, we not only solve these complex problems but also achieve greater than $100\times$ speedup over the state-of-the-art simulators in all of our works. Thus, the work presented in this thesis significantly advances the state-of-the-art in ultra-fast thermal estimation.

सार

जैसे-जैसे आधुनिक चिप्स की शक्ति घनत्व में वृद्धि हुई है, चिप के तापमान में भी लगातार वृद्धि हुई है। उच्च तापमान चिप के प्रदर्शन और विश्वसनीयता को प्रभावित करते हुए कई प्रतिकूल प्रभाव डालता है। यह रिसाव शक्ति (leakage power) को भी बढ़ाता है, जो आगे ऑन-चिप तापमान बढ़ाता है, जिसके परिणामस्वरूप प्रतिक्रिया प्रभावित होती है।

तापमान-जागरूक डिज़ाइन अनुकूलन करने के लिए अक्सर डिज़ाइन चक्र के विभिन्न चरणों में हजारों तापमान सिमुलेशन करना आवश्यक होता है। इस लिए अनुकरण की गति बिना सहवर्ती सटीकता नुकसान के अत्यन्त आवश्यक है।

तापमान प्रतिमान में अत्याधुनिक कार्यों की कुछ गंभीर सीमाएं हैं। इसके साथ ही ये तरीके धीमे हैं। हम अत्याधुनिक कार्यों की सीमाओं पर काबू पाने के लिए अत्यन्त-तेज ग्रीन फंक्शन पर आधारित विश्लेषणात्मक तरीकों को विकसित करते हैं। ग्रीन-फंक्शन-आधारित तरीकों को पारंपरिक परिमित-अंतर (finite-difference) तथा परिमित-तत्व विधियों (finite-element) की तुलना में तेज़ दिखाया जा चुका है, लेकिन उनका उपयोग अतीत में साधारण समस्याओं तक ही सीमित रहा है।

इस थीसिस (thesis) में हम जटिल तापमान प्रतिमान की विस्तृत श्रेणी के लिए ग्रीन-फंक्शन-आधारित तरीकों का विकास करते हैं। साथ ही हम मौजूदा कार्यों की तुलना में उनकी बेहतर दक्षता का प्रदर्शन करते हैं। पहले काम में हमने 3D चिप्स के लिए ग्रीन फंक्शन (एक इकाई शक्ति स्रोत की आवेग प्रतिक्रिया) का उपयोग कर एक तेज़ और विश्लेषणात्मक लीकेज-जागरूक थर्मल सिम्युलेटर विकसित किया। हम ग्रीन-फंक्शन के लिए एक बंद-रूप अभिव्यक्ति पर पहुंचने के लिए कई नवीन बीजगणितीय तकनीक और परिवर्तन का उपयोग करते हैं।

इसके बाद हम इस काम का माइक्रोचैनल-सहित 3D चिप्स के तापमान-प्रतिमान के लिए एनाइसोट्रोपिसिटी (anisotropy) का ख्याल रखते हुए अधिक विस्तार करते हैं।

जैसे-जैसे डिवाइस के आयाम कम होते जा रहे हैं, प्रक्रिया भिन्नता (process variation) तेजी से महत्वपूर्ण होती जा रही है। इस लिए अपने दूसरे काम में हम तापमान पर प्रक्रिया भिन्नता के प्रभावों को मॉडल करते हैं। हम एक विश्लेषणात्मक थर्मल सिम्युलेटर प्रस्तावित करते हैं जो तापमान-निर्भर प्रक्रिया भिन्नता, तापमान-निर्भर प्रवाहकत्व (conductivity) तथा तापमान-निर्भर रिसाव शक्ति के प्रभावों को पूरी तरह से ध्यान में रखते हुए तापमान प्रतिमान करता है।

हमारी प्रस्तावित विश्लेषणात्मक प्रक्रिया भिन्नता-जागरूक पद्धति को जटिल ज्यामिति तक नहीं बढ़ाया जा सकता है। इसलिए अपने अंतिम कार्य में, हम प्रक्रिया भिन्नता-जागरूक तापमान प्रतिमान के लिए एक कॉवोलुशनल न्यूरल नेटवर्क (convolutional neural network) का प्रस्ताव करते हैं जो इस समस्या के समाधान के लिए गर्मी हस्तांतरण (heat transfer) के भौतिकी का लाभ उठाता है। यह विधि आधुनिक समय के माइक्रोचैनल वाली 3D चिप्स में सटीक रिसाव शक्ति मॉडल सहित तापमान अनुमान में करने में सक्षम है।

संक्षेप में, दुर्लभ गणितीय परिवर्तनों के संयोजन का उपयोग करके, बीजीय जोड़तोड़ और संख्यात्मक और मशीन लर्निंग-आधारित विधियों का इस्तेमाल कर, हम न केवल इन जटिल समस्याओं को हल करते हैं बल्कि अपने सभी कार्यों में हम अत्याधुनिक सिमुलेटर पर 100× से अधिक स्पीडअप भी प्राप्त करते हैं। इस प्रकार इस थीसिस में प्रस्तुत कार्य अत्यंत-तेज़ तापमान प्रतिमान की अत्याधुनिक विधियों को महत्वपूर्ण रूप से आगे बढ़ाता है।

Contents

Certificate	i
Acknowledgements	iii
Abstract	v
1 Introduction	1
1.1 Challenges in Thermal Modeling	2
1.2 Problems Solved: A Brief Overview	3
1.2.1 A Survey of Chip-level Thermal Simulators	3
1.2.2 Leakage-Aware Thermal Estimation in 3D Chips with Microchannels .	4
1.2.3 Analytical Variability-Aware Thermal Simulation for 2D chips	4
1.2.4 Variability-Aware Thermal Simulation Using CNNs	5
1.3 Organization of This Thesis	5
2 A Survey of Architectural Thermal Simulators	7
2.1 Introduction	7
2.2 Background	8
2.2.1 Structure of a Chip in an IC	9

2.2.2	Power Consumption in a Chip	11
2.2.3	Modes of Heat Transfer	11
2.2.4	Heat Transfer in a Chip	13
2.2.5	Heat transfer at the Device Level	14
2.2.6	Analogy with Electrical Circuits	16
2.2.7	Numerical Methods	17
2.2.8	Green's Functions	20
2.3	Taxonomy of Solutions	21
2.3.1	Scope of the Survey	21
2.3.2	Classification Criteria	22
2.4	Finite Element based Techniques	23
2.4.1	Device-Level Thermal Modeling Using FEM	24
2.4.2	System-Level Thermal Modeling Using FEM	25
2.5	Finite Difference based Techniques	25
2.5.1	Simulators Solving the Heat Conduction Equation	25
2.5.2	Simulators Invoking the Analogy to Electrical Techniques	26
2.5.3	Simulators for 3D ICs	31
2.5.4	Modeling Leakage Power	35
2.6	Spectral or Transform based Techniques	36
2.6.1	Analytically Computing the Green's Functions	37
2.6.2	Using Pre-Obtained Green's Functions	38
2.6.3	Modeling Leakage Power	39
2.6.4	Miscellaneous Techniques	42

2.7	Machine Learning Techniques	42
2.7.1	Feature Extraction based Techniques	43
2.7.2	Neural Network based Techniques	45
2.8	Discussion and Chapter Conclusion	47
2.8.1	Accuracy Needed at the Architectural Level	47
2.8.2	Leakage Power	48
2.8.3	Tradeoffs of the Various Techniques	50
2.8.4	A Word on FinFETs and SoI Transistors	52
2.8.5	Conclusion	52
3	Leakage-Aware Thermal Estimation in 3D Chips with Microchannels	55
3.1	Introduction	55
3.2	Background	57
3.2.1	Power Dissipation in a Chip	57
3.2.2	Thermal Modeling	58
3.3	Methodology	60
3.3.1	Model of the Chip	60
3.3.2	Overview of our Thermal Modeling Approach	61
3.3.3	Obtaining the Green's Functions without Leakage	63
3.3.4	3D Leakage Aware Green's Function	64
3.3.5	Convolution with the Power Profile	70
3.3.6	Full-Chip Transient Thermal Profile	70
3.3.7	Corrections for Edges and Corners	71
3.3.8	Techniques used for Further Speed-up	71

3.4	Modified Thermal Modeling Approach for Microchannels	72
3.4.1	Residual Green's Functions for Chips with Microchannels	72
3.4.2	Corrections for Edges and Corners	74
3.5	Piecewise Linear Leakage Model	75
3.6	Evaluation	78
3.6.1	Architecture of the 3D Chip	78
3.6.2	Setup	79
3.6.3	Error Metric	80
3.6.4	Results for 3D Chip with Heat Sink	80
3.6.5	Results for 3D Chip with Microchannels	85
3.6.6	Scalability Analysis of the Proposed Approach	89
3.6.7	Error Upon Using a Linear Leakage Model	90
3.6.8	Chapter Conclusion	90
4	VarSim: A Fast Process Variation-Aware Thermal Modeling Methodology	91
4.1	Introduction	91
4.2	Theoretical Preliminaries	93
4.2.1	Process and Temperature Variation	93
4.2.2	Leakage power	94
4.2.3	Conductivity of Silicon	95
4.3	Related Work	95
4.3.1	Effects of Process Variation on Temperature	96
4.3.2	Modeling the Temperature Dependence of Conductivity	97
4.4	Thermal Estimation Considering Variability	99

4.4.1	Overview	99
4.4.2	Variation of the Thermal Conductivity with Temperature	101
4.4.3	Modified Green's Functions for the Steady-State	102
4.4.4	Modified Green's function for the Transient Case	107
4.4.5	Full-chip Transient Thermal Profile for Time-varying Sources	109
4.4.6	Thermal Estimation at the Edges and Corners	111
4.5	Evaluation	112
4.5.1	Setup	112
4.5.2	Calibration of the Setup	113
4.5.3	Steady-State Results	113
4.5.4	Transient Results	118
4.5.5	Analysis of Results	123
4.6	Chapter Conclusion	124
5	Variability-Aware Thermal Simulation Using CNNs	125
5.1	Introduction	125
5.2	Theoretical Preliminaries	127
5.2.1	Leakage Power	127
5.2.2	Convolutional Neural Networks (CNN)	128
5.2.3	Related Work	129
5.3	Green's functions and CNNs	131
5.3.1	Convolutional Neural Network (CNN)	131
5.3.2	Green's Function-based Thermal Estimation	132
5.3.3	Motivation for a CNN based approach	133

5.4	A Learning based Thermal Estimation Framework Considering Variability . . .	134
5.4.1	Overview	134
5.4.2	Modeling Process Variation	135
5.4.3	Architecture of the CNN: <i>CNNbasic</i>	136
5.4.4	Architecture of a more advanced CNN: <i>CNN3D</i>	137
5.4.5	Data Collection	138
5.4.6	Data Normalization	138
5.4.7	Training of the CNN	139
5.4.8	GPU Implementation	139
5.4.9	Transient Thermal Estimation	140
5.5	Evaluation	140
5.5.1	Setup	140
5.5.2	Accuracy of the tool 3D-ICE	143
5.5.3	Results	143
5.5.4	GPU vs CPU Implementation	147
5.5.5	Generalization to new Floorplans	147
5.5.6	Effects of Process Variation	147
5.5.7	Comparison with State-of-the-art Approaches	149
5.6	Chapter Conclusion	150
6	Conclusion and Future Work	151
6.1	Conclusion	151
6.2	Future Work	152

CONTENTS

xv

Bibliography

153

List of Publications

167

Biography

169

List of Figures

1.1	An overview of the work done	3
2.1	A typical 3D chip	9
2.2	Organization of the survey	10
2.3	An electrical circuit	16
2.4	Taxonomy of solutions	22
2.5	Evolution of finite difference based techniques	26
2.6	Modeling technique of HotSpot	29
2.7	Transform based technique	36
2.8	Leakage power vs temperature using the BSIM equation	50
3.1	3D chip with microchannels	61
3.2	Overview of our approach	62
3.3	Handling the anisotropy in the Green's function	73
3.4	Edge effects in chips with microchannels	75
3.5	Overview of our piecewise linear leakage modeling approach	76
3.6	Results for test case 1 for a chip with heat sink	81
3.7	Power and thermal profiles for test case 2 for a chip with heat sink	81

3.8	Power and thermal profiles for test case 2 for layers 1 and 4 for a chip with microchannels	84
3.9	Power and temperature map for test case 3 for a chip with microchannels . . .	87
3.10	Power and temperature map, transient	88
3.11	Transient temperature map for test case 1	88
3.12	Transient power and thermal distribution, layer 4	89
4.1	Overview of our approach	98
4.2	Leakage power in the presence of variability. Note the spatial correlation in the leakage power values.	110
4.3	Evaluation for test case 1: Alpha21264	115
4.4	Evaluation for test case 3: high variance	116
4.5	Error in various scenarios: considering variability in leakage power is as important as considering the temperature dependence of leakage	117
4.6	Transient Green's function considering all effects	120
4.7	Power and temperature map, transient, actual	121
4.8	Power and temperature map, transient, calculated	122
4.9	Transient evolution of temperature: test case 1	122
4.10	Transient evolution of temperature: test case 2	123
5.1	2D CNN Architecture	134
5.2	3D CNN Architecture	135
5.3	CNN Architecture for transient thermal estimation	139
5.4	Floorplans used for training and testing. A wide variety of floorplans are chosen to ensure that the results generalize to unseen floorplans easily.	142
5.5	Power and thermal profiles of the training data for one of the cache layers . . .	144

5.6	Power and thermal profiles for training data for the core layer	144
5.7	Power and thermal profiles for test data for cache layer	145
5.8	Power and thermal profiles for test data for core layer	145
5.9	Error upon ignoring process variation	148

List of Tables

2.1	Notations used in this survey	9
2.2	Variability in power, performance and temperature	48
2.3	Leakage models commonly used	49
2.4	Error and computation time for the BSIM leakage model and its approximations (40-80°C at increments of 0.1°C)	49
2.5	Percentage error for the SPICE leakage model and its approximations	49
2.6	Representative techniques from each category and their timing	52
3.1	Glossary	60
3.2	Parameters of the chip	78
3.3	Location and magnitude of dynamic power sources on the chip in test case 1 . .	82
3.4	Speed and accuracy of popular simulators (steady state)	83
3.5	Test cases used for evaluating chips with microchannels	83
3.6	Scalability to different grid sizes (steady state)	90
4.1	Glossary	100
4.2	Parameters of the chip [1]	112
4.3	Speed of the studied simulators	113
4.4	Errors in various scenarios	114

4.5	Steady-state test cases summary	118
4.6	Speed of existing simulators	119
5.1	Different combinations used for training and testing	146