TECHNIQUES TO ESTIMATE AND MITIGATE THE EFFECTS OF THERMAL COUPLING AND PROCESS VARIATION IN MANYCORE PROCESSORS

GAYATHRI A

SCHOOL OF INFORMATION TECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY DELHI
AUGUST 2017

TECHNIQUES TO ESTIMATE AND MITIGATE THE EFFECTS OF THERMAL COUPLING AND PROCESS VARIATION IN MANYCORE PROCESSORS

by

GAYATHRI A.

SCHOOL OF INFORMATION TECHNOLOGY

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

Indian Institute of Technology Delhi
August 2017

Certificate

This is to certify that the thesis titled **TECHNIQUES TO ESTIMATE AND MITIGATE THE EFFECTS OF THERMAL COUPLING AND PROCESS VARIATION IN MANYCORE PROCESSORS** being submitted by **Gayathri A** for the award of **Doctor of Philosophy** in **School of Information Technology** is a record of bona-fide work carried out by her under our guidance and supervision at the **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.

M. Balakrishnan
Professor
Department of Computer Science
and Engineering
Indian Institute of Technology Delhi
New Delhi, India 110 016

Smruti R. Sarangi
Assistant Professor
Department of Computer Science
and Engineering
Indian Institute of Technology Delhi
New Delhi, India 110 016

Acknowledgments

I would like to thank all the people who contributed in some way to the work described in this thesis. First and Foremost, I thank my advisors Prof. M. Balakrishnan and Dr. Smruti R Sarangi for their continued support and guidance throughout my Ph.D. I started my research working with Dr. Smruti R Sarangi. I am really thankful to him for his immense help during the initial days of my Ph.D. Words cannot express the profound gratitude that I have for Prof Balakrishnan. During the very trying times of my life, his words of wisdom, constant encouragement and faith in me, gave me enough strength to come out of the lean patch and bounce back to normalcy.

I would also like to thank my thesis committee members: Prof. Anshul Kumar, Prof. Preeti R Panda and Prof. Shouribrata Chatterjee for their insightful comments, suggestions and feedback on my work. In addition, I would like to thank Dr. Turbo Majumder and Prof. Kolin Paul for all the active discussions on my research work. I am immensely grateful to Dr. Turbo Majumder for giving me unrestricted access to use the `Tilera`™ machine for my research work.

I would also like to thank Mrs. Vandana Ahluwalia and Mr. S. D. Sharma for their help with the lab resources. I also thank Mr. Rajesh Kumar, Mr. Suresh and Mrs. Sunita for their assistance in all administrative matters.

I am immensely blessed to have a wonderful and great set of friends at IIT Delhi. It made all the difference during my stay at IIT Delhi. I would also like to thank the family of Prof Balakrishnan for all the care, warmth and affection they had shown on me.

Thank you Prof. Ashoke Kumar Sarkar and Prof. Aditya Prasad Koley. Thanks for everything. Lastly, I would like to mention my family. They have been extremely supportive of all my endeavors, right from a very young age and I am grateful to them for all the warmth, support and care over the years.

Gayathri A.

Abstract

Rapid evolution in the transistor fabrication technology has resulted in transistors with less than 10nm gate length. This advancement in the CMOS technology scaling along with the Moore's law has enabled the designers to put more cores inside the same die. Modern processor architects use this phenomenon to build highly complex and sophisticated cores. This has resulted in a steady increase in performance over the last few decades. But, upon the death of Dennard scaling, the power density does not remain constant with reduction in the transistor size making it impossible to sustain the increase in performance. The continuous increase in the power envelope, elevates peak temperatures of the chip, resulting in increased leakage power dissipation. This increase in power dissipation not only constrains the design options but also affects the chip reliability and increases the packaging and cooling costs. With the continuous increase in the number of cores and its complexity, power and temperature have emerged as first order design constraints. Fabrication of transistors to accurate dimensions is another serious challenge in the nanometer era. Because of the inherent limits introduced by the manufacturing process and the underlying device physics, variability in transistor dimensions and its electrical characteristics is inevitable. In this thesis, we study the different factors that can potentially degrade the scaling trends of multicore and manycore processors, quantify their effects and propose techniques to mitigate or reduce their impact on performance and power consumption.

We first study the nature of lateral heat conduction in large multicore processors and its associated implications on leakage power. We then propose a novel technique based on Hankel transforms to efficiently evaluate onchip temperature in the presence of lateral heat conduction, and leakage power. We formulate the heat-spread aware task assignment problem as an optimization problem and use insights from the derived mathematical model and design a family of algorithms to solve it heuristically. We then consider the effects of process variation. We semi-analytically demonstrate that the number of *usable cores decrease as the amount of process variation increases. We show that it is possible to mitigate this trend to some extent by adopting novel multicore architectures such as dynamic and asymmetric cores. Now, instead of considering temperature and process variation separately, we consider both their effects together. Consequently, we consider a variant of our placement problem in which we consider process variation. For manufacturing processes that have a high degree of process variation, our heuristics can measurably reduce leakage power, as well as maximum temperature reached by the chip.*

Overall, the objective of this work is to study the impact on power and performance due to thermal linkages as well as process variations in modern manycore processors.

ट्रांजिस्टर निर्माण प्रौद्योगिकी में तेजी से विकास के परिणामस्वरूप ट्रांजिस्टर में 10 नैनोमीटर से कम गेट की लंबाई होती है। मूर के कानून के साथ सीएमओएस टेक्नोलॉजी स्केलिंग में यह प्रगति ने डिजाइनरों को एक ही मरने के अंदर और अधिक कोर डाल दिया है। आधुनिक प्रोसेसर आर्किटेक्ट अत्यधिक जटिल और परिष्कृत कोर के निर्माण के लिए इस घटना का उपयोग करते हैं। इससे पिछले कुछ दशकों में प्रदर्शन में लगातार वृद्धि हुई है। लेकिन, डेंनार्ड स्केलिंग की मृत्यु के बाद, ट्रांजिस्टर आकार में कमी के साथ बिजली घनत्व स्थिर नहीं रहता है जिससे प्रदर्शन में वृद्धि को बनाए रखना असंभव हो। पावर लिफाफे में निरंतर वृद्धि, चिप के चोटी के तापमान को बढ़ाता है, जिसके परिणामस्वरूप वृद्धि हुई रिसाव पावर अपव्यय होता है। पावर अपव्यय में यह वृद्धि न केवल डिजाइन विकल्प को रोकता है बल्कि चिप विश्वसनीयता को भी प्रभावित करती है पैकेजिंग और कूलिंग लागत बढ़ जाती है कोर की संख्या और इसकी जटिलता में निरंतर वृद्धि के साथ, शक्ति और तापमान पहले आदेश डिजाइन की कमी के रूप में उभरा है। सटीक आयामों के लिए ट्रांजिस्टर का निर्माण नैनोमीटर युग में एक और गंभीर चुनौती है। विनिर्माण द्वारा शुरू की गई अंतर्निहित सीमाओं के कारण प्रक्रिया और अंतर्निहित डिवाइस भौतिकी, ट्रांजिस्टर आयामों में परिवर्तनशीलता और इसकी विद्युत विशेषताओं अनिवार्य है। इस थीसिस में, हम विभिन्न कारकों का अध्ययन करते हैं जो संभावित रूप से मल्टीकोर और कईकोर प्रोसेसर के स्केलिंग रुझान को कम कर सकते हैं, उनके प्रभाव का अनुमान लगा सकते हैं और प्रदर्शन को कम करने या उनके प्रभाव को कम करने के लिए तकनीक का प्रस्ताव और बिजली की खपत हम सबसे पहले बड़े मल्टीकोर प्रोसेसर में पार्श्व गर्मी प्रवाहकत्त्व की प्रकृति और रिसाव शक्ति पर इसके संबंधित निहितार्थ का अध्ययन करते हैं। हम तो हेंकेल पर आधारित एक उपन्यास तकनीक का प्रस्ताव करते हैं जो कि पार्श्व की उपस्थिति में कुशलता से ऑनचिप तापमान का मूल्यांकन करता है गर्मी प्रवाहकत्त्व, और रिसाव शक्ति हम एक ऑप्टिमाइज़ेशन की समस्या के रूप में गर्मी फैल जाने वाले कार्य कार्य की समस्या को तैयार करते हैं और व्युत्पन्न गणितीय मॉडल से अंतर्दृष्टि का उपयोग करते हैं और एल्गोरिदम के एक परिवार को यह आनुषंगिक रूप से हल करने के लिए डिजाइन करते हैं। हम तब प्रक्रिया भिन्नता के प्रभावों पर विचार करते हैं। हम अर्ध–विश्लेषणात्मक रूप से प्रदर्शित करते हैं कि उपयोग की जाने वाली कोर की संख्या में प्रक्रिया भिन्नता बढ़ जाती है। हम दिखाते हैं कि यह है गतिशील और असममित कोर जैसे उपन्यास मल्टीकोर आर्किटेक्चर को अपनाने के द्वारा कुछ हद तक इस प्रवृत्ति को कम करना संभव है। अब, तापमान और प्रक्रिया भिन्नता को अलग से परिलक्षित करने के बजाय, हम दोनों के प्रभाव दोनों पर विचार करते हैं। नतीजतन, हम अपनी प्लेसमेंट समस्या का एक प्रकार देखते हैं जिसमें हम प्रक्रिया भिन्नता पर विचार करते हैं विनिर्माण प्रक्रियाओं के लिए जो प्रक्रिया भिन्नता का उच्च स्तर है, हमारे ह्यूरिस्टिक्स कर सकते हैं मापने से रिसाव की शक्ति कम हो जाती है, साथ ही चिप द्वारा अधिकतम तापमान तक पहुंच जाता है। कुल मिलाकर, इस काम का उद्देश्य थर्मल संबंधों के कारण शक्ति और प्रदर्शन पर प्रभाव के साथ–साथ आधुनिक कई कोर प्रोसेसर में प्रक्रिया विविधताओं का अध्ययन करना है।

Contents

List of Figures

List of Tables