Guest Editors' Introduction

Hot Chips 30

John Kubiatowicz
University of California, Berkeley

Stefan Rusu
TSMC North America

Over its 30-year history, the Hot Chips conference has become a leading forum for the latest computing, interface, and networking chips. The conference covers technical details about these chips, including technology, fundamental algorithms, packaging techniques, architecture, and circuit details. The emphasis is on real chips and applications which are part of actual products; although we do have our share of research talks, we explicitly avoid theoretical talks and marketing hype.

The 2018 conference received sixty (60) submissions from many corners of the globe, including Europe, Asia, and North America. For the final program, we selected 24 hot presentations, along with two riveting keynote presentations—including a first-time ever tag-team keynote on Spectre/Meltdown security flaws. In addition to high-end CPU's and networking chips, this year's Hot Chips included an important focus on user-facing topics such as security, edge computing, neural network acceleration, and graphics processing. Hotchips even included novel use of technology for storage (nanotube memory) and analog computation using FLASH memory.

In addition to high-end CPU's and networking chips, this year's Hot Chips included an important focus on user-facing topics such as security, edge computing, neural network acceleration, and graphics processing.

Digital Object Identifier 10.1109/MM.2019.2899510
Date of current version 15 March 2019.

All-in-all, this year's Hot Chips was exciting and well attended.

As in previous years, we have selected a sampling of the best presentations from Hot Chips 30 to expand to full papers. This issue of IEEE Micro brings you six such papers, spanning the full gamut of topics from this year.

SECURITY BECOMES FRONT AND CENTER

The microprocessor world was shocked to its core (pun intended?) in 2018 with the appearance of practical timing channel attacks that exploited speculative execution to break the kernel/user, user/user, and user/enclave protection boundaries. These attacks strike at the heart of the last 20 years of computer architectural optimizations and prompted David Patterson and John Hennessey to propose the need for new architectural paradigms (which they loosely called "Architecture 2.0") in their Turing Award talk at ISCA 2018. Two such attacks, called "Spectre" and "Meltdown," are the topic of our first paper, titled "On the Spectre and Meltdown Processor Security Vulnerabilities" by Hill *et al.* This paper details the source of these vulnerabilities and postulates what is necessary for the microprocessor industry to move forward. This paper is a must-read for those who wish to understand the brave new world in which we find ourselves and to see how something as apparently innocuous as speculative execution can lead to a complete breakdown in security boundaries.

Of course, security concerns reach far beyond machine rooms and desktop devices. The last decade has seen the embedding of processors

0272-1732 © 2019 IEEE
Published by the IEEE Computer Society

into everything from smart environments, to cars, to fitness devices, to refrigerators, and toasters. The so-called "Internet of Things" (IoT) has been thrust upon the world to great fanfare, brave new applications, and in some cases to great security breaches. Several high-profile distributed denial of service attacks have already been perpetrated by exploiting the weak security of IoT devices. Thus, our second paper, titled "The Hardware Security Behind Azure Sphere," is timely. In it, Doug Stiles from Microsoft describes a new class of security-conscious microcontrollers which support best-in-class security utilizing roots of trust, attested code, and collaborative (cloud-based) fault detection.

HOT CHIPS BREAD AND BUTTER: HIGH-END CHIPS

At its core, the Hot Chips conference is about hot new silicon. Consequently, we are happy to present two papers that describe the state of the art in desktop and mobile processors. Our third paper, titled "Cascade Lake: Next Generation Intel Xeon Scalable Processor," by Arafa *et al.*, describes Intel's latest server-class processor. The authors discuss the features and performance of the Cascade Lake CPU, which includes newly minted instructions for neural networking applications and special support for persistent memory. It also includes mitigations for side-channel attacks as described by Hill *et al.* in our first paper.

In our fourth paper, titled "Samsung M3 Processor," Rupley *et al.* describe Samsung's latest processor targeted at the mobile Android environments. This chip has a variety of advantages over previous generations. The paper discusses the features and performance of the new pipeline, as well as exploring some of the design processes that went into this new chip. It highlights how sophisticated our mobile CPUs have become.

We have selected a sampling of the best presentations from Hot Chips 30 to expand to full papers. This issue of IEEE MICRO brings you six such papers, spanning the full gamut of topics from this year.

PERSISTENCE, PERSISTENCE, EVERYWHERE

Low-power persistent storage is the holy grail for today's mobile and embedded devices. Furthermore, high-density, fast, cost-effective persistent storage is likewise important to warehouse-scale computing environments. Although FLASH has been the gold-standard for persistent memory storage technology, it exhibits a variety of undesirable properties, including limited lifetime (and the resulting need for complex wear-leveling techniques) and relatively high write energy. FLASH is also not as fast as DRAM and thus is often relegated to lower levels of the memory hierarchy, such as SSD. A variety of replacements have been proposed for FLASH over the years, but each of them has disadvantages of one sort or another. Our fifth paper, titled "Will Carbon Nanotube Memory Replace DRAM?" by Gervasi, describes an exciting new memory technology utilizing carbon nanotubes, called "NRAM." NRAM is persistent and able to be rewritten arbitrarily often without degradation, while being strictly better than DRAM on other axes such as performance, density, and write energy. It is further on the cusp of availability. Do not miss this exciting development.

AGILE PROCESSOR DESIGN

Finally, it is no secret that the design of complex processor pipelines has become so challenging that only large teams of designers and validators can hope to produce some of the more sophisticated modern processors. Our final paper shifts the complexity needle back toward smaller design teams by discussing the process behind the design and implementation of the Berkeley BROOM processor. This academic paper, titled "BROOM: An Open-Source Out-of-Order Processor With Resilient Low-Voltage Operation in 28-nm CMOS," is by Celio *et al.* It discusses the *agile* approach that the Berkeley design team utilized to design and implement the BROOM processor, complete with an out-of-order execution pipeline and low-voltage SRAM-based memories. As described by the authors, the BROOM prototype taped out after a four-month design cycle.

Space limitations prevent us from including more presentations from Hot Chips 30. Most of these presentations, as well as others from previous years, are available at http://www.hotchips.org. We hope that you find these papers as exciting as we do. We also encourage you to attend Hot Chips 31 in August of this year.

John Kubiatowicz is Professor of Electrical Engineering and Computer Science at the University of California at Berkeley, Berkeley, CA, USA. His specialties include computer architecture, operating systems, and networking. His research interests include speculative approaches for computer design, such as quantum, biological, and autonomic computing, as well as issues in internet-scale systems design, namely security, privacy, and denial-of-service resilience. He received the Ph.D. degree in electrical engineering and computer science from MIT and the dual B.S. degrees in electrical engineering and physics, as well as the M.S. degree in electrical engineering and computer science from MIT. One of his best known projects from the early 2000s, titled "OceanStore," recently received a "Most Influential Paper" award from the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2018). At this time, he is particularly interested in data-centric approaches to securing information in environments with secure edge computing. Contact him at kubitron@cs.berkeley.edu.

Stefan Rusu is a Senior Director at TSMC North America, San Jose, CA, USA. His VLSI circuit design experience includes more than 23 years with Intel Corporation and six years at Sun Microsystems. He developed key technologies that were adopted across multiple processor designs, including differential clocking, thermal mapping, adaptive clock deskew, cache leakage shut-off, Serial VID, and yield recovery for core, cache and I/O circuits. He has authored more than 90 papers on VLSI circuit technology and holds 50 U.S. patents. He is an IEEE Fellow and a member of the Technical Program Committee for ESSCIRC and A-SSCC conferences. He served as an Associate Editor of the IEEE Journal of Solid-State Circuits between 2005 and 2015. He was the recipient of the ISSCC Beatrice Winner Award for Editorial Excellence in 2009 and is a member of the Solid-State Circuits Society AdCom. Contact him at smrusu@gmail.com.