GUEST EDITOR'S INTRODUCTION

Top Picks From the 2020 Computer Architecture Conferences

Daniel A. Jiménez, *Texas A&M University, College Station, TX, 77843, USA*

Every year,* IEEE Micro *selects some of the most notable articles published in the previous year's computer architecture conferences to highlight in this Special Issue. In early 2021, the selection committee chose 12 articles to appear as "Top Picks" and an additional 12 deserving an "Honorable Mention" from among the 2020 conferences.

It is my pleasure to introduce to you *IEEE Micro*'s Top Picks from the 2020 computer architecture conferences. Every year, *IEEE Micro* publishes a special issue with the previous year's Top Picks as chosen by a selection committee. It was my privilege this year to be chosen as the selection committee chair and special issue guest editor. This year's selections showcase a great diversity of research topics making an impact in our community. From traditional microarchitecture, to machine learning, nonvolatile memory management, security, and quantum computing, our community has been very busy redefining and reinventing computer architecture research. *IEEE Micro* imposes a hard limit of 12 articles that can be chosen as Top Picks. It was very hard to summarize all the great research of 2020 in just a dozen articles but the selection committee did an excellent job. The committee chose another 12 articles as "Honorable Mentions" described in the "Honorable Mentions" sidebar. I encourage you to read each of the Top Picks and Honorable Mentions.

REVIEW PROCESS

The review process for Top Picks was similar to previous years with the exception that the selection committee meeting was held remotely for the first time. I chose a somewhat larger committee than previous years to reduce individual workload, take advantage of more specialized expertise, and select some recently graduated rising stars to give them an insight into the process. I invited half of the previous year's committee as well as the program chairs of 2020's top architecture conferences and recent Top Picks selection committee chairs. I also invited a diverse set of other researchers who regularly publish in our top conferences. The final selection committee had 42 people.

THIS YEAR'S SELECTIONS SHOWCASE A GREAT DIVERSITY OF RESEARCH TOPICS MAKING AN IMPACT IN OUR COMMUNITY. FROM TRADITIONAL MICROARCHITECTURE, TO MACHINE LEARNING, NONVOLATILE MEMORY MANAGEMENT, SECURITY, AND QUANTUM COMPUTING, OUR COMMUNITY HAS BEEN VERY BUSY REDEFINING AND REINVENTING COMPUTER ARCHITECTURE RESEARCH.

Authors were asked to prepare three-page descriptions of their articles, including a citation for a hypothetical "test of time" award their article could win in ten years. Authors submitted these three-pagers along with the original conference article. The intent of Top Picks is to choose noteworthy articles from the top computer architecture conferences, but I allowed a rather broad interpretation as our community has grown significantly. For example, this year one of the Honorable Mentions is from OSDI and another is from the IEEE Symposium on Security and Privacy.

The review process took place in two rounds. In the first round, each article was assigned four reviews. After the reviews were finished, each article was assigned a discussion lead to guide an online discussion during which each group of reviewers decided whether the article should proceed to the second round. Out of 120 submissions, 47 went on to the second round. These articles

0272-1732 © 2021 IEEE
Digital Object Identifier 10.1109/MM.2021.3074795
Date of current version 25 May 2021.

were assigned an additional five reviews, so that every article was assigned nine reviews. The vast majority of these articles actually received nine reviews; three received only eight due to unforeseen reviewer circumstances. After the second round of reviewers, there was another online discussion period led again by discussion leads. This time, articles were classified into categories to help decide which articles should be discussed. During the selection committee meeting, articles were discussed in an order determined by an aggregate of merit scores normalized by an algorithm that takes into account reviewer bias. At the beginning of each discussion, those with conflicts were moved to a Zoom break-out room and were not able to hear or see the discussion of the conflicted article. At the end of each discussion, a decision was made to make the article a "Top Pick," "Honorable Mention," or "Reject." After the decision was recorded and the shared screen was cleared of information from the discussion, conflicts were allowed to reenter the main Zoom room. According to guidelines introduced at the beginning of the meeting, a procedure was in place to upgrade Honorable Mentions or Rejects if there were fewer than 12 Top Picks or Honorable Mentions at the end of the discussion. I was advised by the *IEEE Micro* Editor-In-Chief that 12 is the absolute limit for Top Picks and another 12 for Honorable Mentions, and I wanted to award as many of these distinguished awards as possible to highlight the great research going on in the architecture community. I am gratified but not surprised that we filled all the slots for 12 Top Picks and an additional 12 Honorable Mentions.

SELECTED ARTICLES

This issue has an eclectic set of articles reflecting our diverse community.

Machine Learning

Interest in machine learning has blossomed in recent years, and architecture is on the forefront of this new revolution. One article from a large team spanning many companies including Google, Microsoft, Facebook, Intel, NVIDIA, ARM, and others introduces what has already become an industry-standard benchmark inference suite: MLPerf, addressing the need for an industry-wide, fair, and reproducible machine learning benchmarking methodology. Another article from Kyushu University, Seoul University, and Nagoya University proposes an optimized neural processing unit made with single-flux quantum technology.

Security

As industry explores more complex processor designs, inadvertent exposure of security risks remains a top concern. One article from researchers at Illinois, MIT, EPFL, and NVIDIA gives techniques for attacking the security of compressed caches, showing that compressed caches are vulnerable to side-channel attacks.

Accelerators

One article shows the potential for accelerators to improve performance in datacenters. This study from Facebook and the University of Michigan characterizes workloads at Facebook in terms of the time spent in portions that could be offloaded to accelerators, and then projects the performance improvement achievable by acceleration. Unlike previous work, the study is done in the context of increasingly common microservices. Another article from Seoul National University, Berkeley, Duke, and SungKyunKwan University looks at acceleration in the genomics domain. It presents a framework for flexible genomic data processing using SQL-like queries as a domain-specific language targeting FPGAs. Another fascinating article from the University of Wisconsin and the University of Toronto provides a novel microarchitecture for unary computing focused on matrix multiply for power-constrained environments, showing a very different way to compute with simple circuits.

Address Translation

One article from a team at Illinois and NVIDIA gives a new more efficient method for address translation for containerized workloads by deduplicating page table entries within translation-lookaside buffers and smarter page table sharing techniques.

New Technologies

Several articles describe architectural aspects of new and emerging technologies. One article from the University of California, San Diego, characterizes the first widely available nonvolatile memory technology, compares its behavior against previous less accurate models used in academia, and proposes modifications to make those models more accurate. Another article from UC Santa Barbara, Lawrence Berkeley National Labs, NIST, and the University of Maryland shows how to design logic and processors using superconducting substrates, where pulses rather than voltage levels are the fundamental physical phenomenon that carry information. Another article from Seoul National University examines the design of cryogenically cooled processors, looking at the performance benefits and power tradeoffs and how to design a processor to better leverage a cryogenic environment. Another article from Yale, Rutgers, and Brown describes adapting a traditional architecture to a new extreme domain:

HONORABLE MENTIONS

Article	Summary
"NeuMMU: Architectural Support for Efficient Address Translations in Neural Processing Units" by Hyun *et al.* (ASPLOS 2020).	This article explores the design of memory management units for emerging neural network accelerators.
"Hermes: A Fast, Fault-Tolerant and Linearizable Replication Protocol" by Katsarakis *et al.* (ASPLOS 2020)	Combining insights from cache coherence and consensus protocols, Hermes is a fault-tolerant replication protocol that achieves high performance and strong consistency.
"DSAGEN: Synthesizing Programmable Spatial Accelerators" by Weng *et al.* (ISCA 2020)	This article addresses the cost of developing specialized accelerators, proposing a framework for programmable accelerator synthesis for a decoupled spatial architecture.
"Printed Machine Learning Classifiers" by Husnain Mubarik *et al.* (MICRO 2020).	Inkjet printed circuits are an interesting new technology. This article gives a deep dive into in a synergistic application domain: classifiers.
"Agamotto: How Persistent is your Persistent Memory Application?" by Neal *et al.* (OSDI 2020)	Agamotto is a persistent memory debugging framework that uses symbolic execution to test persistent memory systems.
"Graphene: Strong yet Lightweight Row Hammer Protection" by Park (MICRO 2020)	This article introduces a technique to prevent DRAM Rowhammer attacks based on detecting frequently accessed elements. The idea features low overhead and strong security guarantees.
"Systematic Crosstalk Migitation for Superconducting Qubits via Frequency-Aware Compilation" by Ding *et al.* (MICRO 2020)	This article mitigates crosstalk in quantum computing systems with a hardware–software solution that tunes the frequency of each qubit.
"TRRespass: Exploiting the Many Sides of Target Row Refresh" by Frigo *et al.* (SP (Oakland) 2020)	This article gives a comprehensive study of Target Row Refresh, a DRAM Rowhammer mitigation technique, and shows that new DRAM chips continue to be vulnerable to Rowhammer attacks.
"SQUARE: Strategic Quantum Ancilla Reuse for Modular Quantum Programs via Cost-Effective Uncomputation" by Ding *et al.* (ISCA 2020)	This article shows it is possible to use ancilla (i.e., scratch) qubits by doing uncomputation to set the ancilla back to the original state, allowing longer programs to be run on a quantum computer.
"Leaking Information Through Cache LRU States" by Xiong and Szefer (HPCA 2020)	This article shows how to use LRU replacement state to create a cache covert channel between agents sharing a cache. It works for a variety of scenarios and pseudo-LRU designs.
"Elastic Cuckoo Page Tables: Rethinking Virtual Memory Translation for Parallelism" by Skarlatos *et al.* (ASPLOS 2020)	This article gives a novel page table design based on cuckoo hashing. Sequentially walked radix tables are replaced by a set of cuckoo tables searched in parallel. The scheme allows page tables to scale efficiently.
"The Architectural Implications of Facebook's DNN-based Personalized Recommendation" by Gupta *et al.* (HPCA 2020)	This article gives a characterization of Facebook servers running synthetic recommendation models. It introduces open source recommendation models representing a large class of services important in modern datacenters.

brain–computer interfaces. It gives optimizations to make power-efficient hardware by designing a chip for this unique application domain. Finally, an article from the University of Chicago proposes a 2.5D virtualized architecture for enabling the efficient implementation of surface codes, which are error correcting codes that enable fault-tolerant quantum computing.

THANKS

I would like to give special thanks to Elba Garza and Gino Chacon, doctoral students at Texas A&M University, who were tremendously helpful running the HotCRP and Zoom software during the selection committee meeting. I would also like to thank Emery Berger, ASPLOS 2021 co-PC-chair who shared with me scripts for parsing HotCRP output to provide files for managing conflicts of interest in Zoom; Boris Grot and Paul Gratz who acted as selection chair for papers with which I was conflicted; Kathryn McKinley for helpful suggestions based on previous experiences with Zoom PC meetings that made our meeting more enjoyable; Lizy Kurian John, *IEEE Micro* EIC, for asking me to be the selection chair and guest editor, and for her helpful advice and guidance throughout this process; and the selection committee whose hard work over the Christmas and New Year's holidays made this selection process possible. Most of all, I would like to thank the authors of the 120 submissions, whose remarkable contributions to research continue to make computer architecture an adventure for all of us.

DANIEL A. JIMÉNEZ is a Professor in the Department of Computer Science & Engineering at Texas A&M University. His research is in microarchitecture, including microarchitectural prediction and cache management. He invented perceptron-based branch prediction which is used in millions of processors. He is an NSF CAREER award winner, and a member of the MICRO, HPCA, and (as of June 2021) ISCA Halls of Fame. Contact him at djimenez@tamu.edu.

SELECTION COMMITTEE

- Nael Abu-Ghazaleh, University of California Riverside
- Abhishek Bhattacharjee, Yale University
- Marc Casas, Barcelona Supercomputing Center
- Fred Chong, University of Chicago
- Reetuparna Das, University of Michigan
- Christina Delimitrou, Cornell University
- Ronald Dreslinski, University of Michigan
- Lieven Eeckhout, Ghent University
- Hadi Esmaeilzadeh, University of California, San Diego
- Michael Ferdman, Stony Brook University
- Christopher Fletcher, University of Illinois-Urbana Champaign
- Paul Gratz, Texas A&M University
- Boris Grot, University of Edinburgh
- Engin Ipek, Qualcomm
- Akanksha Jain, Arm Research
- Aamer Jaleel, NVIDIA Research
- Jose Joao, Arm Research
- Samira Khan, University of Virginia
- Martha Kim, Columbia University
- Hyesoon Kim, Georgia Tech
- John Kim, KAIST
- Gabriel Loh, Advanced Micro Devices
- Kathryn S McKinley, Google
- Onur Mutlu, ETH Zurich
- Prashant Nair, University of British Columbia
- Gilles A. Pokam, Intel
- Moinuddin Qureshi, Georgia Tech
- Parthasarathy Ranganathan, Google
- Adrian Sampson, Cornell University
- Daniel Sanchez, MIT
- Joshua San Miguel, University of Wisconsin-Madison
- Simha Sethumadhavan, Columbia University/Chip Scan
- Sophia Shao, UC Berkeley
- Yan Solihin, University of Central Florida
- Viji Srinivasan, IBM Research
- Josep Torrellas, University of Illinois-Urbana Champaign
- Caroline Trippel, Stanford University
- Thomas F. Wenisch, University of Michigan and Google
- Lisa Wu Wills, Duke University
- Yuan Xie, University of California, Santa Barbara
- Mengjia Yan, MIT
- Jishen Zhao, University of California, San Diego