

# **2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA 2020)**

**Valencia, Spain  
30 May – 3 June 2020**

**Pages 1-555**



**IEEE Catalog Number: CFP20030-POD  
ISBN: 978-1-7281-4662-1**

**Copyright © 2020 by the Institute of Electrical and Electronics Engineers, Inc.  
All Rights Reserved**

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

***\*\*\* This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP20030-POD
ISBN (Print-On-Demand):	978-1-7281-4662-1
ISBN (Online):	978-1-7281-4661-4

**Additional Copies of This Publication Are Available From:**

Curran Associates, Inc  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: (845) 758-0400  
Fax: (845) 758-2633  
E-mail: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA) **ISCA 2020**

## Table of Contents

Message from the ISCA 2020 General Co-Chairs .....	xvi
Message from the ISCA 2020 Program Chair .....	xvii
Foreword about the New ISCA Industry Track .....	xxiii
ISCA 2020 Organizing Committee .....	xxiv
ISCA 2020 Program Committee .....	xxvi
ISCA 2020 External Review Committee .....	xxvii
ISCA 2020 External Reviewers .....	xxix

### Session 1: Industry Track

Data Compression Accelerator on IBM POWER9 and z15 Processors .....	1
<i>Bulent Abali (IBM), Bart Blaner (IBM), John Reilly (IBM), Matthias Klein (IBM), Ashutosh Mishra (IBM), Craig B. Agricola (IBM), Bedri Sendir (IBM), Alper Buyuktosunoglu (IBM), Christian Jacobi (IBM), William J. Starke (IBM), Haren Myneni (IBM), and Charlie Wang (IBM)</i>	
High-Performance Deep-Learning Coprocessor Integrated into x86 SoC with Server-Class CPUs ....	15
<i>Glenn Henry (Centaur Technology), Parviz Palangpour (Centaur Technology), Michael Thomson (Centaur Technology), J Scott Gardner (Advantage Engineering LLC), Bryce Arden (Centaur Technology), Jim Donahue (Centaur Technology), Kimble Houck (Centaur Technology), Jonathan Johnson (Centaur Technology), Kyle O'Brien (Centaur Technology), Scott Petersen (Centaur Technology), Benjamin Seroussi (Centaur Technology), and Tyler Walker (Centaur Technology)</i>	
The IBM z15 High Frequency Mainframe Branch Predictor .....	27
<i>Narasimha Adiga (IBM), James Bonanno (IBM), Adam Collura (IBM), Matthias Heizmann (IBM), Brian R. Prasky (IBM), and Anthony Saporito (IBM)</i>	
Evolution of the Samsung Exynos CPU Microarchitecture .....	40
<i>Brian Grayson (SiFive), Jeff Rupley (Centaur), Gerald Zuraski Jr. (Independent Consultant), Eric Quinnell (ARM), Daniel Jimenez (Texas A&amp;M University), Tarun Nakra (AMD), Paul Kitchin (Nuvia), Ryan Hensley (Goodix), Edward Brekelbaum (SiFive), Vikas Sinha (Nuvia), and Ankit Ghiya (ARM)</i>	

Xuantie-910: A Commercial Multi-Core 12-Stage Pipeline Out-of-Order 64-bit High Performance RISC-V Processor with Vector Extension .52.....  
*Chen Chen (Alibaba Cloud), Xiaoyan Xiang (Alibaba Cloud), Chang Liu (Alibaba Cloud), Yunhai Shang (Alibaba Cloud), Ren Guo (Alibaba Cloud), Dongqi Liu (Alibaba Cloud), Yimin Lu (Alibaba Cloud), Ziyi Hao (Alibaba Cloud), Jiahui Luo (Alibaba Cloud), Zhijian Chen (Alibaba Cloud), Chunqiang Li (Alibaba Cloud), Yu Pu (Alibaba Cloud), Jianyi Meng (Alibaba Cloud), Xiaolang Yan (Alibaba Cloud), Yuan Xie (Alibaba Cloud), and Xiaoning Qi (Alibaba Cloud)*

## Session 2A: Microarchitecture

Divide and Conquer Frontend Bottleneck .65.....  
*Ali Ansari (Sharif University of Technology), Pejman Lotfi-Kamran (Institute for Research in Fundamental Sciences (IPM)), and Hamid Sarbazi-Azad (Sharif University of Technology, Institute for Research in Fundamental Sciences (IPM))*

Focused Value Prediction .79.....  
*Sumeet Bandishte (Intel Labs, India), Jayesh Gaur (Intel Labs, India), Zeev Sperber (Intel Corporation, Israel), Lihu Rappoport (Intel Corporation, Israel), Adi Yoaz (Intel Corporation, Israel), and Sreenivas Subramoney (Intel Labs, India)*

Auto-Predication of Critical Branches .92.....  
*Adarsh Chauhan (Intel Labs, India), Jayesh Gaur (Intel Labs, India), Zeev Sperber (Intel Corporation, Israel), Franck Sala (Intel Corporation, Israel), Lihu Rappoport (Intel Corporation, Israel), Adi Yoaz (Intel Corporation, Israel), and Sreenivas Subramoney (Intel Labs, India)*

Slipstream Processors Revisited: Exploiting Branch Sets .105.....  
*Vinesh Srinivasan (North Carolina State University), Rangeen Basu Roy Chowdhury (Intel Corporation), and Eric Rotenberg (North Carolina State University)*

Bouquet of Instruction Pointers: Instruction Pointer Classifier-Based Spatial Hardware Prefetching .118.....  
*Samuel Pakalapati (Intel Technology Pvt. Ltd. and BITS Pilani) and Biswabandan Panda (Indian Institute of Technology Kanpur)*

MuonTrap: Preventing Cross-Domain Spectre-Like Attacks by Capturing Speculative State .132....  
*Sam Ainsworth (University of Cambridge, UK) and Timothy M. Jones (University of Cambridge, UK)*

Think Fast: A Tensor Streaming Processor (TSP) for Accelerating Deep Learning Workloads .....	145
<i>Dennis Abts (Groq, Inc.), Jonathan Ross (Groq, Inc.), Jonathan Sparling (Groq, Inc.), Mark Wong-VanHaren (Groq, Inc.), Max Baker (Groq, Inc.), Tom Hawkins (Groq, Inc.), Andrew Bell (Groq, Inc.), John Thompson (Groq, Inc.), Temesghen Kahsai (Groq, Inc.), Garrin Kimmell (Groq, Inc.), Jennifer Hwang (Groq, Inc.), Rebekah Leslie-Hurd (Groq, Inc.), Michael Bye (Groq, Inc.), E. R. Creswick (Groq, Inc.), Matthew Boyd (Groq, Inc.), Mahitha Venigalla (Groq, Inc.), Evan Laforge (Groq, Inc.), Jon Purdy (Groq, Inc.), Purushotham Kamath (Groq, Inc.), Dinesh Maheshwari (Groq, Inc.), Michael Beidler (Groq, Inc.), Geert Rosseel (Groq, Inc.), Omar Ahmad (Groq, Inc.), Gleb Gagarin (Groq, Inc.), Richard Czekalski (Groq, Inc.), Ashay Rane (Groq, Inc.), Sahil Parmar (Groq, Inc.), Jeff Werner (Groq, Inc.), Jim Sproch (Groq, Inc.), Adrian Macias (Groq, Inc.), and Brian Kurtz (Groq, Inc.)</i>	

## Session 2B: Paralellism/IoT and Mobile

T4: Compiling Sequential Code for Effective Speculative Parallelization in Hardware .....	159
<i>Victor A. Ying (Massachusetts Institute of Technology, USA), Mark C. Jeffrey (University of Toronto, Canada), and Daniel Sanchez (Massachusetts Institute of Technology, USA)</i>	
Efficiently Supporting Dynamic Task Parallelism on Heterogeneous Cache-Coherent Systems .....	173
<i>Moyang Wang (Cornell University, USA), Tuan Ta (Cornell University, USA), Lin Cheng (Cornell University, USA), and Christopher Batten (Cornell University, USA)</i>	
Flick: Fast and Lightweight ISA-Crossing Call for Heterogeneous-ISA Environments .....	187
<i>Shengsun Cho (Stony Brook University, USA), Han Chen (Stony Brook University, USA), Sergey Madaminov (Stony Brook University, USA), Michael Ferdman (Stony Brook University, USA), and Peter Milder (Stony Brook University, USA)</i>	
The NeBuLa RPC-Optimized Architecture .....	199
<i>Mark Sutherland (EcoCloud, EPFL), Siddharth Gupta (EcoCloud, EPFL), Babak Falsafi (EcoCloud, EPFL), Virendra Marathe (Oracle Labs), Dionisios Pneumatikatos (National Technical University of Athens), and Alexandros Daglis (Georgia Institute of Technology)</i>	
Printed Microprocessors .....	213
<i>Nathaniel Bleier (UIUC), Muhammad Husnain Mubarak (UIUC), Farhan Rasheed (KIT), Jasmin Aghassi-Hagmann (KIT), Mehdi B Tahoori (KIT), and Rakesh Kumar (UIUC)</i>	
SysScale: Utilizing Holistic Multi-domain DVFS to Improve the Energy Efficiency of Mobile Processors .....	227
<i>Jawad Haj-Yahya (ETH Zürich), Mohammed Alser (ETH Zürich), Jeremie Kim (ETH Zürich), A. Giray Yağlıkçı (ETH Zürich), Nandita Vijaykumar (ETH Zürich; Intel; University of Toronto), Efraim Rotem (Intel), and Onur Mutlu (ETH Zürich)</i>	

Déjà View: Spatio-Temporal Compute Reuse for Energy-Efficient 360° VR Video Streaming .....	241
<i>Shulin Zhao (The Pennsylvania State University), Haibo Zhang (The Pennsylvania State University), Sandeepa Bhuyan (The Pennsylvania State University), Cyan Mishra (The Pennsylvania State University), Ziyu Ying (The Pennsylvania State University), Mahmut Kandemir (The Pennsylvania State University), Anand Sivasubramaniam (The Pennsylvania State University), and Chita Das (The Pennsylvania State University)</i>	

## Session 3A: Accelerator-Based/Application-Specific Archs.

Genesis: A Hardware Acceleration Framework for Genomic Data Analysis .....	254
<i>Tae Jun Ham (Seoul National University), David Bruns-Smith (University of California, Berkeley), Brendan Sweeney (University of California, Berkeley), Yejin Lee (Seoul National University), Seong Hoon Seo (Seoul National University), U Gyeong Song (Seoul National University), Young H. Oh (Sungkyunkwan University), Krste Asanovic (University of California, Berkeley), Jae W. Lee (Seoul National University), and Lisa Wu Wills (Duke University)</i>	
DSAGEN: Synthesizing Programmable Spatial Accelerators .....	268
<i>Jian Weng (University of California, Los Angeles, USA), Sihao Liu (University of California, Los Angeles, USA), Vidushi Dadu (University of California, Los Angeles, USA), Zhengrong Wang (University of California, Los Angeles, USA), Preyas Shah (Simple Machines Inc., USA), and Tony Nowatzki (University of California, Los Angeles, USA)</i>	
Bonsai: High-Performance Adaptive Merge Tree Sorting .....	282
<i>Nikola Samardzic (University of California, Los Angeles), Weikang Qiao (University of California, Los Angeles), Vaibhav Aggarwal (University of California, Los Angeles), Mau-Chung Frank Chang (University of California, Los Angeles), and Jason Cong (University of California, Los Angeles)</i>	
SOFF: An OpenCL High-Level Synthesis Framework for FPGAs .....	295
<i>Gangwon Jo (ManyCoreSoft, Korea), Heehoon Kim (Seoul National University, Korea), Jeesoo Lee (Seoul National University, Korea), and Jaejin Lee (Seoul National University, Korea)</i>	
Gorgon: Accelerating Machine Learning from Relational Data .....	309
<i>Matthew Vilim (Stanford University), Alexander Rucker (Stanford University), Yaqi Zhang (Stanford University), Sophia Liu (Stanford University), and Kunle Olukotun (Stanford University)</i>	
A Specialized Architecture for Object Serialization with Applications to Big Data Analytics .....	322
<i>Jaeyoung Jang (Sungkyunkwan University, Korea), Sung Jun Jung (Seoul National University, Korea), Sunmin Jeong (Seoul National University, Korea), Jun Heo (Seoul National University, Korea), Hoon Shin (Seoul National University, Korea), Tae Jun Ham (Seoul National University, Korea), and Jae W. Lee (Seoul National University, Korea)</i>	

CryoCore: A Fast and Dense Processor Architecture for Cryogenic Computing .335.....  
*Ilkwon Byun (Seoul National University, Republic of Korea), Dongmoon Min (Seoul National University, Republic of Korea), Gyu-hyeon Lee (Seoul National University, Republic of Korea), Seongmin Na (Seoul National University, Republic of Korea), and Jangwoo Kim (Seoul National University, Republic of Korea)*

## **Session 3B: Non-Traditional Computing / Graph Processing**

SpinalFlow: An Architecture and Dataflow Tailored for Spiking Neural Networks .349.....  
*Surya Narayanan (University of Utah, USA), Karl Taht (University of Utah, USA), Rajeev Balasubramonian (University of Utah, USA), Édouard Giacomin (University of Utah, USA), and Pierre-Emmanuel Gaillardon (University of Utah, USA)*

NEBULA: A Neuromorphic Spin-Based Ultra-Low Power Architecture for SNNs and ANNs .363..  
*Sonali Singh (The Pennsylvania State University, USA), Anup Sarma (The Pennsylvania State University, USA), Nicholas Jao (The Pennsylvania State University, USA), Ashutosh Pattnaik (The Pennsylvania State University, USA), Sen Lu (The Pennsylvania State University, USA), Kezhou Yang (The Pennsylvania State University, USA), Abhronil Sengupta (The Pennsylvania State University, USA), Vijaykrishnan Narayanan (The Pennsylvania State University, USA), and Chita R. Das (The Pennsylvania State University, USA)*

uGEMM: Unary Computing Architecture for GEMM Applications .377.....  
*Di Wu (University of Wisconsin–Madison), Jingjie Li (University of Wisconsin–Madison), Ruokai Yin (University of Wisconsin–Madison), Hsuan Hsiao (University of Toronto), Younghyun Kim (University of Wisconsin–Madison), and Joshua San Miguel (University of Wisconsin–Madison)*

Hardware-Software Co-Design for Brain-Computer Interfaces .391.....  
*Ioannis Karageorgos (Yale University, USA), Karthik Sriram (Yale University, USA), Ján Veselý (Rutgers University, USA and Yale University, USA), Michael Wu (Rutgers University, USA), Marc Powell (Brown University, USA), David Borton (Brown University, USA), Rajit Manohar (Yale University, USA), and Abhishek Bhattacharjee (Yale University, USA)*

Heat to Power: Thermal Energy Harvesting and Recycling for Warm Water-Cooled Datacenters .405  
*Xinhui Zhu (Huazhong University of Science and Technology, China), Weixiang Jiang (Huazhong University of Science and Technology, China), Fangming Liu (Huazhong University of Science and Technology, China), Qixia Zhang (Huazhong University of Science and Technology, China), Li Pan (Huazhong University of Science and Technology, China), Qiong Chen (Huazhong University of Science and Technology, China), and Ziyang Jia (Huazhong University of Science and Technology, China)*

GraphABCD: Scaling Out Graph Analytics with Asynchronous Block Coordinate Descent 419.....  
*Yifan Yang (Tsinghua University, China), Zhaoshi Li (Tsinghua University, China), Yangdong Deng (Tsinghua University, China), Zhiwei Liu (Tsinghua University, China), Shouyi Yin (Tsinghua University, China), Shaojun Wei (Tsinghua University, China), and Leibo Liu (Tsinghua University, China)*

GaaS-X: Graph Analytics Accelerator Supporting Sparse Data Representation Using Crossbar Architectures 433.....  
*Nagadastagiri Challapalle (The Pennsylvania State University), Sahithi Rampalli (The Pennsylvania State University), Linghao Song (Duke University), Nandhini Chandramoorthy (IBM Research), Karthik Swaminathan (IBM Research), John Sampson (The Pennsylvania State University), Yiran Chen (Duke University), and Vijaykrishnan Narayanan (The Pennsylvania State University)*

## Session 4A: Performance Evaluation / Virtualization

MLPerf Inference Benchmark 446.....  
*Vijay Janapa Reddi (Harvard University), Christine Cheng (Intel), David Kanter (Real World Insights), Peter Mattson (Google), Guenther Schmuelling (Microsoft), Carole-Jean Wu (Facebook), Brian Anderson (Google), Maximilien Breughe (NVIDIA), Mark Charlebois (Qualcomm), William Chou (Qualcomm), Ramesh Chukka (Intel), Cody Coleman (Stanford), Sam David (Myrtle.ai), Pan Deng (Tencent), Greg Diamos (Landing AI), Jared Duke (Google), Dave Fick (Mythic), Scott Gardner (Advantage Engineering), Itay Hubara (Habana Labs), Sachin Idgunji (NVIDIA), Thomas Jablin (Google), Jeff Jiao (Alibaba T-Head), Tom St. John (Tesla), Pankaj Kanwar (Google), David Lee (MediaTek), Jeffery Liao (OPPO), Anton Lokhmotov (dividiti), Francisco Massa (Facebook), Peng Meng (Tencent), Paulius Micikevicius (NVIDIA), Colin Osborne (ARM), Gennady Pekhimenko (University of Toronto & Vector Institute), Arun Tejusve Raghunath Rajan (Intel), Dilip Sequeira (NVIDIA), Ashish Sirasao (Xilinx), Fei Sun (Alibaba), Hanlin Tang (Intel), Michael Thomson (Centaur Technology), Frank Wei (Alibaba Cloud), Ephrem Wu (Xilinx), Lingjie Xu (Biren Research), Koichi Yamada (Intel), Bing Yu (Facebook), George Yuan (NVIDIA), Aaron Zhong (Alibaba T-Head), Peizhao Zhang (Facebook), and Yuchen Zhou (General Motors)*

Mocktails: Capturing the Memory Behaviour of Proprietary Mobile Architectures 460.....  
*Mario Badr (University of Toronto, Canada), Carlo Delconte (Arm, United Kingdom), Isak Edo (University of Toronto, Canada), Radhika Jagtap (Arm, United Kingdom), Matteo Andreozzi (Arm, United Kingdom), and Natalie Enright Jerger (University of Toronto, Canada)*

Accel-Sim: An Extensible Simulation Framework for Validated GPU Modeling 473.....  
*Mahmoud Khairy (Purdue University, US), Zhesheng Shen (Purdue University, US), Tor M. Aamodt (University of British Columbia, Canada), and Timothy G. Rogers (Purdue University, US)*

HyperTRIO: Hyper-Tenant Translation of I/O Addresses 487.....  
*Alexey Lavrov (Princeton University, USA) and David Wentzclaff (Princeton University, USA)*



BabelFish: Fusing Address Translations for Containers .501.....  
*Dimitrios Skarlatos (University of Illinois at Urbana-Champaign), Umur Darbaz (University of Illinois at Urbana-Champaign), Bhargava Gopireddy (University of Illinois at Urbana-Champaign), Nam Sung Kim (University of Illinois at Urbana-Champaign), and Josep Torrellas (University of Illinois at Urbana-Champaign)*

Enhancing and Exploiting Contiguity for Fast Memory Virtualization .515.....  
*Chloe Alverti (National Technical University of Athens), Stratos Psomadakis (National Technical University of Athens), Vasileios Karakostas (National Technical University of Athens), Jayneel Gandhi (VMware Research), Konstantinos Nikas (National Technical University of Athens), Georgios Goumas (National Technical University of Athens), and Nectarios Koziris (National Technical University of Athens)*

## Session 4B: DRAM / Quantum Computing

Architecting Noisy Intermediate-Scale Trapped Ion Quantum Computers .529.....  
*Prakash Murali (Princeton University, USA), Dripto M. Debroy (Duke University, USA), Kenneth R. Brown (Duke University, USA), and Margaret Martonosi (Princeton University, USA)*

AccQOC : Accelerating Quantum Optimal Control Based Pulse Generation .543.....  
*Jinglei Cheng (University of Southern California), Haoqing Deng (University of Southern California), and Xuehai Qian (University of Southern California)*

NISQ+: Boosting Quantum Computing Power by Approximating Quantum Error Correction .556..  
*Adam Holmes (The University of Chicago), Mohammad Reza Jokar (The University of Chicago), Ghasem Pasandi (The University of Southern California), Yongshan Ding (The University of Chicago), Massoud Pedram (The University of Southern California), and Frederic T. Chong (The University of Chicago)*

SQUARE: Strategic Quantum Ancilla Reuse for Modular Quantum Programs via Cost-Effective Uncomputation .570.....  
*Yongshan Ding (University of Chicago, USA), Xin-Chuan Wu (University of Chicago, USA), Adam Holmes (Intel Corporation, USA), Ash Wiseth (University of Chicago, USA), Diana Franklin (University of Chicago, USA), Margaret Martonosi (Princeton University, USA), and Frederic Chong (University of Chicago, USA)*

## Session 5A: Non-Volatile / Persistent Memory and Storage

HOOP: Efficient Hardware-Assisted Out-of-Place Update for Non-Volatile Memory .584.....  
*Miao Cai (Nanjing University, China), Chance Coats (University of Illinois at Urbana-Champaign, USA), and Jian Huang (University of Illinois at Urbana-Champaign, USA)*

Lelantus: Fine-Granularity Copy-On-Write Operations for Secure Non-Volatile Memories .597.....  
*Jian Zhou (University of Central Florida), Amro Awad (University of Central Florida), and Jun Wang (University of Central Florida)*

MorLog: Morphable Hardware Logging for Atomic Persistence in Non-Volatile Main Memory .....	610
<i>Xueliang Wei (Huazhong University of Science and Technology, China), Dan Feng (Huazhong University of Science and Technology, China), Wei Tong (Huazhong University of Science and Technology, China), Jingning Liu (Huazhong University of Science and Technology, China), and Liuqing Ye (Huazhong University of Science and Technology, China)</i>	
Tvarak: Software-Managed Hardware Offload for Redundancy in Direct-Access NVM Storage .....	624
<i>Rajat Kateja (Carnegie Mellon University), Nathan Beckmann (Carnegie Mellon University), and Gregory R. Ganger (Carnegie Mellon University)</i>	
Revisiting RowHammer: An Experimental Analysis of Modern DRAM Devices and Mitigation Techniques .....	638
<i>Jeremie Kim (ETH Zurich; Carnegie Mellon University), Minesh Patel (ETH Zurich), A. Giray Yaglikci (ETH Zurich), Hasan Hassan (ETH Zurich), Roknoddin Azizi (ETH Zurich), Lois Orosa (ETH Zurich), and Onur Mutlu (ETH Zurich; Carnegie Mellon University)</i>	
Relaxed Persist Ordering Using Strand Persistency .....	652
<i>Vaibhav Gogte (University of Michigan, USA), William Wang (Arm Research, UK), Stephan Diestelhorst (Xilinx DCG, UK), Peter M. Chen (University of Michigan, USA), Satish Narayanasamy (University of Michigan, USA), and Thomas F. Wenisch (University of Michigan, USA)</i>	
CLR-DRAM: A Low-Cost DRAM Architecture Enabling Dynamic Capacity-Latency Trade-Off .....	666
<i>Haocong Luo (ETH Zürich; ShanghaiTech University, China), Taha Shahroodi (ETH Zürich), Hasan Hassan (ETH Zürich), Minesh Patel (ETH Zürich), A. Giray Yağlıkçı (ETH Zürich), Lois Orosa (ETH Zürich), Jisung Park (ETH Zürich), and Onur Mutlu (ETH Zürich)</i>	
Hardware-Based Domain Virtualization for Intra-Process Isolation of Persistent Memory Objects .....	680
<i>Yuanchao Xu (North Carolina State University, USA), ChenCheng Ye (Huazhong University of Science and Technology, China), Yan Solihin (University of Central Florida, USA), and Xipeng Shen (North Carolina State University, USA)</i>	
Check-In: In-Storage Checkpointing for Key-Value Store System Leveraging Flash-Based SSDs .....	693
<i>Joohyeong Yoon (Yonsei University and Samsung Electronics), Won Seob Jeong (Yonsei University and Samsung Electronics), and Won Woo Ro (Yonsei University)</i>	

## Session 5B: Security

Speculative Data-Oblivious Execution: Mobilizing Safe Prediction for Safe and Efficient Speculative Execution .....	707
<i>Jiyong Yu (University of Illinois at Urbana-Champaign), Namrata Mantri (University of Illinois at Urbana-Champaign), Josep Torrellas (University of Illinois at Urbana-Champaign), Adam Morrison (Tel Aviv University), and Christopher W. Fletcher (University of Illinois at Urbana-Champaign)</i>	
Packet Chasing: Spying on Network Packets over a Cache Side-Channel .....	721
<i>Mohammadkazem Taram (University of California San Diego), Ashish Venkat (University of Virginia), and Dean Tullsen (University of California San Diego)</i>	

Compact Leakage-Free Support for Integrity and Reliability .735.....	
	<i>Meysam Taassori (University of Utah, Salt Lake City, USA), Rajeev Balasubramonian (University of Utah), Siddhartha Chhabra (Intel), Alaa Alameldeen (Intel), Manjula Peddireddy (Intel), Rajat Agarwal (Intel), and Ryan Stutsman (University of Utah)</i>
A Bus Authentication and Anti-Probing Architecture Extending Hardware Trusted Computing Base Off CPU Chips and Beyond .749.....	
	<i>Zhenyu Xu (University of Rhode Island), Thomas Mauldin (University of Rhode Island), Zheyi Yao (University of Rhode Island), Shuyi Pei (University of Rhode Island), Tao Wei (University of Rhode Island), and Qing Yang (University of Rhode Island)</i>
CHEX86: Context-Sensitive Enforcement of Memory Safety via Microcode-Enabled Capabilities .762	
	<i>Rasool Sharifi (University of Virginia, USA) and Ashish Venkat (University of Virginia, USA)</i>
Nested Enclave: Supporting Fine-Grained Hierarchical Isolation with SGX .776.....	
	<i>Joongun Park (KAIST), Naegyong Kang (KAIST), Taehoon Kim (KAIST), Youngjin Kwon (KAIST), and Jaehyuk Huh (KAIST)</i>

## Session 6A: Near-Data Processing / Processing-in-Memory

RecNMP: Accelerating Personalized Recommendation with Near-Memory Processing .790.....	
	<i>Liu Ke (Facebook/Washington University in St Louis), Udit Gupta (Harvard), Benjamin Youngjae Cho (Facebook/The University of Texas at Austin), David David (Harvard), Vikas Chandra (Facebook), Utku Diril (Facebook), Amin Firoozshahian (Facebook), Kim Hazelwood (Facebook), Bill Jia (Facebook), Hsien-Hsin Sean Lee (Facebook), Meng Li (Facebook), Bert Maher (Facebook), Dheevatsa Mudigere (Facebook), Maxim Naumov (Facebook), Martin Schatz (Facebook), Mikhail Smelyanskiy (Facebook), Xiaodong Wang (Facebook), Brandon Reagen (Facebook), Carole-Jean Wu (Facebook), Mark Hempstead (Facebook / Tufts University), and Xuan Zhang (Facebook/Washington University in St. Louis)</i>
iPIM: Programmable In-Memory Image Processing Accelerator Using Near-Bank Architecture .804	
	<i>Peng Gu (University of California, Santa Barbara), Xinfeng Xie (University of California, Santa Barbara), Yufei Ding (University of California, Santa Barbara), Guoyang Chen (Alibaba Cloud Infrastructure), Weifeng Zhang (Alibaba Cloud Infrastructure), Dimin Niu (Alibaba DAMO Academy), and Yuan Xie (University of California, Santa Barbara)</i>
Near Data Acceleration with Concurrent Host Access .818.....	
	<i>Benjamin Cho (The University of Texas at Austin), Yongkee Kwon (The University of Texas at Austin), Sangkug Lym (The University of Texas at Austin), and Mattan Erez (The University of Texas at Austin)</i>
TIMELY: Pushing Data Movements and Interfaces in PIM Accelerators Towards Local and in Time Domain .832.....	
	<i>Weitao Li (Rice University; University of California, Santa Barbara), Pengfei Xu (Rice University), Yang Zhao (Rice University), Haitong Li (Stanford University), Yuan Xie (University of California, Santa Barbara), and Yingyan Lin (Rice University)</i>

- Hyper-AP: Enhancing Associative Processing through a Full-Stack Optimization .846.....  
*Yue Zha (University of Pennsylvania, USA) and Jing Li (University of Pennsylvania, USA)*
- JPEG-ACT: Accelerating Deep Learning via Transform-Based Lossy Compression .860.....  
*R. David Evans (University of British Columbia, Canada), Lufei Liu (University of British Columbia, Canada), and Tor M. Aamodt (University of British Columbia, Canada)*

## Session 6B: Coherence, Consistency, and Memory

- TransForm: Formally Specifying Transistency Models and Synthesizing Enhanced Litmus Tests .874  
*Naorin Hossain (Princeton University, USA), Caroline Trippel (Stanford University, USA), and Margaret Martonosi (Princeton University, USA)*
- HieraGen: Automated Generation of Concurrent, Hierarchical Cache Coherence Protocols .888.....  
*Nicolai Oswald (The University of Edinburgh), Vijay Nagarajan (The University of Edinburgh), and Daniel J. Sorin (Duke University)*
- Tailored Page Sizes .900.....  
*Faruk Gwoenilir (The University of Texas at Austin) and Yale N. Patt (The University of Texas at Austin)*
- Perforated Page: Supporting Fragmented Memory Allocation for Large Pages .913.....  
*Chang Hyun Park (Uppsala University, Sweden), Sanghoon Cha (KAIST, Republic of Korea), Bokyeong Kim (KAIST, Republic of Korea), Youngjin Kwon (KAIST, Republic of Korea), David Black-Schaffer (Uppsala University, Sweden), and Jaehyuk Huh (KAIST, Republic of Korea)*
- Buddy Compression: Enabling Larger Memory for Deep Learning and HPC Workloads on GPUs .926  
*Esha Choukse (Microsoft), Michael Sullivan (NVIDIA), Mike O'Connor (NVIDIA/University of Texas at Austin), Mattan Erez (University of Texas at Austin), Jeff Pool (NVIDIA), David Nellans (NVIDIA), and Stephen Keckler (NVIDIA)*

## Session 7A: Architectural Support for Machine Learning

- A Multi-Neural Network Acceleration Architecture .940.....  
*Eunjin Baek (Seoul National University, Republic of Korea), Dongup Kwon (Seoul National University, Republic of Korea), and Jangwoo Kim (Seoul National University, Republic of Korea)*
- SmartExchange: Trading Higher-Cost Memory Storage/Access for Lower-Cost Computation .954...  
*Yang Zhao (Rice University, USA), Xiaohan Chen (Texas A&M University, USA), Yue Wang (Rice University, USA), Chaojian Li (Rice University, USA), Haoran You (Rice University, USA), Yonggan Fu (Rice University, USA), Yuan Xie (University of California, Santa Barbara, USA), Zhangyang Wang (Texas A&M University, USA), and Yingyan Lin (Rice University, USA)*

Centaur: A Chiplet-Based, Hybrid Sparse-Dense Accelerator for Personalized Recommendations.968	
	<i>Ranggi Hwang (Korea Advanced Institute of Science and Technology, South Korea), Taehun Kim (Korea Advanced Institute of Science and Technology, South Korea), Youngeun Kwon (Korea Advanced Institute of Science and Technology, South Korea), and Minsoo Rhu (Korea Advanced Institute of Science and Technology, South Korea)</i>
DeepRecSys: A System for Optimizing End-to-End At-scale Neural Recommendation Inference .982	
	<i>Udit Gupta (Harvard University/Facebook), Samuel Hsia (Harvard University), Vikram Saraph (Facebook), Xiaodong Wang (Facebook), Brandon Reagen (Facebook), Gu-Yeon Wei (Harvard University), Hsien-Hsin Sean Lee (Facebook), David Brooks (Harvard University/Facebook), and Carole-Jean Wu (Facebook)</i>
An In-network Architecture for Accelerating Shared-Memory Multiprocessor Collectives .996.....	
	<i>Benjamin Klenk (NVIDIA), Nan Jiang (NVIDIA), Greg Thorson (NVIDIA), and Larry Dennison (NVIDIA)</i>
DRQ: Dynamic Region-Based Quantization for Deep Neural Network Acceleration .1010.....	
	<i>Zhuoran Song (Shanghai Jiao Tong University), Bangqi Fu (Shanghai Jiao Tong University), Feiyang Wu (Shanghai Jiao Tong University), Zhaoming Jiang (Shanghai Jiao Tong University), Li Jiang (Shanghai Jiao Tong University), Naifeng Jing (Shanghai Jiao Tong University), and Xiaoyao Liang (Shanghai Jiao Tong University)</i>

## Session 7B: GPUs / Memory

Independent Forward Progress of Work-Groups .1022.....	
	<i>Alexandru Duțu (AMD Research), Matthew D. Sinclair (AMD Research, University of Wisconsin-Madison), Bradford M. Beckmann (AMD Research), David A. Wood (AMD Research, University of Wisconsin-Madison), and Marcus Chow (AMD Research, University of California-Riverside)</i>
ScoRD: A Scoped Race Detector for GPUs .1036.....	
	<i>Aditya Kamath (Indian Institute of Science, India), Alvin George (Indian Institute of Science, India), and Arkaprava Basu (Indian Institute of Science, India)</i>
The Virtual Block Interface: A Flexible Alternative to the Conventional Virtual Memory Framework .1050.....	
	<i>Nastaran Hajinazar (Simon Fraser University , ETH Zürich), Pratyush Patel (University of Washington), Minesh Patel (ETH Zürich), Konstantinos Kanellopoulos (ETH Zürich), Saugata Ghose (Carnegie Mellon University), Rachata Ausavarungnirun (King Mongkut’s University of Technology North Bangkok), Geraldo Francisco de Oliveira Junior (ETH Zürich), Jonathan Appavoo (Boston University), Vivek Seshadri (Microsoft Research India), and Onur Mutlu (ETH Zürich)</i>
ZnG: Architecting GPU Multi-Processors with New Flash for Scalable Data Analysis .1064.....	
	<i>Jie Zhang (KAIST) and Myoungsoo Jung (KAIST)</i>

Commutative Data Reordering: A New Technique to Reduce Data Movement Energy on Sparse Inference Workloads .....	1076
<i>Ben Feinberg (Sandia National Laboratories, USA), Benjamin C. Heyman (University of Rochester, USA), Darya Mikhailenko (University of Rochester, USA), Ryan Wong (University of Rochester, USA), An C. Ho (University of Rochester, USA), and Engin Ipek (University of Rochester, USA)</i>	
Echo: Compiler-based GPU Memory Footprint Reduction for LSTM RNN Training .....	1089
<i>Bojian Zheng (University of Toronto, Vector Institute), Nandita Vijaykumar (University of Toronto, Intel), and Gennady Pekhimenko (University of Toronto, Vector Institute)</i>	
A Case for Hardware-Based Demand Paging .....	1103
<i>Gyusun Lee (Sungkyunkwan University, South Korea), Wenjing Jin (Seoul National University, South Korea), Wonsuk Song (Sungkyunkwan University, South Korea), Jeonghun Gong (Seoul National University, South Korea), Jonghyun Bae (Seoul National University, South Korea), Tae Jun Ham (Seoul National University, South Korea), Jae W. Lee (Seoul National University, South Korea), and Jinkyu Jeong (Sungkyunkwan University, South Korea)</i>	
<b>Author Index .....</b>	<b>1117</b>