

2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO 2022)

**Chicago, Illinois, USA
1-5 October 2022**

Pages 1-726



**IEEE Catalog Number: CFP22071-POD
ISBN: 978-1-6654-7428-3**

**Copyright © 2022 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:	CFP22071-POD
ISBN (Print-On-Demand):	978-1-6654-7428-3
ISBN (Online):	978-1-6654-6272-3

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
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO) **MICRO 2022**

Table of Contents

Message from the MICRO 2022 General Chairs	xx
Message from the MICRO 2022 Program Chairs	xxii
Organizing Committee	xxx
Technical Program Committee	xxxi
External Review Committee	xxxiv
Additional External Reviewers	xxxvii
Steering Committee	xxxviii
MICRO-55 Sponsors	xxxix
Keynotes	xl
Panel	xlii

Best Papers Session

Hermes: Accelerating Long-Latency Load Requests via Perceptron-Based Off-Chip Load Prediction	1
<i>Rahul Bera (ETH Zürich, Switzerland), Konstantinos Kanellopoulos (ETH Zürich, Switzerland), Shankar Balachandran (Intel Processor Architecture Research Lab, Intel Labs, India), David Novo (LIRMM, Univ. Montpellier, CNRS), Ataberk Olgun (ETH Zürich), Mohammad Sadrosadati (ETH Zürich), and Onur Mutlu (ETH Zürich, Switzerland)</i>	
Whisper: Profile-Guided Branch Misprediction Elimination for Data Center Applications	19
<i>Tanvir Ahmed Khan (University of Michigan, USA), Muhammed Ugur (University of Michigan, USA), Krishnendra Nathella (ARM, USA), Dam Sunwoo (ARM, USA), Heiner Litz (University of California, Santa Cruz, USA), Daniel A. Jiménez (Texas A&M University, USA), and Baris Kasikci (University of Michigan, USA)</i>	
OverGen: Improving FPGA Usability through Domain-Specific Overlay Generation	35
<i>Sihao Liu (University of California, Los Angeles), Jian Weng (University of California, Los Angeles), Dylan Kupsh (University of California, Los Angeles), Atefeh Sohrabizadeh (University of California, Los Angeles), Zhengrong Wang (University of California, Los Angeles), Licheng Guo (University of California, Los Angeles), Jiuyang Liu (Institute of Software, Chinese Academy of Sciences), Maxim Zhulin (University of California, Los Angeles), Rishabh Mani (University of California, Los Angeles), Lucheng Zhang (Institute of Software, Chinese Academy of Sciences), Jason Cong (University of California, Los Angeles), and Tony Nowatzki (University of California, Los Angeles)</i>	

Cambricon-P: A Bitflow Architecture for Arbitrary Precision Computing	57
<i>Yifan Hao (SKL of Processors, Institute of Computing Technology, CAS, China; Cambricon Technologies Co., Ltd., China), Yongwei Zhao (SKL of Processors, Institute of Computing Technology, CAS, China; Cambricon Technologies Co., Ltd., China), Chenxiao Liu (SKL of Processors, Institute of Computing Technology, CAS, China; Cambricon Technologies Co., Ltd., China; University of Chinese Academy of Science, China), Zidong Du (SKL of Processors, Institute of Computing Technology, CAS, China; Cambricon Technologies Co., Ltd., China), Shuyao Cheng (SKL of Processors, Institute of Computing Technology, CAS, China; Cambricon Technologies Co., Ltd., China; University of Chinese Academy of Science, China), Xiaqing Li (SKL of Processors, Institute of Computing Technology, CAS, China), Xing Hu (SKL of Processors, Institute of Computing Technology, CAS, China), Qi Guo (SKL of Processors, Institute of Computing Technology, CAS, China), Zhiwei Xu (SKL of Processors, Institute of Computing Technology, CAS, China; University of Chinese Academy of Science, China), and Tianshi Chen (Cambricon Technologies Co., Ltd., China)</i>	

Session 1A: Security I

Revisiting Residue Codes for Modern Memories	73
<i>Evgeny Manzhosov (Columbia University, USA), Adam Hastings (Columbia University, USA), Meghna Pancholi (Columbia University, USA), Ryan Piersma (Columbia University, USA), Mohamed Tarek Ibn Ziad (Columbia University, USA), and Simha Sethumadhavan (Columbia University, USA)</i>	
PageORAM : An Efficient DRAM Page Aware ORAM Strategy	91
<i>Rachit Rajat (University of Southern California, USA), Yongqin Wang (University of Southern California, USA), and Murali Annavaram (University of Southern California, USA)</i>	
AQUA: Scalable Rowhammer Mitigation by Quarantining Aggressor Rows at Runtime	108
<i>Anish Saxena (Georgia Institute of Technology, USA), Gururaj Saileshwar (Georgia Institute of Technology, USA), Prashant J. Nair (University of British Columbia, Canada), and Moinuddin Qureshi (Georgia Institute of Technology, USA)</i>	
CRONUS: Fault-Isolated, Secure and High-Performance Heterogeneous Computing for Trusted Execution Environment	124
<i>Jianguo Jiang (The University of Hong Kong, China), Ji Qi (The University of Hong Kong, China), Tianxiang Shen (The University of Hong Kong, China), Xusheng Chen (The University of Hong Kong, China), Shixiong Zhao (The University of Hong Kong, China), Sen Wang (Huawei Technologies, China), Li Chen (Huawei Technologies, China), Gong Zhang (Huawei Technologies, China), Xiapu Luo (The Hong Kong Polytechnic University, China), and Heming Cui (The University of Hong Kong, China; Pujiang Lab, China)</i>	

Session 1B: Microarchitecture I

Reconstructing Out-of-Order Issue Queue	144
<i>Ipoom Jeong (Yonsei University, Korea), Jiwon Lee (Yonsei University, Korea), Myung Kuk Yoon (Ewha Womans University, Korea), and Won Woo Ro (Yonsei University, Korea)</i>	
Speculative Code Compaction: Eliminating Dead Code via Speculative Microcode Transformations	162
<i>Logan Moody (University of Virginia), Wei Qi (University of Virginia), Abdolrasoul Sharifi (University of Virginia), Layne Berry (University of Virginia), Joey Rudek (University of Virginia), Jayesh Gaur (Processor Architecture Lab, Intel Labs), Jeff Parkhurst (Intel Corporation), Sreenivas Subramoney (Processor Architecture Lab, Intel Labs), Kevin Skadron (University of Virginia), and Ashish Venkat (University of Virginia)</i>	
big.VLITTLE: On-Demand Data-Parallel Acceleration for Mobile Systems on Chip	181
<i>Tuan Ta (Cornell University, USA), Khalid Al-Hawaj (Cornell University, USA), Nick Cebry (Cornell University, USA), Yanghui Ou (Cornell University, USA), Eric Hall (Cornell University, USA), Courtney Golden (Cornell University, USA), and Christopher Batten (Cornell University, USA)</i>	
Exploring Instruction Fusion Opportunities in General Purpose Processors	199
<i>Sawan Singh (University of Murcia, Spain), Arthur Perais (University Grenoble Alpes, CNRS, Grenoble INP, TIMA, Grenoble, France), Alexandra Jimborean (University of Murcia, Spain), and Alberto Ros (University of Murcia, Spain)</i>	

Session 2A: GPUs

DTeXL: Decoupled Raster Pipeline for Texture Locality	213
<i>Diya Joseph (Universitat Politècnica de Catalunya, Spain), Juan L. Aragón (Universidad de Murcia, Spain), Joan-Manuel Parcerisa (Universitat Politècnica de Catalunya, Spain), and Antonio González (Universitat Politècnica de Catalunya, Spain)</i>	
Morpheus: Extending the Last Level Cache Capacity in GPU Systems Using Idle GPU Core Resources	228
<i>Sina Darabi (Sharif University of Technology), Mohammad Sadrosadati (ETH Zurich), Negar Akbarzadeh (Sharif University of Technology), Joël Lindegger (ETH Zurich), Mohammad Hosseini (Institute for Research in Fundamental Sciences (IPM)), Jisung Park (ETH Zurich; POSTECH), Juan Gómez-Luna (ETH Zurich), Onur Mutlu (ETH Zurich), and Hamid Sarbazi-Azad (Sharif University of Technology; Institute for Research in Fundamental Sciences (IPM))</i>	
Featherweight Soft Error Resilience for GPUs	245
<i>Yida Zhang (Purdue University) and Changhee Jung (Purdue University)</i>	

Vulkan-Sim: A GPU Architecture Simulator for Ray Tracing	263
<i>Mohammadreza Saed (University of British Columbia, Canada), Yuan Hsi Chou (University of British Columbia, Canada), Lufei Liu (University of British Columbia, Canada), Tyler Nowicki (Huawei Technologies, Canada), and Tor M. Aamodt (University of British Columbia, Canada)</i>	

Session 2B: Embedded/IoT/Mobile Systems

Pushing Point Cloud Compression to the Edge	282
<i>Ziyu Ying (The Pennsylvania State University, USA), Shulin Zhao (The Pennsylvania State University, USA), Sandeepa Bhuyan (The Pennsylvania State University, USA), Cyan Subhra Mishra (The Pennsylvania State University, USA), Mahmut T. Kandemir (The Pennsylvania State University, USA), and Chita R. Das (The Pennsylvania State University, USA)</i>	
Automatic Domain-Specific SoC Design for Autonomous Unmanned Aerial Vehicles	300
<i>Srivatsan Krishnan (Harvard University), Zishen Wan (Georgia Institute of Technology), Kshitij Bhardwaj (Lawrence Livermore National Lab), Paul Whatmough (ARM Research), Aleksandra Faust (Google Research, Brain Team), Sabrina Neuman (Harvard University), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University), and Vijay Janapa Reddi (Harvard University)</i>	
An Architectural Charge Management Interface for Energy-Harvesting Systems	318
<i>Emily Ruppel (Bosch Research and Technology Center), Milijana Surbatovich (Carnegie Mellon University), Harsh Desai (Carnegie Mellon University), Kiwan Maeng (Pennsylvania State University), and Brandon Lucia (Carnegie Mellon University)</i>	
ROG: A High Performance and Robust Distributed Training System for Robotic IoT	336
<i>Xiuxian Guan (The University of Hong Kong, China), Zekai Sun (The University of Hong Kong, China; Pujiang Lab, China), Shengliang Deng (The University of Hong Kong, China), Xusheng Chen (The University of Hong Kong, China), Shixiong Zhao (The University of Hong Kong, China), Zongyuan Zhang (The University of Hong Kong, China), Tianyang Duan (The University of Hong Kong, China), Yuexuan Wang (The University of Hong Kong, China), Chenshu Wu (The University of Hong Kong, China), Yong Cui (Tsinghua University, China), Libo Zhang (Chinese Academy of Sciences, China), Yanjun Wu (Chinese Academy of Sciences, China), Rui Wang (Southern University of Science and Technology, China), and Heming Cui (The University of Hong Kong, China; Pujiang Lab, China)</i>	

Session 3A: Virtual Memory/Storage

ASSASIN: Architecture Support for Stream Computing to Accelerate Computational Storage	354
<i>Chen Zou (University of Chicago) and Andrew A. Chien (University of Chicago; Argonne National Laboratory)</i>	

DaxVM: Stressing the Limits of Memory as a File Interface	369
<i>Chloe Alverti (National Technical University of Athens), Vasileios Karakostas (University of Athens), Nikhita Kunati (NVIDIA), Georgios Goumas (National Technical University of Athens), and Michael Swift (University of Wisconsin-Madison)</i>	
Networked SSD: Flash Memory Interconnection Network for High-Bandwidth SSD	388
<i>Jiho Kim (KAIST, Republic of Korea), Seokwon Kang (Hanyang University, Republic of Korea), Yongjun Park (Yonsei University, Republic of Korea), and John Kim (KAIST, Republic of Korea)</i>	
Designing Virtual Memory System of MCM GPUs	404
<i>Pratheek B (Indian Institute of Science, India), Neha Jawalkar (Indian Institute of Science, India), and Arkaprava Basu (Indian Institute of Science, India)</i>	

Session 3B: Data Centers/Networks

ALTOCUMULUS: Scalable Scheduling for Nanosecond-Scale Remote Procedure Calls	423
<i>Jiechen Zhao (University of Toronto, Canada), Iris Uwizeyimana (University of Toronto, Canada), Karthik Ganesan (University of Toronto, Canada), Mark C. Jeffrey (University of Toronto, Canada), and Natalie Enright Jerger (University of Toronto, Canada)</i>	
SIMR: Single Instruction Multiple Request Processing for Energy-Efficient Data Center Microservices	441
<i>Mahmoud Khairy (Purdue University), Ahmad Alawneh (Purdue University), Aaron Barnes (Purdue University), and Timothy G. Rogers (Purdue University)</i>	
Patching up Network Data Leaks with Sweeper	464
<i>Marina Vemou (Georgia Institute of Technology, USA), Albert Cho (Georgia Institute of Technology, USA), and Alexandros Daglis (Georgia Institute of Technology, USA)</i>	
IDIO: Network-Driven, Inbound Network Data Orchestration on Server Processors	480
<i>Mohammad Alian (University of Kansas), Siddharth Agarwal (University of Illinois), Jongmin Shin (DGIST), Neel Patel (University of Kansas), Yifan Yuan (University of Illinois), Daehoon Kim (DGIST), Ren Wang (Intel Labs), and Nam Sung Kim (University of Illinois)</i>	

Session 4A: Compilers/Code Generation

Treebeard: An Optimizing Compiler for Decision Tree Based ML Inference	494
<i>Ashwin Prasad (Indian Institute of Science, India), Sampath Rajendra (Microsoft Research, India), Kaushik Rajan (Microsoft Research, India), R. Govindarajan (Indian Institute of Science, India), and Uday Bondhugula (Indian Institute of Science, India)</i>	
GCD^2: A Globally Optimizing Compiler for Mapping DNNs to Mobile DSPs	512
<i>Wei Niu (William & Mary, USA), Jiexiong Guan (William & Mary, USA), Xipeng Shen (North Carolina State University, USA), Yanzhi Wang (Northeastern University, USA), Gagan Agrawal (Augusta University, USA), and Bin Ren (William & Mary, USA)</i>	

OCOLOS: Online COde Layout OptimizationS	530
<i>Yuxuan Zhang (University of Pennsylvania, USA), Tanvir Ahmed Khan (University of Michigan, USA), Gilles Pokam (Intel Corporation, USA), Baris Kasikci (University of Michigan, USA), Heiner Litz (University of California, Santa Cruz, USA), and Joseph Devietti (University of Pennsylvania, USA)</i>	
RipTide: A Programmable, Energy-Minimal Dataflow Compiler and Architecture	546
<i>Graham Gobieski (Carnegie Mellon University), Souradip Ghosh (Carnegie Mellon University), Marijn Heule (Carnegie Mellon University), Todd Mowry (Carnegie Mellon University), Tony Nowatzki (University of California Los Angeles), Nathan Beckmann (Carnegie Mellon University), and Brandon Lucia (Carnegie Mellon University)</i>	

Session 4B: Machine Learning I

Skipper: Enabling Efficient SNN Training Through Activation-Checkpointing and Time-Skipping	565
<i>Sonali Singh (Pennsylvania State University, USA), Anup Sarma (Pennsylvania State University, USA), Sen Lu (Pennsylvania State University, USA), Abhronil Sengupta (Pennsylvania State University, USA), Mahmut T. Kandemir (Pennsylvania State University, USA), Emre Neftci (RWTH Aachen, Germany), Vijaykrishnan Narayanan (Pennsylvania State University, USA), and Chita R. Das (Pennsylvania State University, USA)</i>	
Going Further With Winograd Convolutions: Tap-Wise Quantization for Efficient Inference on 4x4 Tiles	582
<i>Renzo Andri (Computing Systems Lab, Huawei Zurich Research Center), Beatrice Bussolino (Computing Systems Lab, Huawei Zurich Research Center; DET, Politecnico di Torino, Italy), Antonio Cipolletta (Computing Systems Lab, Huawei Zurich Research Center), Lukas Cavigelli (Computing Systems Lab, Huawei Zurich Research Center), and Zhe Wang (Computing Systems Lab, Huawei Zurich Research Center)</i>	
Adaptable Butterfly Accelerator for Attention-Based NNs via Hardware and Algorithm Co-Design	599
<i>Hongxiang Fan (Imperial College London, UK), Thomas Chau (Samsung AI Center Cambridge, UK), Stylianos I. Venieris (Samsung AI Center Cambridge, UK), Royson Lee (University of Cambridge, UK), Alexandros Kouris (Samsung AI Center Cambridge, UK), Wayne Luk (Imperial College London, UK), Nicholas D. Lane (Samsung AI Cambridge and University of Cambridge, UK), and Mohamed S. Abdelfattah (Cornell University, United States)</i>	
DFX: A Low-Latency Multi-FPGA Appliance for Accelerating Transformer-Based Text Generation	616
<i>Seongmin Hong (KAIST, South Korea), Seungjae Moon (KAIST, South Korea), Junsoo Kim (KAIST, South Korea), Sungjae Lee (NAVER CLOVA, South Korea), Minsub Kim (NAVER CLOVA, South Korea), Dongsoo Lee (NAVER CLOVA, South Korea), and Joo-Young Kim (KAIST, South Korea)</i>	

HARMONY: Heterogeneity-Aware Hierarchical Management for Federated Learning System	631
<i>Chunlin Tian (University of Macau, China), Li Li (University of Macau, China), Zhan Shi (The University of Texas at Austin, USA), Jun Wang (Futurewei Technology, USA), and ChengZhong Xu (University of Macau, China)</i>	

Session 5A: Security II

Leaky Way: A Conflict-Based Cache Covert Channel Bypassing Set Associativity	646
<i>Yanan Guo (University of Pittsburgh, USA), Xin Xin (University of Pittsburgh, USA), Youtao Zhang (University of Pittsburgh, USA), and Jun Yang (University of Pittsburgh, USA)</i>	
SwiftDir: Secure Cache Coherence Without Overprotection	662
<i>Chenlu Miao (Zhejiang University, China), Kai Bu (Zhejiang University, China), Mengming Li (Zhejiang University, China), Shaowu Mao (Huawei Technologies, China), and Jianwei Jia (Huawei Technologies, China)</i>	
Self-Reinforcing Memoization for Cryptography Calculations in Secure Memory Systems	678
<i>Xin Wang (Virginia Tech, USA), Daulet Talapkaliyev (Virginia Tech, USA), Matthew Hicks (Virginia Tech, USA), and Xun Jian (Virginia Tech, USA)</i>	
Eager Memory Cryptography in Caches	693
<i>Xin Wang (Virginia Tech), Jagadish B. Kotra (AMD Research), and Xun Jian (Virginia Tech)</i>	

Session 5B: Accelerators in/near Memory

GenPIP: In-Memory Acceleration of Genome Analysis via Tight Integration of Basecalling and Read Mapping	710
<i>Haiyu Mao (ETH Zurich), Mohammed Alser (ETH Zurich), Mohammad Sadrosadati (ETH Zurich), Can Firtina (ETH Zurich), Akanksha Baranwal (ETH Zurich), Damla Senol Cali (Bionano Genomics (San Diego, CA, USA)), Aditya Manglik (ETH Zurich), Nour Almadhoun Alser (ETH Zurich), and Onur Mutlu (ETH Zurich)</i>	
BEACON: Scalable Near-Data-Processing Accelerators for Genome Analysis near Memory Pool with the CXL Support	727
<i>Wenqin Huangfu (University of California, Santa Barbara), Krishna T. Malladi (Samsung Semiconductor, Inc.), Andrew Chang (Samsung Semiconductor, Inc.), and Yuan Xie (University of California, Santa Barbara)</i>	
Sparse Attention Acceleration with Synergistic In-Memory Pruning and On-Chip Recomputation..	744
<i>Amir Yazdanbakhsh (Google Research, Brain Team), Ashkan Moradifirouzabadi (University of California, San Diego; Google Research, Brain Team), Zheng Li (University of California, San Diego: Google Research, Brain Team), and Mingu Kang (University of California, San Diego)</i>	

ICE: An Intelligent Cognition Engine with 3D NAND-Based In-Memory Computing for Vector Similarity Search Acceleration	763
<p><i>Han-Wen Hu (Macronix International Co., Ltd., Taiwan, National Tsing Hua University, Taiwan), Wei-Chen Wang (Massachusetts Institute of Technology, USA, Macronix International Co., Ltd., Taiwan, National Taiwan University, Taiwan), Yuan-Hao Chang (Academia Sinica, Taiwan), Yung-Chun Lee (Macronix International Co., Ltd., Taiwan), Bo-Rong Lin (Macronix International Co., Ltd., Taiwan), Huai-Mu Wang (Macronix International Co., Ltd., Taiwan), Yen-Po Lin (Macronix International Co., Ltd., Taiwan, National Taiwan University, Taiwan), Yu-Ming Huang (Macronix International Co., Ltd., Taiwan), Chong-Ying Lee (Macronix International Co., Ltd., Taiwan), Tzu-Hsiang Su (Macronix International Co., Ltd., Taiwan), Chih-Chang Hsieh (Macronix International Co., Ltd., Taiwan), Chia-Ming Hu (Macronix International Co., Ltd., Taiwan), Yi-Ting Lai (Macronix International Co., Ltd., Taiwan), Chung-Kuang Chen (Macronix International Co., Ltd., Taiwan), Han-Sung Chen (Macronix International Co., Ltd., Taiwan), Hsiang-Pang Li (Macronix International Co., Ltd., Taiwan), Tei-Wei Kuo (National Taiwan University, Taiwan), Meng-Fan Chang (National Tsing Hua University, Taiwan), Keh-Chung Wang (Macronix International Co., Ltd., Taiwan), Chun-Hsiung Hung (Macronix International Co., Ltd., Taiwan), and Chih-Yuan Lu (Macronix International Co., Ltd., Taiwan)</i></p>	

Session 6A: Potpourri

CORUSCANT: Fast Efficient Processing-in-Racetrack Memories	784
<p><i>Sebastien Ollivier (University of Pittsburgh, USA), Stephen Longofono (University of Pittsburgh, USA), Prayash Dutta (University of South Florida, USA), Jingtong Hu (University of Pittsburgh, USA), Sanjukta Bhanja (University of South Florida, USA), and Alex K. Jones (University of Pittsburgh, USA)</i></p>	
IDLD: Instantaneous Detection of Leakage and Duplication of Identifiers used for Register Renaming	799
<p><i>Yiannakis Sazeides (University of Cyprus, Cyprus), Alex Gerber (Google, Israel), Ron Gabor (NVIDIA, Israel), Arkady Bramnik (Intel, Israel), George Papadimitriou (University of Athens, Greece), Dimitris Gizopoulos (University of Athens, Greece), Chrysostomos Nicopoulos (University of Cyprus, Cyprus), Giorgos Dimitrakopoulos (Democritus University of Thrace, Greece), and Karyofyllis Patsidis (Democritus University of Thrace, Greece)</i></p>	
HiRA: Hidden Row Activation for Reducing Refresh Latency of Off-the-Shelf DRAM Chips	815
<p><i>A. Giray Yaglikci (ETH Zürich), Ataberk Olgun (ETH Zurich; TOBB University of Economics and Technology), Minesh Patel (ETH Zürich), Haocong Luo (ETH Zürich), Hasan Hassan (ETH Zürich), Lois Orosa (ETH Zürich; Galicia Supercomputing Center (CESGA)), Oguz Ergin (TOBB University of Economics and Technology), and Onur Mutlu (ETH Zürich)</i></p>	

Session 6B: Energy

AgileWatts: An Energy-Efficient CPU Core Idle-State Architecture for Latency-Sensitive Server Applications	835
<i>Jawad Haj-Yahya (Rivos Inc., USA), Haris Volos (University of Cyprus), Davide B. Bartolini (Huawei), Georgia Antoniou (University of Cyprus), Jeremie S. Kim (ETH Zurich), Zhe Wang (Huawei), Kleovoulos Kalaitzidis (Huawei), Tom Rollet (Huawei), Chen Zhirui (Huawei), Geng Ye (Huawei), Onur Mutlu (ETH Zurich, Switzerland), and Yiannakis Sazeides (University of Cyprus)</i>	
AgilePkgC: An Agile System Idle State Architecture for Energy Proportional Datacenter Servers	851
<i>Georgia Antoniou (University of Cyprus, Cyprus), Haris Volos (University of Cyprus, Cyprus), Davide B. Bartolini (Huawei, Switzerland), Tom Rollet (Huawei, Switzerland), Yiannakis Sazeides (University of Cyprus, Cyprus), and Jawad Haj-Yahya (Rivos Inc., United States)</i>	
Realizing Emotional Interactions to Learn User Experience and Guide Energy Optimization for Mobile Architectures	868
<i>Xueliang Li (Shenzhen University, China), Zhuobin Shi (Shenzhen University, China), Junyang Chen (Shenzhen University, China), and Yepang Liu (Southern University of Science and Technology, China)</i>	

Session 7A: PiM

FracDRAM: Fractional Values in Off-the-Shelf DRAM	885
<i>Fei Gao (Princeton University, USA), Georgios Tziantzioulis (Princeton University, USA), and David Wentzloff (Princeton University, USA)</i>	
pLUTo: Enabling Massively Parallel Computation in DRAM via Lookup Tables	900
<i>João Dinis Ferreira (ETH Zürich), Gabriel Falcao (University of Coimbra), Juan Gómez-Luna (ETH Zürich), Mohammed Alser (ETH Zürich), Lois Orosa (ETH Zürich; Galicia Supercomputing Center), Mohammad Sadrosadati (ETH Zürich), Jeremie S. Kim (ETH Zürich), Geraldo F. Oliveira (ETH Zürich), Taha Shahroodi (TU Delft), Anant Nori (Intel), and Onur Mutlu (ETH Zürich)</i>	
Multi-layer In-memory Processing	920
<i>Daichi Fujiki (Keio University), Alireza Khadem (University of Michigan), Scott Mahlke (University of Michigan, Nvidia Research), and Reetuparna Das (University of Michigan)</i>	
Flash-Cosmos: In-Flash Bulk Bitwise Operations Using Inherent Computation Capability of NAND Flash Memory	937
<i>Jisung Park (POSTECH; ETH Zürich), Roknoddin Azizi (ETH Zürich), Geraldo F. Oliveira (ETH Zürich), Mohammad Sadrosadati (ETH Zürich), Rakesh Nadig (ETH Zürich), David Novo (LIRMM, Univ Montpellier, CNRS), Juan Gómez Luna (ETH Zürich), Myungsuk Kim (Kyungpook National University), and Onur Mutlu (ETH Zürich)</i>	

Session 7B: Microarchitecture II

Page Size Aware Cache Prefetching	956
<i>Georgios Vavouliotis (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain), Gino Chacon (Texas A&M University, USA), Lluc Alvarez (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain), Paul V. Gratz (Texas A&M University, USA), Daniel A. Jiménez (Texas A&M University, USA), and Marc Casas (Barcelona Supercomputing Center, Spain; Universitat Politècnica de Catalunya, Spain)</i>	
Berti: An Accurate Local-Delta Data Prefetcher	975
<i>Agustin Navarro-Torres (Universidad de Zaragoza, Spain), Biswabandan Panda (Indian Institute of Technology Bombay, India), Jesús Alastruey-Benedé (Universidad de Zaragoza, Spain), Pablo Ibáñez (Universidad de Zaragoza, Spain), Victor Viñals-Yúfera (Universidad de Zaragoza, Spain), and Alberto Ros (University of Murcia, Spain)</i>	
Translation-Optimized Memory Compression for Capacity	992
<i>Gagandeep Panwar (Virginia Tech), Muhammad Laghari (Virginia Tech), David Bears (Virginia Tech), Yuqing Liu (Virginia Tech), Chandler Jearls (Virginia Tech; Apple), Esha Choukse (Microsoft Research), Kirk W. Cameron (Virginia Tech), Ali R. Butt (Virginia Tech), and Xun Jian (Virginia Tech)</i>	
Merging Similar Patterns for Hardware Prefetching	1012
<i>Shizhi Jiang (University of Chinese Academy of Sciences; Institute of Software, Chinese Academy of Sciences, China), Qiusong Yang (Institute of Software, Chinese Academy of Sciences, China), and Yiwei Ci (Institute of Software, Chinese Academy of Sciences, China)</i>	

Session 8: Quantum Computing

AutoComm: A Framework for Enabling Efficient Communication in Distributed Quantum Programs..	1027
<i>Anbang Wu (University of California, USA), Hezi Zhang (University of California, USA), Gushu Li (University of California, USA), Alireza Shabani (Cisco Research, Los Angeles, USA), Yuan Xie (University of California, USA), and Yufei Ding (University of California, USA)</i>	
Let Each Quantum Bit Choose Its Basis Gates	1042
<i>Sophia Fuhui Lin (University of Chicago, USA), Sara Sussman (Princeton University, USA), Casey Duckering (University of Chicago, USA), Pranav S. Mundada (Princeton University, USA), Jonathan M. Baker (University of Chicago, USA), Rohan S. Kumar (University of Chicago, USA), Andrew A. Houck (Princeton University, USA), and Frederic T. Chong (University of Chicago, USA)</i>	
COMPAQT: Compressed Waveform Memory Architecture for Scalable Qubit Control	1059
<i>Satvik Maurya (University of Wisconsin-Madison, USA) and Swamit Tannu (University of Wisconsin-Madison, USA)</i>	

Qubit Mapping and Routing via MaxSAT	1078
<i>Abtin Molavi (University of Wisconsin-Madison, USA), Amanda Xu (University of Wisconsin-Madison, USA), Martin Diges (University of Wisconsin-Madison, USA), Lauren Pick (University of Wisconsin-Madison, USA), Swamit Tannu (University of Wisconsin-Madison, USA), and Aws Albarghouthi (University of Wisconsin-Madison, USA)</i>	
Scaling Superconducting Quantum Computers with Chiplet Architectures	1092
<i>Kaitlin N. Smith (University of Chicago, USA), Gokul Subramanian Ravi (University of Chicago, USA), Jonathan M. Baker (University of Chicago, USA), and Frederic T. Chong (University of Chicago, USA)</i>	
Q3DE: A Fault-Tolerant Quantum Computer Architecture for Multi-bit Burst Errors by Cosmic Rays	1110
<i>Yasunari Suzuki (Nippon Telegraph and Telephone, Japan; JST, PRESTO, Japan), Takanori Sugiyama (The University of Tokyo, Japan), Tomochika Arai (Nippon Telegraph and Telephone, Japan; The University of Tokyo, Japan), Wang Liao (The University of Tokyo, Japan), Koji Inoue (Kyushu University, Japan), and Teruo Tanimoto (Kyushu University, Japan; JST, PRESTO, Japan)</i>	

Session 9A: Design Methodology

RemembERR: Leveraging Microprocessor Errata for Design Testing and Validation	1126
<i>Flavien Solt (ETH Zürich), Patrick Jattke (ETH Zürich), and Kaveh Razaavi (ETH Zürich)</i>	
Datamime: Generating Representative Benchmarks by Automatically Synthesizing Datasets	1144
<i>Hyun Ryong Lee (Massachusetts Institute of Technology, USA) and Daniel Sanchez (Massachusetts Institute of Technology, USA)</i>	
An Architecture Interface and Offload Model for low-Overhead, Near-Data, Distributed Accelerators	1160
<i>Saambhavi Baskaran (Pennsylvania State University, USA), Mahmut Taylan Kandemir (Pennsylvania State University, USA), and Jack Sampson (Pennsylvania State University, USA)</i>	

Towards Developing High Performance RISC-V Processors Using Agile Methodology 1178

Yinan Xu (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Zihao Yu (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Dan Tang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Guokai Chen (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Lu Chen (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Lingrui Gou (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Yue Jin (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Qianruo Li (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Xin Li (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Zuojun Li (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Jiawei Lin (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Tong Liu (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Zhigang Liu (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Jiazhan Tan (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Huaqiang Wang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Huizhe Wang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Kaifan Wang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Chuanqi Zhang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Fawang Zhang (Shenzhen University, China), Linjuan Zhang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Zifei Zhang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Yangyang Zhao (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Yaoyang Zhou (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Yike Zhou (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Jiangrui Zou (Shenzhen University, China), Ye Cai (Shenzhen University, China), Dandan Huan (Beijing VCore Technology Co., Ltd., China), Zusong Li (Beijing VCore Technology Co., Ltd., China), Jiye Zhao (Beijing VCore Technology Co., Ltd., China), Zihao Chen (Peng Cheng Laboratory, China), Wei He (Peng Cheng Laboratory, China), Qiyuan Quan (Peng Cheng Laboratory, China), Xingwu Liu (Dalian University of Technology, China), Sa Wang (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Kan Shi (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), Ninghui Sun (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, China), and Yungang

Bao (State Key Lab of Processors, Institute of Computing Technology,
Chinese Academy of Sciences, China)

Session 9B: Accelerators for Security

DiVa: An Accelerator for Differentially Private Machine Learning	1200
<i>Beomsik Park (Korea Advanced Institute of Science and Technology (KAIST), South Korea), Ranggi Hwang (Korea Advanced Institute of Science and Technology (KAIST), South Korea), Dongho Yoon (Korea Advanced Institute of Science and Technology (KAIST), South Korea), Yoonhyuk Choi (Korea Advanced Institute of Science and Technology (KAIST), South Korea), and Minsoo Rhu (Korea Advanced Institute of Science and Technology (KAIST), South Korea)</i>	
EVAX: Towards a Practical, Pro-Active & Adaptive Architecture for High Performance & Security	1218
<i>Samira Mirbagher Ajorpaz (North Carolina State University, USA), Daniel Moghimi (University of California San Diego, USA), Jeff Collins (Independent, USA), Gilles Pokam (Intel Labs, USA), Nael Abu-Ghazaleh (University of California Riverside, USA), and Dean Tullsen (University of California San Diego, USA)</i>	
ARK: Fully Homomorphic Encryption Accelerator with Runtime Data Generation and Inter-Operation Key Reuse	1237
<i>Jongmin Kim (Seoul National University, South Korea), Gwangho Lee (Seoul National University, South Korea), Sangpyo Kim (Seoul National University, South Korea), Gina Sohn (Seoul National University, South Korea), Minsoo Rhu (KAIST, South Korea), John Kim (KAIST, South Korea), and Jung Ho Ahn (Seoul National University, South Korea)</i>	
Horus: Persistent Security for Extended Persistence-Domain Memory Systems	1255
<i>Xijing Han (North Carolina State University, USA), James Tuck (North Carolina State University, USA), and Amro Awad (North Carolina State University, USA)</i>	

Session 10A: Graph Processing

Mint: An Accelerator for Mining Temporal Motifs	1270
<i>Nishil Talati (University of Michigan, USA), Haojie Ye (University of Michigan, USA), Sanketh Vedula (Technion – Israel Institute of Technology, Israel), Kuan-Yu Chen (University of Michigan, USA), Yuhan Chen (University of Michigan, USA), Daniel Liu (University of Michigan, USA), Yichao Yuan (University of Michigan, USA), David Blaauw (University of Michigan, USA), Alex Bronstein (Technion – Israel Institute of Technology, Israel), Trevor Mudge (University of Michigan, USA), and Ronald Dreslinski (University of Michigan, USA)</i>	
DPU-v2: Energy-Efficient Execution of Irregular Directed Acyclic Graphs	1288
<i>Nimish Shah (KU Leuven, Belgium), Wannes Meert (KU Leuven, Belgium), and Marian Verhelst (KU Leuven, Belgium)</i>	

XPGraph: XPLine-Friendly Persistent Memory Graph Stores for Large-Scale Evolving Graphs	1308
<i>Rui Wang (Zhejiang University, China), Shuibing He (Zhejiang University, China), Weixu Zong (Zhejiang University, China), Yongkun Li (University of Science and Technology of China, China), and Yinlong Xu (University of Science and Technology of China, China)</i>	
A Data-Centric Accelerator for High-Performance Hypergraph Processing	1326
<i>Qinggang Wang (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China; Zhejiang Lab, Hangzhou, China), Long Zheng (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China; Zhejiang Lab, Hangzhou, China), Ao Hu (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China; Zhejiang Lab, Hangzhou, China), Yu Huang (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China; Zhejiang Lab, Hangzhou, China), Pengcheng Yao (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China; Zhejiang Lab, Hangzhou, China), Chuangyi Gui (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China; Zhejiang Lab, Hangzhou, China), Xiaofei Liao (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China), Hai Jin (National Engineering Research Center for Big Data Technology and System/Services Computing Technology and System Lab/Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China), and Jingling Xue (School of Computer Science and Engineering, University of New South Wales, Sydney, Australia)</i>	
ReGraph: Scaling Graph Processing on HBM-Enabled FPGAs with Heterogeneous Pipelines	1342
<i>Xinyu Chen (National University of Singapore), Yao Chen (Advanced Digital Sciences Center), Feng Cheng (City University of Hong Kong), Hongshi Tan (National University of Singapore), Bingsheng He (National University of Singapore), and Weng-fai Wong (National University of Singapore)</i>	

Session 10B: Machine Learning II

3D-FPIM: An Extreme Energy-Efficient DNN Acceleration System Using 3D NAND Flash-Based In-Situ PIM Unit	1359
<i>Hunjun Lee (Seoul National University, South Korea), Minseop Kim (Seoul National University, South Korea), Dongmoon Min (Seoul National University, South Korea), Joonsung Kim (Google, USA), Jongwon Back (Seoul National University, South Korea), Honam Yoo (Seoul National University, South Korea), Jong-Ho Lee (Ministry of Science and ICT, South Korea), and Jangwoo Kim (Seoul National University, South Korea)</i>	
Sparseloop: An Analytical Approach to Sparse Tensor Accelerator Modeling	1377
<i>Yannan Nellie Wu (MIT, USA), Po-An Tsai (NVIDIA, USA), Angshuman Parashar (NVIDIA, USA), Vivienne Sze (MIT, USA), and Joel S. Emer (MIT / NVIDIA, USA)</i>	
DeepBurning-SEG: Generating DNN Accelerators of Segment-Grained Pipeline Architecture	1396
<i>Xuyi Cai (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Ying Wang (Institute of Computing Technology, Chinese Academy of Sciences, China; Zhejiang Lab, China), Xiaohan Ma (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Yinhe Han (Institute of Computing Technology, Chinese Academy of Sciences, China; Zhejiang Lab, China), and Lei Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China)</i>	
ANT: Exploiting Adaptive Numerical Data Type for Low-bit Deep Neural Network Quantization	1414
<i>Cong Guo (Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute), Chen Zhang (Microsoft Research, China), Jingwen Leng (Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute), Zihan Liu (Shanghai Jiao Tong University, China), Fan Yang (Microsoft Research, China), Yunxin Liu (Institute for AI Industry Research (AIR), Tsinghua University, China), Minyi Guo (Shanghai Jiao Tong University, China; Shanghai Qi Zhi Institute), and Yuhao Zhu (University of Rochester, US)</i>	
Ristretto: An Atomized Processing Architecture for Sparsity-Condensed Stream Flow in CNN	1434
<i>Gang Li (Shanghai Jiao Tong University, China), Weixiang Xu (Institute of Automation, CAS, China), Zhuoran Song (Shanghai Jiao Tong University, China), Naifeng Jing (Shanghai Jiao Tong University, China), Jian Cheng (Institute of Automation, CAS, China), and Xiaoyao Liang (Shanghai Jiao Tong University, China)</i>	

Author Index