

2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA 2021)

**Virtual Conference
27 February - 3 March 2021**

Pages 1-466



**IEEE Catalog Number: CFP21013-POD
ISBN: 978-1-6654-4670-9**

**Copyright © 2021 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:	CFP21013-POD
ISBN (Print-On-Demand):	978-1-6654-4670-9
ISBN (Online):	978-1-6654-2235-2
ISSN:	1530-0897

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

2021 IEEE International Symposium on High- Performance Computer Architecture (HPCA) HPCA 2021

Table of Contents

Message from the General Chairs .xvi.....	
Message from the Program Chair .xviii.....	
Message from the Industry Track Program Chair .xxi.....	
Organizing Committee .xxii.....	
Program Committee .xxiii.....	
External Review Committee .xxv.....	
Industry Session Committee .xxvii.....	

Security Architectures

Common Counters: Compressed Encryption Counters for Secure GPU Memory .1.....	
<i>Seonjin Na (KAIST), Sunho Lee (KAIST), Yeonjae Kim (KAIST), Jongse Park (KAIST), and Jaehyuk Huh (KAIST)</i>	
Streamline Ring ORAM Accesses Through Spatial and Temporal Optimization .14.....	
<i>Dingyuan Cao (Tsinghua University, China), Mingzhe Zhang (State Key Laboratory of Computer Architecture, ICT, CAS, Beijing, China), Hang Lu (State Key Laboratory of Computer Architecture, ICT, CAS, Beijing, China), Xiaochun Ye (State Key Laboratory of Computer Architecture, ICT, CAS, Beijing, China), Dongrui Fan (State Key Laboratory of Computer Architecture, ICT, CAS, Beijing, China), Yuezhi Che (Illinois Institute of Technology), and Rujia Wang (Illinois Institute of Technology)</i>	
Cheetah: Optimizing and Accelerating Homomorphic Encryption for Private Inference .26.....	
<i>Brandon Reagen (New York University; Facebook AI Research), Woo-Seok Choi (Seoul National University), Yeongil Ko (Harvard University), Vincent Lee (Facebook Reality Labs Research), Hsien-Hsin Lee (Facebook AI Research), Gu-Yeon Wei (Harvard University), and David Brooks (Harvard University)</i>	
New Models for Understanding and Reasoning About Speculative Execution Attacks .40.....	
<i>Zecheng He (Princeton University, USA), Guangyuan Hu (Princeton University, USA), and Ruby Lee (Princeton University, USA)</i>	

Accelerators for Machine Learning 1

A Computational Stack for Cross-Domain Acceleration .54.....
Sean Kinzer (University of California San Diego), Joon Kyung Kim (University of California San Diego), Soroush Ghodrati (University of California San Diego), Brahmendra Yatham (University of California San Diego), Alric Althoff (Tortuga Logic), Divya Mahajan (Microsoft), Sorin Lerner (University of California San Diego), and Hadi Esmaeilzadeh (University of California San Diego)

Heterogeneous Dataflow Accelerators for Multi-DNN Workloads .71.....
Hyoukjun Kwon (Facebook, USA; Georgia Institute of Technology, USA), Liangzhen Lai (Facebook, USA), Michael Pellauer (NVIDIA, USA), Tushar Krishna (Georgia Institute of Technology, USA), Yu-Hsin Chen (Facebook, USA), and Vikas Chandra (Facebook, USA)

SPAGHETTI: Streaming Accelerators for Highly Sparse GEMM on FPGAs .84.....
Reza Hojabr (University of Tehran, Iran and Simon Fraser University, Canada), Ali Sedaghati (Simon Fraser University, Canada), Amirali Sharifian (Untether AI, Canada), Ahmad Khonsari (University of Tehran, Iran and IPM, Iran), and Arvindh Shriraman (Simon Fraser University, Canada)

SpAtten: Efficient Sparse Attention Architecture with Cascade Token and Head Pruning .97.....
Hanrui Wang (Massachusetts Institute of Technology, USA), Zhekai Zhang (Massachusetts Institute of Technology, USA), and Song Han (Massachusetts Institute of Technology, USA)

Storage Systems

BBB: Simplifying Persistent Programming Using Battery-Backed Buffers .111.....
Mohammad Alshboul (North Carolina State University), Prakash Ramrakhiani (Arm Research), William Wang (Arm Research), James Tuck (North Carolina State University), and Yan Solihin (University of Central Florida)

TSOPER: Efficient Coherence-Based Strict Persistency .125.....
Per Ekemark (Uppsala University, Sweden), Yuan Yao (Uppsala University, Sweden), Alberto Ros (University of Murcia, Spain), Konstantinos Sagonas (Uppsala University, Sweden and National Technical University of Athens, Greece), and Stefanos Kaxiras (Uppsala University, Sweden)

Stealth-Persist: Architectural Support for Persistent Applications in Hybrid Memory Systems .139.....
Mazen Alwadi (University of Central Florida), Vamsee Reddy Kommareddy (University of Central Florida), Clayton Hughes (Sandia National Laboratories), Simon David Hammond (Sandia National Laboratories), and Amro Awad (North Carolina State University)

Quantum Computing

- TILT: Achieving Higher Fidelity on a Trapped-Ion Linear-Tape Quantum Computing Architecture .153.....
Xin-Chuan Wu (University of Chicago, USA), Dripto M. Debroy (Duke University, USA), Yongshan Ding (University of Chicago, USA), Jonathan M. Baker (University of Chicago, USA), Yuri Alexeev (Argonne National Laboratory, USA), Kenneth R. Brown (Duke University, USA), and Frederic T. Chong (University of Chicago, USA)
- QuCloud: A New Qubit Mapping Mechanism for Multi-programming Quantum Computing in Cloud Environment .167.....
Lei Liu (Sys-Inventor Lab, SKLCA, ICT, CAS) and Xinglei Dou (Sys-Inventor Lab, SKLCA, ICT, CAS)
- Systematic Approaches for Precise and Approximate Quantum State Runtime Assertion .179..
Ji Liu (North Carolina State University, USA) and Huiyang Zhou (North Carolina State University, USA)
- Faster Schrödinger-Style Simulation of Quantum Circuits .194.....
Aneeqa Fatima (University of Michigan) and Igor Markov (University of Michigan)

Systems for Machine Learning 1

- Mix and Match: A Novel FPGA-Centric Deep Neural Network Quantization Framework .208.....
Sung-En Chang (Northeastern University, USA), Yanyu Li (Northeastern University, USA), Mengshu Sun (Northeastern University, USA), Runbin Shi (The University of Hong Kong, China), Hayden K.-H. So (The University of Hong Kong, China), Xuehai Qian (University of Southern California, USA), Yanzhi Wang (Northeastern University, USA), and Xue Lin (Northeastern University, USA)
- Revisiting HyperDimensional Learning for FPGA and Low-Power Architectures .221.....
Mohsen Imani (University of California Irvine; University of California San Diego), Zhuowen Zou (University of California San Diego), Samuel Bosch (Massachusetts Institute of Technology), Sanjay Anantha Rao (University of California San Diego), Sahand Salamat (University of California San Diego), Venkatesh Kumar (University of California San Diego), Yeseong Kim (Daegu Institue of Science and Technology), and Tajana Rosing (University of California San Diego)
- Tensor Casting: Co-Designing Algorithm-Architecture for Personalized Recommendation Training .235.....
Youngeun Kwon (KAIST), Yunjae Lee (KAIST), and Minsoo Rhu (KAIST)
- GradPIM: A Practical Processing-in-DRAM Architecture for Gradient Descent .249.....
Heesu Kim (Seoul National University, South Korea), Hanmin Park (Samsung Electronics, South Korea), Taehyun Kim (Seoul National University, South Korea), Kwanheum Cho (Yonsei University, South Korea), Eojin Lee (Samsung Electronics, South Korea), Soojung Ryu (Seoul National University, South Korea), Hyuk-Jae Lee (Seoul National University, South Korea), Kiyong Choi (Seoul National University, South Korea), and Jinho Lee (Yonsei University, South Korea)

Cache Design

- SynCron: Efficient Synchronization Support for Near-Data-Processing Architectures .263.....
Christina Giannoula (National Technical University of Athens, ETH Zürich), Nandita Vijaykumar (University of Toronto, ETH Zürich), Nikela Papadopoulou (National Technical University of Athens), Vasileios Karakostas (National Technical University of Athens), Ivan Fernandez (University of Malaga, ETH Zürich), Juan Gómez-Luna (ETH Zürich), Lois Orosa (ETH Zürich), Nectarios Koziris (National Technical University of Athens), Georgios Goumas (National Technical University of Athens), and Onur Mutlu (ETH Zürich)
- Zero Directory Eviction Victim: Unbounded Coherence Directory and Core Cache Isolation .277
Mainak Chaudhuri (Indian Institute of Technology Kanpur)
- Designing a Cost-Effective Cache Replacement Policy Using Machine Learning .291.....
Subhash Sethumurugan (University of Minnesota, Twin Cities), Jieming Yin (Lehigh University), and John Sartori (University of Minnesota, Twin Cities)
- WiDir: A Wireless-Enabled Directory Cache Coherence Protocol .304.....
Antonio Franques (University of Illinois at Urbana-Champaign (UIUC), USA), Apostolos Kokolis (University of Illinois at Urbana-Champaign (UIUC), USA), Sergi Abadal (Universitat Politècnica de Catalunya (UPC), Spain), Vimuth Fernando (University of Illinois at Urbana-Champaign (UIUC), USA), Sasa Misailovic (University of Illinois at Urbana-Champaign (UIUC), USA), and Josep Torrellas (University of Illinois at Urbana-Champaign (UIUC), USA)

Security Attacks

- Heat Behind the Meter: A Hidden Threat of Thermal Attacks in Edge Colocation Data Centers .318
Zhihui Shao (University of California), Mohammad Islam (University of Texas at Arlington), and Shaolei Ren (University of California)
- Trident: A Hybrid Correlation-Collision GPU Cache Timing Attack for AES Key Recovery .332.....
Jaeguk Ahn (KAIST), Cheolgyu Jin (KAIST), Jiho Kim (KAIST), Minsoo Rhu (KAIST), Yunsi Fei (Northeastern University), David Kaeli (Northeastern University), and John Kim (KAIST)
- BlockHammer: Preventing RowHammer at Low Cost by Blacklisting Rapidly-Accessed DRAM Rows .345.....
A. Giray Yaglıkçı (ETH Zürich), Minesh Patel (ETH Zürich), Jeremie S. Kim (ETH Zürich), Roknoddin Azizi (ETH Zürich), Ataberk Olgun (ETH Zürich), Lois Orosa (ETH Zürich), Hasan Hassan (ETH Zürich), Jisung Park (ETH Zürich), Konstantinos Kanellopoulos (ETH Zürich), Taha Shahroodi (ETH Zürich), Saugata Ghose (University of Illinois at Urbana-Champaign), and Onur Mutlu (ETH Zürich)
- A Write-Friendly and Fast-Recovery Scheme for Security Metadata in Non-Volatile Memories .359
Jianming Huang (Huazhong University of Science and Technology, China) and Yu Hua (Huazhong University of Science and Technology, China)

Hardware Accelerators Beyond Machine Learning

- DepGraph: A Dependency-Driven Accelerator for Efficient Iterative Graph Processing .371.....
Yu Zhang (Huazhong University of Science and Technology, China), Xiaofei Liao (Huazhong University of Science and Technology, China), Hai Jin (Huazhong University of Science and Technology, China), Ligang He (University of Warwick, United Kingdom), Bingsheng He (National University of Singapore, Singapore), Haikun Liu (Huazhong University of Science and Technology, China), and Lin Gu (Huazhong University of Science and Technology, China)
- QEI: Query Acceleration Can be Generic and Efficient in the Cloud .385.....
Yifan Yuan (University of Illinois at Urbana-Champaign, USA), Yipeng Wang (Intel Labs, USA), Ren Wang (Intel Labs, USA), Rangeen Basu Roy Chowdhury (Intel Corporation, USA), Charlie Tai (Intel Labs, USA), and Nam Sung Kim (University of Illinois at Urbana-Champaign, USA)
- EXMA: A Genomics Accelerator for Exact-Matching .399.....
Lei Jiang (Indiana University Bloomington, USA) and Farzaneh Zokaei (Indiana University Bloomington, USA)
- Ultra-Elastic CGRAs for Irregular Loop Specialization .412.....
Christopher Torng (Cornell University), Peitian Pan (Cornell University), Yanghui Ou (Cornell University), Cheng Tan (Cornell University), and Christopher Batten (Cornell University)

Memory and Storage Architectures

- GSSA: A Resource Allocation Scheme Customized for 3D NAND SSDs .426.....
Chun-Yi Liu (The Pennsylvania State University), Yunju Lee (The Pennsylvania State University), Wonil Choi (The Pennsylvania State University), Myoungsoo Jung (KAIST), Mahmut Kandemir (The Pennsylvania State University), and Chita Das (The Pennsylvania State University)
- Memristive Data Ranking .440.....
Ananth Krishna Prasad (University of Utah), Morteza Rezaalipour (K.N. Toosi University of Technology), Masoud Dehyadegari (K.N. Toosi University of Technology), and Mahdi Nazm Bojnordi (University of Utah)
- DeACT: Architecture-Aware Virtual Memory Support for Fabric Attached Memory Systems .453
Vamsee Reddy Kommareddy (University of Central Florida, USA), Clayton Hughes (Sandia National Laboratories, USA), Simon David Hammond (Sandia National Laboratories, USA), and Amro Awad (North Carolina State University, USA)

High Throughput Architectures

- Analyzing and Leveraging Decoupled L1 Caches in GPUs .467.....
Mohamed Ibrahim (William & Mary), Onur Kayiran (Advanced Micro Devices, Inc.), Yasuko Eckert (Advanced Micro Devices, Inc.), Gabriel Loh (Advanced Micro Devices, Inc.), and Adwait Jog (William & Mary)

Deadline-Aware Offloading for High-Throughput Accelerators .479.....
Tsung Tai Yeh (National Chiao Tung University), Matthew D. Sinclair (University of Wisconsin), Bradford Beckmann (Advanced Micro Devices, Inc), and Timothy G. Rogers (Purdue University)

LazyBatching: An SLA-Aware Batching System for Cloud Machine Learning Inference .493.....
Yujeong Choi (Korea Advanced Institute of Science and Technology (KAIST), South Korea), Yunseong Kim (Korea Advanced Institute of Science and Technology (KAIST), South Korea), and Minsoo Rhu (Korea Advanced Institute of Science and Technology (KAIST), South Korea)

Dead Page and Dead Block Predictors: Cleaning TLBs and Caches Together .507.....
Chandrashis Mazumdar (Indian Institute of Science, India), Prachatos Mitra (Indian Institute of Science, India), and Arkaprava Basu (Indian Institute of Science, India)

Power Efficiency and Resiliency

ParaDox: Eliminating Voltage Margins via Heterogeneous Fault Tolerance .520.....
Sam Ainsworth (University of Edinburgh, UK), Lionel Zoubritzky (École Normale Supérieure – PSL, France), Alan Mycroft (University of Cambridge, UK), and Timothy M. Jones (University of Cambridge, UK)

CARE: Coordinated Augmentation for Elastic Resilience on DRAM Errors in Data Centers .533..
Jian Chen (Alibaba Group, China), Xiaowei Jiang (Alibaba Group, China), Ying Zhang (Alibaba Group, China), Liyin Liu (Alibaba Group, China), Huifeng Xu (Alibaba Group, China), and Qiang Liu (Alibaba Group, China)

Automatic Microprocessor Performance Bug Detection .545.....
Erick Carvajal Barboza (Texas A&M University), Sara Jacob (Texas A&M University), Mahesh Ketkar (Intel Corporation), Michael Kishinevsky (Intel Corporation), Paul Gratz (Texas A&M University), and Jiang Hu (Texas A&M University)

CAPE: A Content-Addressable Processing Engine .557.....
Helena Caminal (Cornell University), Kailin Yang (Cornell University), Srivatsa Srinivasa (Intel Labs), Akshay Krishna Ramanathan (Pennsylvania State University), Khalid Al-Hawaj (Cornell University), Tianshu Wu (Cornell University), Vijaykrishnan Narayanan (Pennsylvania State University), Christopher Batten (Cornell University), and José F. Martínez (Cornell University)

Systems for Machine Learning 2

SpaceA: Sparse Matrix Vector Multiplication on Processing-in-Memory Accelerator .570.....
Xinfeng Xie (University of California, Santa Barbara, USA), Zheng Liang (University of California, Santa Barbara, USA), Peng Gu (University of California, Santa Barbara, USA), Abanti Basak (University of California, Santa Barbara, USA), Lei Deng (University of California, Santa Barbara, USA), Ling Liang (University of California, Santa Barbara, USA), Xing Hu (University of California, Santa Barbara, USA), and Yuan Xie (University of California Santa Barbara, USA)

Layerweaver: Maximizing Resource Utilization of Neural Processing Units via Layer-Wise Scheduling .584.....
Young H. Oh (Sungkyunkwan University), Seonghak Kim (Seoul National University), Yunho Jin (Seoul National University), Sam Son (Seoul National University), Jonghyun Bae (Seoul National University), Jongsung Lee (Seoul National University), Yeonhong Park (Seoul National University), Dong Uk Kim (Seoul National University), Tae Jun Ham (Seoul National University), and Jae W. Lee (Seoul National University)

Sentinel: Efficient Tensor Migration and Allocation on Heterogeneous Memory Systems for Deep Learning .598.....
Jie Ren (University of California, Merced), Jiaolin Luo (University of California, Merced), Kai Wu (University of California, Merced), Minjia Zhang (Microsoft), Hyeran Jeon (University of California, Merced), and Dong Li (University of California, Merced)

CSCNN: Algorithm-Hardware Co-Design for CNN Accelerators Using Centrosymmetric Filters .612
Jiajun Li (George Washington University), Ahmed Louri (George Washington University), Avinash Karanth (Ohio University), and Razvan Bunescu (University of North Carolina at Charlotte)

Best Paper Nominees

Improving GPU Multi-tenancy with Page Walk Stealing .626.....
Pratheek B (Indian Institute of Science), Neha Jawalkar (Indian Institute of Science), and Arkaprava Basu (Indian Institute of Science)

Stream Floating: Enabling Proactive and Decentralized Cache Optimizations .640.....
Zhengrong Wang (University of California, Los Angeles), Jian Weng (University of California, Los Angeles), Jason Lowe-Power (University of California, Davis), Jayesh Gaur (Intel), and Tony Nowatzki (University of California, Los Angeles)

Prodigy: Improving the Memory Latency of Data-Indirect Irregular Workloads Using Hardware-Software Co-Design .654.....
Nishil Talati (University of Michigan, USA), Kyle May (University of Michigan, USA; University of Wisconsin, Madison, USA), Armand Behroozi (University of Michigan, USA), Yichen Yang (University of Michigan, USA), Kuba Kaszyk (University of Edinburgh, UK), Christos Vasiladiotis (University of Edinburgh, UK), Tarunesh Verma (University of Michigan, USA), Lu Li (University of Edinburgh, UK), Brandon Nguyen (University of Michigan, USA), Jiawen Sun (University of Edinburgh, UK), John Magnus Morton (University of Edinburgh, UK), Agreeen Ahmadi (University of Michigan, USA), Todd Austin (University of Michigan, USA), Michael O'Boyle (University of Edinburgh, UK), Scott Mahlke (University of Michigan, USA; Nvidia Research, USA), Trevor Mudge (University of Michigan, USA), and Ronald Dreslinski (University of Michigan, USA)

P-OPT: Practical Optimal Cache Replacement for Graph Analytics .668.....
Vignesh Balaji (Carnegie Mellon University, USA), Neal Crago (NVIDIA, USA), Aamer Jaleel (NVIDIA, USA), and Brandon Lucia (Carnegie Mellon University, USA)

Network on Chip

- Pitstop: Enabling a Virtual Network Free Network-on-Chip .682.....
Hossein Farrokhbakht (University of Toronto), Henry Kao (University of Toronto), Kamran Hasan (University of Toronto), Paul Gratz (Texas A&M University), Tushar Krishna (Georgia Institute of Technology), Joshua San Miguel (University of Wisconsin-Madison), and Natalie Enright Jerger (University of Toronto)
- BoomGate: Deadlock Avoidance in Non-Minimal Routing for High-Radix Networks .696.....
Gyuyoung Kwauk (KAIST), Seungkwan Kang (KAIST), Hans Kasan (KAIST), Hyojun Son (KAIST), and John Kim (KAIST)
- CHOPIN: Scalable Graphics Rendering in Multi-GPU Systems via Parallel Image Composition .709
Xiaowei Ren (University of British Columbia) and Mieszko Lis (University of British Columbia)
- Adapt-NoC: A Flexible Network-on-Chip Design for Heterogeneous Manycore Architectures .723
Hao Zheng (George Washington University, USA), Ke Wang (George Washington University, USA), and Ahmed Louri (George Washington University, USA)

Emerging Technologies and Applications

- Hardware-Based Address-Centric Acceleration of Key-Value Store .736.....
Chencheng Ye (Huazhong University of Science and Technology), Yuanchao Xu (North Carolina State University), Xipeng Shen (North Carolina State University), Xiaofei Liao (Huazhong University of Science and Technology), Hai Jin (Huazhong University of Science and Technology), and Yan Solihin (University of Central Florida)
- BRIM: Bistable Resistively-Coupled Ising Machine .749.....
Richard Afoakwa (University of Rochester, USA), Yiqiao Zhang (University of Rochester, USA), Uday Kumar Reddy Vengalam (University of Rochester, USA), Zeljko Ignjatovic (University of Rochester, USA), and Michael Huang (University of Rochester, USA)
- An Analog Preconditioner for Solving Linear Systems .761.....
Ben Feinberg (Sandia National Laboratories), Ryan Wong (University of Rochester), Tianyao Xiao (Sandia National Laboratories), Christopher Bennett (Sandia National Laboratories), Jacob Rohan (University of Texas at Austin), Erik Boman (Sandia National Laboratories), Matthew Marinella (Sandia National Laboratories), Sapan Agarwal (Sandia National Laboratories), and Engin Ipek (Qualcomm Inc)
- GCNAX: A Flexible and Energy-Efficient Accelerator for Graph Convolutional Neural Networks.775
Jiajun Li (George Washington University), Ahmed Louri (George Washington University), Avinash Karanth (Ohio University), and Razvan Bunescu (University of North Carolina at Charlotte)

Industry Track 1

- Ascend: A Scalable and Unified Architecture for Ubiquitous Deep Neural Network Computing 789
Heng Liao (Huawei, China), Jiajin Tu (Huawei, China), Jing Xia (Huawei, China), Hu Liu (Huawei, China), Xiping Zhou (Huawei, China), Honghui Yuan (Huawei, China), and Yuxing Hu (Huawei, China)
- Understanding Training Efficiency of Deep Learning Recommendation Models at Scale .802....
Bilge Acun (Facebook AI Research), Matthew Murphy (Facebook), Xiaodong Wang (Facebook), Jade Nie (Facebook), Carole-Jean Wu (Facebook AI Research), and Kim Hazelwood (Facebook AI Research)
- LIBRA: Clearing the Cloud Through Dynamic Memory Bandwidth Management .815.....
Ying Zhang (Alibaba Group, China), Jian Chen (Alibaba Group, China), Xiaowei Jiang (Alibaba Group, China), Qiang Liu (Alibaba Group, China), Ian Steiner (Intel Corporation, USA), Andrew Herdrich (Intel Corporation, USA), Kevin Shu (Intel Corporation, USA), Ripan Das (Intel Corporation, USA), Long Cui (Intel Corporation, USA), and Litrin Jiang (Intel Corporation, USA)
- Eudoxus: Characterizing and Accelerating Localization in Autonomous Machines .827.....
Yiming Gan (University of Rochester), Yu Bo (PerceptIn), Boyuan Tian (University of Rochester), Leimeng Xu (PerceptIn), Wei Hu (PerceptIn), Shaoshan Liu (PerceptIn), Qiang Liu (Tianjin University), Yanjun Zhang (Beijing Institute of Technology), Jie Tang (South China University of Technology), and Yuhao Zhu (University of Rochester)

Industry Track 2

- NeuroMeter: An Integrated Power, Area, and Timing Modeling Framework for Machine Learning Accelerators .841.....
Tianqi Tang (University of California, Santa Barbara), Sheng Li (Google), Lifeng Nai (Google), Norm Jouppi (Google), and Yuan Xie (University of California, Santa Barbara)
- Chasing Carbon: The Elusive Environmental Footprint of Computing .854.....
Udit Gupta (Harvard University/Facebook Inc.), Young Geun Kim (Arizona State University), Sylvia Lee (Facebook Inc.), Jordan Tse (Facebook Inc.), Hsien-Hsin S. Lee (Facebook Inc.), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University), and Carole-Jean Wu (Facebook Inc.)
- Need for Speed: Experiences Building a Trustworthy System-Level GPU Simulator .868.....
Oreste Villa (NVIDIA), Daniel Lustig (NVIDIA), Zi Yan (NVIDIA), Evgeny Bolotin (NVIDIA), Yaosheng Fu (NVIDIA), Niladrish Chatterjee (NVIDIA), Nan Jiang (NVIDIA), and David Nellans (NVIDIA)
- Operating Liquid-Cooled Large-Scale Systems: Long-Term Monitoring, Reliability Analysis, and Efficiency Measures .881.....
Rohan Basu Roy (Northeastern University), Tirthak Patel (Northeastern University), Raj Kettimuthu (Argonne National Laboratory), William Allcock (Argonne National Laboratory), Paul Rich (Argonne National Laboratory), Adam Scovel (Argonne National Laboratory), and Devesh Tiwari (Northeastern University)

Best of CAL

Accelerators for Machine Learning 2

- FuseKNA: Fused Kernel Convolution Based Accelerator for Deep Neural Networks 894
Jianxun Yang (Tsinghua University, China), Zhao Zhang (Tsinghua University, China), Zhuangzhi Liu (Tsinghua University, China), Jing Zhou (Tsinghua University, China), Leibo Liu (Tsinghua University, China), Shaojun Wei (Tsinghua University, China), and Shouyi Yin (Tsinghua University, China)
- FAFNIR: Accelerating Sparse Gathering by Using Efficient Near-Memory Intelligent Reduction.. 908
Bahar Asgari (Georgia Institute of Technology), Ramyad Hadidi (Georgia Institute of Technology), Jiashen Cao (Georgia Institute of Technology), Da Eun Shim (Georgia Institute of Technology), Sung-Kyu Lim (Georgia Institute of Technology), and Hyesoon Kim (Georgia Institute of Technology)
- VIA: A Smart Scratchpad for Vector Units with Application to Sparse Matrix Computations 921
Julian Pavon (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya), Ivan Vargas Valdivieso (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya), Adrian Barredo (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya), Joan Marimon (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya), Miquel Moreto (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya), Francesc Moll (Universitat Politecnica de Catalunya), Osman Unsal (Barcelona Supercomputing Center), Mateo Valero (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya), and Adrian Cristal (Barcelona Supercomputing Center; Universitat Politecnica de Catalunya)

Author Index