

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Aamer Jaleel

October 16, 2016 (Sunday)	
18:00-20:00 Reception	
October 17, 2016 (Monday)	
7:00-8:00 Breakfast	
8:00-8:20 Opening remarks	
8:20-9:20	Keynote I: Internet of Things, Hologram and Hyper-Technology and Policy Margaret Martonosi (University of Maryland)
9:20-10:00 Lightning Session I	
10:00-10:20 Break	
Session 1a: Microarchitecture	Session 1b: Cloud & Storage
Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. Binabandhan Panda (NVIDIA Research), Anshu Saxena (NVIDIA Research)	SAFEB: Atomic Object Reads for In-Memory Rack-Scale Computing. Alexander Daglis (EPFL), Dimitri Usigov (EPFL), Marko Novakovic (EPFL), Eduard Bugren (EPFL), Bekah Fahedi (EPFL), Boris Grosz (University of Michigan)
Perception Learning for Reuse Prediction. Uthra Iyeran (Texas A&M University), Zhe Wang (Intel Labs), Daniel A. Jimenez (Texas A&M University)	A Cloud-Scale Acceleration Architecture. Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angeles (Microsoft), Jeremy Iosseev (Microsoft Research), Michael Heuleman (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joon-Young Kim (Microsoft Research), Daniel Li (Microsoft Research), Jodi Maswengil (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papernschall (Microsoft Research), Lin Woods (Microsoft Research), Sitaram Lenka (Microsoft), Derek Chou (Microsoft), Doug Burger (Microsoft Research)
pTask: A Smart Prefetching Scheme for OS Intensive Applications. Prathmesh Kulkarni (IIIT Delhi), Suresh R. Sarangi (IIIT Delhi)	Towards Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. Yang Hu, and Tao Li (University of Florida)
Register Sharing for Equality Prediction. Arthur Perlas (INRIA), Fernando A. Indio (INRIA), André Seneac (INRIA)	Improving the I/O Performance Gap for Big Data Workloads: A New NVMMIO-based Approach. Benjie Chen (The Hong Kong Polytechnic University), Jia Shao (The Hong Kong Polytechnic University), Tao Li (NDU (University of Florida))
Data-Centric Execution of Speculative Parallel Programs. Mark C. Jeffrey (Massachusetts Institute of Technology), Suvraj Subramanian (Massachusetts Institute of Technology), Mahesh Aravamudan (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)	NeSC: Self-Virtualizing Nested Storage Controller. Yonatan Gutfreund (Technion), Yoav Eason (Technion)
12:00-14:00 Lunch	
14:00-15:40 Poster session	
15:40-16:00 Break	
Session 2a: GPU	Session 2b: Neural Networks
MIMD Synchronization on SIMD Architectures. Ahmad Alilawati (The University of British Columbia), Tar M. Amoadit (The University of British Columbia)	From High-Level Deep Neural Models to FPGA. Harsh Sharma (Georgia Institute of Technology), Jorgo Park (Georgia Institute of Technology), Doyu Mahajan (Georgia Institute of Technology), Emmanuel Anato (Georgia Institute of Technology), Jason Kyung Kim (Georgia Institute of Technology), Chawika Shao (Georgia Institute of Technology), Aadu Mohita (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)
Efficient Kernel Synthesis for Performance Portable Programming. Li-Wen Chang (University of Illinois at Urbana-Champaign), Izat El Hag (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Palo Alto Research Labs), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	VDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Hwu (NVIDIA), Natalia Gromakova (NVIDIA), Jason Clemens (NVIDIA), Arden Zilber (NVIDIA), Stephen W. Keckler (NVIDIA)
KLIP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism. Izat El Hag (University of Illinois at Urbana-Champaign), Juan Gómez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dejan Misic (Pewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	Combustion-X: An Accelerator for Sparse Neural Networks. Shih-Zhen Institute of Computing Technology, CAS, Ziding Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huaying Lan (Institute of Computing Technology, CAS), Shaoh Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Henshi Chen (Institute of Computing Technology, CAS), Yunfeng Chen (Institute of Computing Technology, CAS)
Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPGPUs. Naifeng Jing (Shanghai Jiao Tong University), Jianfeng Wang (Shanghai Jiao Tong University), Fengfeng Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Li Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyan Liang (Shanghai Jiao Tong University)	Zorus: A Holistic Approach to Resource Utilization in GPUs. Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gerard Pflueger (Carnegie Mellon University), Semra Khan (University of Virginia), Akshith Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Chiu-Ming Li (Carnegie Mellon University)
GRAPE: Minimizing Energy for GPU Applications with Performance Requirements. Muhammad Husni Sanjaya (Surya University), Henry Hoffmann (University of Michigan)	NEUTRAMS: Neural Network Transformation and Co-design under Neuroeconomic Hardware Constraints. Yu Ji (Tsinghua Univ.), Youfuo Zhang (Tsinghua Univ.), Shuangchen Li (UCSB), Ping Chi (UCSB), Cihang Jiang (Tsinghua Univ.), Peng Du (Tsinghua Univ.), Yuan Xu (UCSB), Wenqiang Chen (Tsinghua Univ.)
18:00-20:00 Business meeting	

October 18, 2016 (Tuesday)	
7:00-8:30 Breakfast	
8:30-9:30	Keynote II: Low Power (LP), Low Mobile to Wearable & IoT (Lemo IoT - Media Lab)
9:30-10:10 Lightning Session II	
10:10-10:20 Break	
Session 3a: Compilation & Memory	Session 3b: Interconnect
Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Armesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)	OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Iw Zhan (UCSB), Omer Kayran (AMD Research), Gabriel H. Lee (AMD Research), Chih-R. Dow (PSU), Yuan Xie (UCSB)
CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jekuty (University of Michigan), Daniel Rings (University of Michigan), Nathan Barnes (University of Michigan), Michael Laurenzano (University of Michigan), Lingjie Tang (University of Michigan), Jason Mars (University of Michigan)	A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Jia Zhan (UCSB), Sir Akshay (UCSB), Jinhui Zhao (UCSB), Ai Devis (HP Labs), Paolo Farabochi (HP Labs), Yungang Wang (Huawei), Yuan Xie (UCSB)
Low-Cost Soft Error Resilience with Unified Data Verification and Free-Generated Recovery for Acoustic Sensor Based Detection. Qingyu Lu (Virginia Tech), Changhui Jung (Virginia Tech), Dongyong Lee (Virginia Tech), Dewesh Imman (Oak Ridge National Lab)	Content-based Congestion Management in Data-Centric Networks. Chengxin Kim (KAIST), Chengxin Kim (KAIST), Hyun Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)
Lazy Release Consistency for GPUs. Johnathan Akrop (University of Illinois at Urbana-Champaign), Ahmad Hameed, Man S. Cho (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research)	Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Dominic DiCiccio (Ohio University), Brian Jordan (Ohio University), Avinash Kati (Ohio University), Ahmed Elmaghrabi (George Washington University)
Improving Energy Efficiency of DRAM by Exploiting Half Page Row Access. Heorhan He (Stanford University), Arslan Padhan (Stanford University), Stephen Richardson (Stanford University), Shaoh Kwiatkowsky (Technion), Mark Horowitz (Stanford University)	Reducing Data Movement Energy via Offline Data Clustering and Encoding. Shihong Wang (University of Rochester), Engin Ipek (University of Rochester)
12:10-14:10 Award Lunch (including Bob Hsu Award, best of Time)	
Session 4a: Multicore	Session 4b: Security
Racer: TSO Consistency via Race Detection. Alberto Iles (Universidad de Murcia), Stefano Karavas (Alibaba University)	Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors. Mikhail Kuznetsov (University of Texas at Austin), Vijay Janapa Reddy (University of Texas at Austin), Minh Toan (University of Texas at Austin)
Exploiting Semantic Commutativity in Hardware Speculation. Guozuo Zhang (MIT CSAIL), Virginia Chu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)	Poinlessly Safe Speculation for Secure Memory. Ihsan Subergilz Lehman (Duke University), Andrew D. Pfitz (Duke University), Benjamin C. Lee (Duke University)
CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. Cheachen Chou (Georgia Tech), Aamer Jalali (NVIDIA), Masumuddin K. Qureshi (Georgia Tech)	ReplyConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay. Mengyao Yan (University of Illinois at Urbana-Champaign), Joseph Torrellas (University of Illinois at Urbana-Champaign)
CSIT: Mitigating the NUMA Bottleneck via Coherent DRAM Caches. Cheng-Chieh Huang (University of Edinburgh), Balashankar Kumar (University of Edinburgh), Marco Isber (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)	Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. Urvay Vaidyanath (SIUPT Binghamton), Dmitry Porozovnev (SIUPT Binghamton), Nael Abu-Ghazwan (SIU Binghamton)
15:30-16:00 Break	
Session 5a: Approximate Computing	Session 5b: Accelerators 1
Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. Aniranth Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shih-Chieh Lin (University of Michigan, Ann Arbor), Muneeb Khan (Ippawa University), Md E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)	HARE: Hardware Accelerator for Regular Expressions. Weibiao Gogte (University of Michigan), Aashesh Koli (University of Michigan), Michael J. Lafferty (University of Michigan), Lena D'Amore (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan)
Approxlyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency. Radha Venkatesan (University of Illinois at Urbana-Champaign), Abhishek Maheshwari (University of Illinois at Urbana-Champaign), Siva Kumar Sastry Hari (Nvidia), Santia V. Adve (University of Illinois at Urbana-Champaign)	The Microarchitecture of a Real-time Robot Motion Planning Accelerator. Sean Murray (Duke University), Will Floyd-Jones (Duke University), Ying Qi (Duke University), George Konstantis (Duke University), Daniel J. Sorin (Duke University), Thomas F. Wenisch (University of Michigan)
Efficient Data Supply for Hardware Accelerators with Prefetching and Access/Execute Decoupling. Joo Chen (Cornell University), G. Edward Suh (Cornell University)	The Bunker Cache for Spatio-Value Approximation. Joshua San-Miguel (University of Toronto), Jorge Albarracín (University of Toronto), Natalie Enright Jerger (University of Toronto), Aamer Jalali (NVIDIA)
17:00-21:00 Banquet	

October 19, 2016 (Wednesday)	
7:00-8:00 Breakfast	
Session 6a: Accelerators 2	Session 6b: Mobile & Power Mgmt
An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Reza Yazdani Ardebilabad (Universidad Politécnica de Catalunya (UPC)), Albert Sarguch (Universidad Politécnica de Catalunya (UPC)), Jose-María Armau (Universidad Politécnica de Catalunya (UPC)), Antonio González (Universidad Politécnica de Catalunya (UPC))	Evaluating Programmable Architectures for Imaging and Vision Applications. Arslan Vassily (Stanford), Nishu Shrivastava (Stanford), Arslan Pedram (Stanford), Stephen Richardson (Stanford), Shaoh Kwiatkowsky (Technion), Mark Horowitz (Stanford)
Co-Designing Accelerators and SoC Interfaces using gem5-SoC. Hamed Ghosh (Harvard University), Vijayalakshmi Srivastava (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)	Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users. Keyue Fan (University of Houston), Xinyang Zhao (University of Houston), Jengwee Yan (University of Houston), Xin Fu (University of Houston)
8:00-9:40 CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Anurag Shrivastava (Simon Fraser University), Srinivasan Kumar (Simon Fraser University), Apala Gulur (Simon Fraser University), Aravindh Shrivastava (Simon Fraser University)	Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks. Dimitrios Skafarots (SIUC), Benj Thomas (Intel), Aditya Agrawal (Nvidia), Shihon Qin (SIUC), Robert Pinnau-Podgorski (SIUC), Uwe H. Karpuzcu (IBM), Radu Ionescu (CSI), Nam Sung Kim (SIUC), Joseph Torrellas (SIUC)
Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Mohammadian (SIUC), Young-Hoon Son (SIUC), Jung Ho Ahn (SIUC), Nam Sung Kim (SIUC)	Throttle: Processor Power Management in the Temperature Inversion Region. Yichun Zu (University of Texas at Austin), Wei Huang (AMD Research), Indran Paul (AMD Research), Vijay Anantha Reddi (University of Texas at Austin)
9:40-10:00 Break	
Session 7: Best Paper Candidates	
Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. Lee Jun-Hwi (Piscataway University), Lina Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation)	
Improving Bank-Level Parallelism for Irregular Applications. Xulong Jiang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapati (Penn State), Agadesh Kotov (Penn State)	
Delegated Perist Order. Aashesh Koli (University of Michigan), Jeff Rosen (Intel/Oracle Computing), Stephen Dieckhoff (IBM), Ali Saedi (IBM), Steven Halay (Intel/Oracle Computing), Sheng Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan)	
Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emissions. Nadia Sebhatkhalid (Georgia Tech), Alexia Nazari (Georgia Tech), Aleksei Zayt (Georgia Tech), Minh Phuc (Georgia Tech)	
Path Confidence based Lookahead Prefetching. Jinchun Kim (Texas A&M University), Seth H. Pugsley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Narasimha Reddy (Texas A&M University), Chris Willerson (Intel Labs), Anshu Chandra (Intel Labs), Omer Mutlu (CMU), Yafe N. Puri (UIUC Austin)	
Session 8: Conference Closing and Best Paper Award	
12:30-18:15 Conference Outing (includes sack lunch)	

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Nataie Enright Jerger and Aamer Jalali

October 16, 2016 (Sunday)	
18:00-20:00 Reception	
October 17, 2016 (Monday)	
7:00-8:00 Breakfast	
8:00-8:20 Opening remarks	
8:20-9:20	Keynote 1: Internet of Things, Mobile and Edge Technology and Policy Margaret Martonosi (University of Maryland)
9:20-10:00 Lightning Session I	
10:00-10:20 Break	
Session 1a: Microarchitecture	Session 1b: Cloud & Storage
Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. Binabandhan Panda (NVIDIA Research), Andri Sasmita (NVIDIA Research)	SABRe: Atomic Object Reads for In-Memory Rack-Scale Computing. Alexander Daglis (EPFL), Dimitri Uzunoglu (EPFL), Marko Novakovic (EPFL), Eduard Bugren (EPFL), Bekah Fahedi (EPFL), Boris Grosz (University of Michigan)
Perception Learning for Reuse Prediction. Uthra Iyer (Iowa State University), Zhe Wang (Intel Labs), Daniel A. Jimenez (Iowa State University)	A Cloud-Scale Acceleration Architecture. Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angeles (Microsoft), Jeremy Lossen (Microsoft Research), Michael Haselman (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joon-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Jodi Maswergil (Microsoft Research), Kain Dytchovov (Microsoft Research), Michael Pappaschall (Microsoft Research), Lisa Wu (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chou (Microsoft), Doug Burger (Microsoft Research)
pTask: A Smart Prefetching Scheme for OS Intensive Applications. Prathmesh Kulkarni (IIIT Delhi), Suresh R. Sarangi (IIIT Delhi)	Networks Efficient Server Architecture for Virtualized Workload Function Deployment: Implications and Implementations. Yang Hu, and Ian Li (University of Florida)
Register Sharing for Equality Prediction. Arthur Peres (IBM), Fernando A. Endo (NRA), Andre Seneo (NRA)	Bridging the I/O Performance Gap for Big Data Workloads: A New NVDIMM-based Approach. Renjie Chen (The Hong Kong Polytechnic University), Ai Shao (The Hong Kong Polytechnic University), Tao Li (PDU (University of Florida))
Data-Centric Execution of Speculative Parallel Programs. Mark C. Jeffrey (Massachusetts Institute of Technology), Suvraj Subramanian (Massachusetts Institute of Technology), Mahesh Anayelapala (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)	NeSC: Self-Virtualizing Nested Storage Controller. Yonatan Gutfreund (Technion), Yoav Eason (Technion)
12:00-14:00 Lunch	
14:00-15:40 Poster session	
15:40-16:00 Break	
Session 2a: GPU	Session 2b: Neural Networks
MIMD Synchronization on SIMT Architectures. Ahmad Bilitiyoti (The University of British Columbia), Sri M. Amootil (The University of British Columbia)	From High-Level Deep Neural Models to FPGA. Herdik Sharma (Georgia Institute of Technology), Jorge Park (Georgia Institute of Technology), Doyu Mahajan (Georgia Institute of Technology), Ermanno Airota (Georgia Institute of Technology), Joon Myung Kim (Georgia Institute of Technology), Chawika Shao (Georgia Institute of Technology), Ashi Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)
Efficient Kernel Synthesis for Performance Portable Programming. Li-Wen Chang (University of Illinois at Urbana-Champaign), Izat El Hag (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Purdue America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	VDRN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Hwu (NVIDIA), Natalia Gromkova (NVIDIA), Jason Clemens (NVIDIA), Arden Zilber (NVIDIA), Stephen W. Koehler (NVIDIA)
KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism. Izat El Hag (University of Illinois at Urbana-Champaign), Juan Gómez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dejan Mijovic (Purdue-Parkland Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	Synapse: Bit-Serial Deep Neural Network Computing. Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Taylor Hetherington (University of British Columbia), Sri Amootil (University of British Columbia), Andrius Moshovos (University of Toronto)
Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPGPUs. Naifeng Jing (Shanghai Jiao Tong University), Jianfei Wang (Shanghai Jiao Tong University), Fengting Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Liang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyan Liang (Shanghai Jiao Tong University)	Combustion-X: An Accelerator for Sparse Neural Networks. Shih-Zhen Institute of Computing Technology, CAS, Ziding Gu (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huaying Lan (Institute of Computing Technology, CAS), Shaoh Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Henshi Chen (Institute of Computing Technology, CAS), Yong Chen (Institute of Computing Technology, CAS)
Zorus: A Holistic Approach to Resource Virtualization in GPUs. Ganendra Venkatesh (Georgia Tech University), Kevin Shah (Georgia Tech University), Gerald P. Morrison (Georgia Tech University), Samira Khan (Georgia Tech University), Adithi Sivaram (Georgia Tech University), Chong-Gee Yong (Georgia Tech University), Adnan Iqbal (Georgia Tech University), William and Maria, Philip S. Gibbons (Georgia Tech University), Ghim Hee Cho (Georgia Tech University)	NEURONs: Neural Network Transformation and Co-Design for Accelerating Deep Learning. Yuhui Chen (University of Illinois at Urbana-Champaign), Li Li (University of Illinois at Urbana-Champaign), Peng Du (University of Illinois at Urbana-Champaign), Sanku Kim (University of Illinois at Urbana-Champaign)
GRAP: Minimizing Energy for GPU Applications. Performance Requirements, Muhammad Hammad (University of Toronto), Henry Holbrook (University of Toronto)	NEURONs: Neural Network Transformation and Co-Design for Accelerating Deep Learning. Yuhui Chen (University of Illinois at Urbana-Champaign), Li Li (University of Illinois at Urbana-Champaign), Peng Du (University of Illinois at Urbana-Champaign), Sanku Kim (University of Illinois at Urbana-Champaign)
18:00-20:00 Business meeting	

(20 min paper)

October 18, 2016 (Tuesday)	
7:00-8:30 Breakfast	
8:30-9:30	Keynote 1: Low Power (CPU, GPU, Mobile to Wearable & IoT) Urooj K. Medala (Intel)
9:30-10:10 Lightning Session II	
10:10-10:20 Break	
Session 3a: Compilation & Memory	Session 3b: Interconnect
Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Armesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)	OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Iw Zhan (UCSB), Omur Kayran (AMD Research), Gabriel H. Luo (AMD Research), Chite R. Das (PSU), Yuan Xie (UCSB)
CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jekuty (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjie Tang (University of Michigan), Jason Mars (University of Michigan)	A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Ja Zhan (UCSB), Sir Agkath (UCSB), Jinhui Zhao (UCSB), Ai Davis (HP Labs), Paolo Farabochi (HP Labs), Yungang Wang (Huawei), Yuan Xie (UCSB)
Low-Cost Soft Error Resilience with Unified Data Verification and Free-Generated Recovery for Acoustic Sensor Based Detection. Qingyu Lu (Virginia Tech), Changhui Jung (Virginia Tech), Dongyong Lee (Virginia Tech), Dewesh Imman (Oak Ridge National Lab)	Content-based Congestion Management in Large-Scale Networks. Ganggang Chen (KAIST), Chengqun Kim (KAIST), Ryan Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)
Lazy Release Consistency for GPUs. Johnathan Akrop (University of Illinois at Urbana-Champaign, AMD Research), Man S. Cho (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research)	Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Dominic Dimouzis (Ohio University), Brian Bonder (Ohio University), Avinash Kati (Ohio University), Ahmed Elson (George Washington University)
Improving Energy Efficiency of DRAM by Exploiting Half Page Row Access. Hoang Ha (Stanford University), Arduwan Padman (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)	Reducing Data Movement Energy via Offline Data Clustering and Encoding. Shihong Wang (University of Rochester), Engin Ipek (University of Rochester)
12:10-14:10 Award Lunch (including Bob Hwu Award, best of Time)	
Session 4a: Multicore	Session 4b: Security
Racer: TSO Consistency via Race Detection. Alberto Iles (Universidad de Murcia), Stefano Karavas (Alibaba University)	Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors. Michael Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Minh Toan (University of Texas at Austin)
Exploiting Semantic Commutativity in Hardware Speculation. Guozuo Zhang (MIT CSAIL), Virginia Chu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)	Polynology: Safe Speculation for Secure Memory. Iyanna Subergeldi Lehman (Duke University), Andrew D. Hillston (Duke University), Benjamin C. Lee (Duke University)
CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. Cheechen Chau (Georgia Tech), Aamer Jaleel (NVIDIA), Muzammid K. Qureshi (Georgia Tech)	ReplayConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay. Mengyao Yan (University of Illinois at Urbana-Champaign), Joseph Torrellas (University of Illinois at Urbana-Champaign)
MITting the NUMA Bottleneck via Coherent DRAM Caches. Cheng-Chieh Huang (University of Edinburgh), Sakshil Kumar (University of Edinburgh), Marco Jure (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)	Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. Urvay Vidyashankar (SUNY Binghamton), Omir Porrovecchi (SUNY Binghamton), Nael Abu-Ghazwan (SUNY Binghamton)
15:30-16:00 Break	
Session 5a: Approximate Computing	Session 5b: Acceleration 1
Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. Aniranth Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shih-Chieh Lin (University of Michigan, Ann Arbor), Muneeb Khan (Ippolito University), Md E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Isaac M. Aberkane (University of Michigan, Ann Arbor)	HARE: Hardware Accelerator for Regular Expressions. Weibiao Gogte (University of Michigan), Aashesh Koli (University of Michigan), Michael J. Lafferty (University of Michigan), Lena D'Antonio (University of Wisconsin-Madison), Jennifer F. Hewitt (University of Michigan)
Approxifyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency. Radha Venkatesh (University of Illinois at Urbana-Champaign), Abdulhamid Mahmood (University of Illinois at Urbana-Champaign), Siva Kumar (Sany Tech (Beijing), Xianke V. Adve (University of Illinois at Urbana-Champaign)	The Bunker Cache for Spatio-Value Approximation. Joshua San-Miguel (University of Toronto), Jorge Albericio (University of Toronto), Subodhrajeev Jeevan (University of Toronto), Aamer Jaleel (NVIDIA)
17:00-21:00 Reception	

October 19, 2016 (Wednesday)	
7:00-8:00 Breakfast	
Session 6a: Accelerators 2	Session 6b: Mobile & Power Mgmt
An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Reza Yazdani Ardebilabad (Universidad Politécnica de Catalunya (UPC)), Albert Sangra (Universidad Politécnica de Catalunya (UPC)), Jose-María Arsuaga (Universidad Politécnica de Catalunya (UPC)), Antonio González (Universidad Politécnica de Catalunya (UPC))	Evaluating Programmable Architectures for Imaging and Vision Applications. Arvind Venkitesh (Stanford), Nishu Shrivastava (Stanford), Arduwan Padman (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)
Co-Designing Accelerators and SoC Interfaces using gem5-BooTwin. Yuhui Chen (Stanford University), Sam (Liang) Jui (Harvard University), David Brooks (Harvard University)	Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users. Keyue Fan (University of Houston), Xinyang Zhang (University of Houston), Jengwee Jia (University of Houston), Xin Fu (University of Houston)
CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Armesh Jain (University of Michigan), Srinivasan Kumar (Simon Fraser University), Apala Gulra (Simon Fraser University), Aravind Shrivastava (Simon Fraser University)	Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 2D Stacks. Simos Sifalopoulos (SUNY), Benji Thomas (Intel), Aditya Agrawal (Nvidia), Shihui Qin (ARM), Robert Pinnau-Podgurski (SUNY), Ulye H. Kapteina (ARM), Radu Ionescu (DSU), Nam Sung Kim (SUNY), Joseph Torrellas (SUNY)
Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Moghaddam (SUNY), Young-Hoon Son (SNU), Jung Ho Ahn (SNU), Nam Sung Kim (SUNY)	Throttle: Processor Power Management in the Temperature Inversion Region. Yichou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indranil Paul (AMD Research), Vijay Janapa Reddi (University of Texas at Austin)
8:00-9:40	
9:40-10:00 Break	
Session 7: Best Paper Candidates	
Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. Lee Jun Hwi (Pinceton University), Lina Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation)	Path Confidence based Lookahead Prefetching. Jinchun Kim (Iowa State University), Seth H. Pugdley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Nareesha Reddy (Texas A&M University), Chris Wilkinson (Intel Labs), Zishan Cheng (Intel Labs)
Improving Bank-Level Parallelism for Irregular Applications. Xulong Jiang (Penn State), Mahmut Kandemir (Penn State), Proveen Yedlapati (Penn State), Agadesh Kotov (Penn State)	Throttle: Processor Power Management in the Temperature Inversion Region. Yichou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indranil Paul (AMD Research), Vijay Janapa Reddi (University of Texas at Austin)
Delegated Perist Ordering. Aashesh Koli (University of Michigan), Jeff Hosen (Intel/Oracle Computing), Stephen Duesterhorst (AIM), Ali Saedi (AIM), Steven Palay (Intel/Oracle Computing), Shang Lu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan)	
Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emissions. Nader Shehatahkhah (Georgia Tech), Alexea Nazari (Georgia Tech), Aleksei Zap (Georgia Tech), Miles Pruckner (Georgia Tech)	
Path Confidence based Lookahead Prefetching. Jinchun Kim (Iowa State University), Seth H. Pugdley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Nareesha Reddy (Texas A&M University), Chris Wilkinson (Intel Labs), Zishan Cheng (Intel Labs)	
Continuous Runahead: Transparent Hardware Accelerator for Memory Intensive Workloads. Mahesh Heman (UT Austin), Omar Mutlu (CMU), Yuhui Chen (UT Austin)	
Session 8: Conference Closing and Best Paper Award	
12:30-18:15 Conference Outing (includes sack lunch)	

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalia Enrique Jerger and Aamer Jaleel

October 16, 2016 (Sunday)	
18:00-20:00 Reception	
October 17, 2016 (Monday)	
7:00-8:00 Breakfast	
8:00-8:20 Opening remarks	
8:20-9:20	Keynote I: Internet of Things, Mobile and Edge Technology and Policy Margaret Martonosi (University of Wisconsin)
9:20-10:00 Lightning Session I	
10:00-10:20 Break	
Session 1a: Microarchitecture	Session 1b: Cloud & Storage
Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. Binabandhan Panda (NVIDIA Research), Anshu Saxena (NVIDIA Research)	SABRe: Atomic Object Reads for In-Memory Rack-Scale Computing. Alexander Daglis (EPFL), Dimitri Usakov (EPFL), Marko Novakovic (EPFL), Eduard Bugren (EPFL), Bekob Fahedi (EPFL), Boris Grosz (University of Edinburgh)
Perception Learning for Reuse Prediction. Huiyu Lian (Intel ASIM University), Zhe Wang (Intel Labs), Daniel A. Jimenez (Intel ASIM University)	A Cloud-Scale Acceleration Architecture. Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Anepudi (Microsoft), Jeremy Lossen (Microsoft Research), Michael Haselman (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joon-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Jodi Maswengil (Microsoft Research), Kain Dytchewich (Microsoft Research), Michael Pappaschall (Microsoft Research), Lin Wu (Microsoft Research), Shyam Lanka (Microsoft), Derek Chou (Microsoft), Qing Burger (Microsoft Research)
pTask: A Smart Prefetching Scheme for OS Intensive Applications. Prathmesh Kulkarni (IIIT Delhi), Sumit R. Sanzgiri (IIIT Delhi)	Towards Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. Yang Hu, and Tao Li (University of Florida)
Register Sharing for Equality Prediction. Arthur Peres (IBM), Fernando A. Endo (NIRA), Andre Sampaio (NIRA)	Bridging the I/O Performance Gap for Big Data Workloads: A New NVDIMM-based Approach. Renjie Chen (The Hong Kong Polytechnic University), Tao Li (FSU (University of Florida))
Data-Centric Execution of Speculative Parallel Programs. Mark C. Jeffrey (Massachusetts Institute of Technology), Suvraj Subramanian (Massachusetts Institute of Technology), Mahesh Anshu (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)	Ne5C: Self-Virtualizing Nested Storage Controller. Yonatan Gutfreund (Microsoft), Yusef Ezzam (Technion)
12:00-14:00 Lunch	
14:00-15:40 Poster session	
15:40-16:00 Break	
Session 2a: GPU	Session 2b: Neural Networks
MIMD Synchronization on SIMT Architectures. Ahmad Bilalatty (The University of British Columbia), Ter M. Aamodi (The University of British Columbia)	From High-Level Deep Neural Models to FPGA. Harsh Sharma (Georgia Institute of Technology), Jorgae Park (Georgia Institute of Technology), Doyu Mahajan (Georgia Institute of Technology), Emmanuel Anato (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chawki Shao (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)
Efficient Kernel Synthesis for Performance Portable Programming. Li-Wen Chang (University of Illinois at Urbana-Champaign), Izat El Hag (University of Illinois at Urbana-Champaign), Christopher Rodriguez (Intel America Research Lab), Juan Gomez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	VDRN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Hwu (NVIDIA), Natalia Gromkova (NVIDIA), Jason Clemens (NVIDIA), Arden Zilic (NVIDIA), Stephen Wu, Ivo Ilic (NVIDIA)
KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism. Izat El Hag (University of Illinois at Urbana-Champaign), Juan Gomez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dejan Mijovic (Phyletel-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	Stripes: Bit-Serial Deep Neural Network Computing. Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Taylor Hetherington (University of British Columbia), Ter Aamodi (University of British Columbia), Andrew Moshchov University of Toronto)
16:00-18:00	Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPGPUs. Naifeng Jing (Shanghai Jiao Tong University), Jianfei Wang (Shanghai Jiao Tong University), Fengfeng Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Liang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyan Liang (Shanghai Jiao Tong University), Zorus: A Holistic Approach to Resource Virtualization in GPUs. Venkatesh Visvakumar (Georgia Tech University), Kevin Ghemawat (Georgia Tech University), Ganesh P. Venkatesh (Georgia Tech University), Sumeet Mittal (Georgia Tech University), Arshad Ahmad (Georgia Tech University), Arshad Ahmad (Georgia Tech University), William and Maria, Philip S. Gibbons (Georgia Tech University), Shuangping Chen (Georgia Tech University)
18:00-20:00 Business meeting	

(20 min) (61 papers)
paper MICRO

October 18, 2016 (Tuesday)	
7:00-8:30 Breakfast	
8:30-9:30	Keynote I: Low Power CPUs: from Mobile to Wearable & IoT Urolog Koz (MediaLabs)
9:30-10:10 Lightning Session II	
10:10-10:30 Break	
Session 3a: Compilation & Memory	Session 3b: Interconnects
Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Armesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)	OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Iw Zhan (UCSB), Omur Kayran (AMD Research), Gabriel H. Lee (AMD Research), Chite R. Das (PSU), Yuan Xie (UCSB)
CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jekuty (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjie Tang (University of Michigan), Jason Mars (University of Michigan)	A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Ja Zhan (UCSB), Sir Akshay (UCSB), Jinhui Zhao (UCSB), Al Devis (HP Labs), Paolo Farabochi (HP Labs), Yuanqiang Wang (Huawei), Yuan Xie (UCSB)
10:30-12:10	Content-based Congestion Management in Large-Scale Networks. Guozeng Chen (KAUST), Chengyuan Kim (KAUST), Ryan Jiang (KAUST), Mike Parker (Intel), John Kim (KAUST)
Low-Cost Soft Error Resilience with Unified Data Verification and Free-Generated Recovery for Acoustic Sense Based Detection. Qingqi Lu (Virginia Tech), Changhui Jung (Virginia Tech), Dongyong Lee (Virginia Tech), Dewesh Siman (Oak Ridge National Lab)	Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Dominic Dimic (Ohio University), Brian Jordan (Ohio University), Avinash Kati (Ohio University), Ahmed Elson (George Washington University)
Lazy Release Consistency for GPUs. Jeevanth Akhup (University of Illinois at Urbana-Champaign), Ameer Alwehedi, Man S. Cho (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research)	Reducing Data Movement Energy via Offline Data Clustering and Encoding. Shihong Wang (University of Rochester), Engin Ipek (University of Rochester)
12:10-14:10	Improving Energy Efficiency of DRAM by Exploiting Half Page Row Access. Hoang Ha (Stanford University), Arduwan Padman (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)
Session 4a: Multicore	Session 4b: Security
Racer: TSO Consistency via Race Detection. Alberto Iles (Universidad de Murcia), Stefano Karavas (Alibaba University)	Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors. Michael Kasdagli (University of Texas at Austin), Vijay Janapa Reddy (University of Texas at Austin), Minh Toan (University of Texas at Austin)
Exploiting Semantic Commutativity in Hardware Speculation. Guozhen Zhang (MIT CSAIL), Virginia Chu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)	Polynoid: Safe Speculation for Secure Memory. Iyanna Subergeldi Lehman (Duke University), Andrew D. Hillson (Duke University), Benjamin C. Lee (Duke University)
14:10-15:10	Reply/Confusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay. Mengyao Yan (University of Illinois at Urbana-Champaign), Joseph Torrellas (University of Illinois at Urbana-Champaign)
CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. Cheechee Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Muzammid K. Qureshi (Georgia Tech)	Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. Urvay Vidyashankar (SUNY Binghamton), Omry Perlmutter (SUNY Binghamton), Naveel Abu-Ghazwan (SUNY Binghamton)
15:30-16:00 Break	
Session 5a: Approximate Computing	Session 5b: Accelerators 1
Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. Aniranth Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shih-Chieh Lin (University of Michigan, Ann Arbor), Muneeb Khan (Ippolito University), Md E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Isaac M. Aberkane (University of Michigan, Ann Arbor)	HARE: Hardware Accelerator for Regular Expressions. Weibiao Gupte (University of Michigan), Aashesh Koli (University of Michigan), Michael J. Laflamme (University of Michigan), Lora D'Antonio (University of Wisconsin-Madison), Jennifer R. Burch (University of Wisconsin-Madison)
16:00-17:00	The Microarchitectural Design of a Real-time Robot Motion Planning Accelerator. Sam Murray (Duke University), Wei-Feng Jones (Duke University), King-qi (Duke University), George Vamvakos (Duke University), Daniel A. Stern (Duke University)
Approximix: Towards A Systematic Framework for Instruction-Level Approximate Computing and Its Application to Hardware Resiliency. Radha Venkatesh (University of Illinois at Urbana-Champaign), Abdulmalik Abd-Elghany (University of Illinois at Urbana-Champaign), Uza Khan, Arshad V. Arshad (University of Illinois at Urbana-Champaign)	Efficient Data Supply for Hardware Accelerators with Prefetching and Access Entropy Decoupling. Ion Ionescu (University of Edinburgh), Edward Slob (Cornell University)
Approximix: Towards A Systematic Framework for Instruction-Level Approximate Computing and Its Application to Hardware Resiliency. Radha Venkatesh (University of Illinois at Urbana-Champaign), Abdulmalik Abd-Elghany (University of Illinois at Urbana-Champaign), Uza Khan, Arshad V. Arshad (University of Illinois at Urbana-Champaign)	

October 19, 2016 (Wednesday)	
7:00-8:00 Breakfast	
Session 6a: Accelerators 2	Session 6b: Mobile & Power Mgmt
An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Reza Yazdani Amnabadi (Universitat Politecnica de Catalunya (UPC)), Albert Sangra (Universitat Politecnica de Catalunya (UPC)), Jose-María Armas (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC))	A Patch Memory System for Image Processing and Computer Vision. Jason Clemens (NVIDIA), Chih-Chih Chen (NVIDIA), Lun Frosio (NVIDIA), Daniel Johnson (NVIDIA), Steve Kackler (NVIDIA)
Co-Designing Accelerators and SoC Interfaces using gem5-SoC. Nikunj Sathya Shao (Harvard University), Sam (Harvard), Wei (Harvard University), David Brooks (Harvard University)	Evaluating Programmable Architectures for Imaging and Vision Applications. Arshad Venkatesh (Stanford), Nishu Ghemawat (Stanford), Arduwan Padman (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)
8:00-9:40	Delegated Perist Ordering. Aashesh Koli (University of Michigan), Jeff Rosen (Intel/Oracle Computing), Stephen Duesterhorst (AIM), Ali Saedi (AIM), Steven Palay (Intel/Oracle Computing), Shang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan)
CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Anrushi Shrivastava (Simon Fraser University), Srinathkumar Kumar (Simon Fraser University), Apala Gulre (Simon Fraser University), Aravindh Shrivastava (Simon Fraser University)	Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 2D Stacks. Simosits Skeltonas (SUJC), Benj Thomas (Intel), Aditya Agrawal (Nvidia), Shihon Qin (ARM), Robert Pinnau-Podgurski (SUJC), Ulye R. Kapurath (ARM), Radu Ionescu (DSU), Nam Sung Kim (SUJC), Joseph Torrellas (SUJC)
Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Moghaddam (SUJC), Young Moon Son (KAIST), Jung Ho Ahn (KAIST), Nam Sung Kim (SUJC)	Throttle: Processor Power Management in the Temperature Inversion Region. Yichun Zu (University of Texas at Austin), Wei Huang (AMD Research), Indranil Paul (AMD Research), Vijay Janapa Reddy (University of Texas at Austin)
9:40-10:00 Break	
Session 7: Best Paper Candidates	
Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. Lee Jun Hwi (Piscataway University), Lina Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation)	Improving Bank-Level Parallelism for Irregular Applications. Kulong Jiang (Penn State), Mahmut Kandemir (Penn State), Proveen Yedlapati (Penn State), Agadesh Kotov (Penn State)
10:00-12:00	Path Confidence Based Lookahead Prefetching. Jinchun Kim (Intel ASIM University), Seth H. Pugsley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Narasimha Reddy (Intel ASIM University), Chris Wilkinson (Intel Labs), Zhihan Chen (Intel Labs), Omar Mutlu (CMU), Yafe N. Puri (UT Austin)
Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emissions. Nader Shehatahkhani (Georgia Tech), Alexrea Nazari (Georgia Tech), Aleksei Zepi (Georgia Tech), Milos Pruckovic (Georgia Tech)	
Session 8: Conference Closing and Best Paper Award	
12:30-18:15 Conference Outing (includes sack lunch)	

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Aamer Jaleel

October 16, 2016 (Sunday)		October 17, 2016 (Monday)		October 18, 2016 (Tuesday)		October 19, 2016 (Wednesday)	
18:00-20:00	Reception			7:00-8:30	Breakfast	7:00-8:00	Breakfast
7:00-8:00	Breakfast			8:30-9:30	Keynote 1: Low Power (EPL) - Sun-Min Lee (University of Washington)	8:00-9:40	Session 6a: Accelerators 2
8:00-8:20	Opening remarks			9:30-10:10	Lightning Session II		
8:20-9:20		Keynote 1: Internet of Things, Hetero and Edge Technology and Policy Margaret Martonosi (University of Michigan)		10:10-10:20	Break		
9:20-10:10	Lightning Session I			Session 3a: Compilation & Memory			
10:00-10:20	Break			Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting , Armaneh Iani (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)	Session 3b: Virtualization		
Session 1a: Microarchitecture	Session 1b: Cloud & Storage			CrystalBall: Statistically Analyzing Runtime Behavior via Deep Sequence Learning , Stephen Jankary (University of Michigan), Daniel Rings (University of Michigan), Nathan Harata (University of Michigan), Michael Laurenzano (University of Michigan), Lingtao Yang (University of Michigan), Jason Mars (University of Michigan)	OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures , Ise Zhan (UCSB), Omer Kayran (AMD Research), Gabriel H. Lee (AMD Research), Chta R. Das (PSU), Yuan-Kai (UCSB)		
Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches, Binabandhan Panda (NVIDIA Research), Andre Saenz (NVIDIA Research)	SABRe: Atomic Object Reads for In-Memory Rack-Scale Computing, Alexander Daglis (EPFL), Dimitri Laskov (EPFL), Marko Novakovic (EPFL), Edward Sproson (EPFL), Bekah Fahri (EPFL), Boris Grot (University of Edinburgh)			A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers , Jai Zhuo (UCSB), Iir Agung (UCSB), Jinho Zhou (UCSB), AI Davis (HP Labs), Paolo Farabochi (HP Labs), Yuangang Wang (Hawaii), Yuan Kai (UCSB)	ChainSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains , Arunesh Sherfan (Simon Fraser University), Srinathkumar Karmali (Simon Fraser University), Apala Guha (Simon Fraser University), Avinash Shrivastava (Simon Fraser University)		
Perception Learning for Reuse Prefetching, Ehsa Imani (Iowa State University), Zhu Wang (Intel Labs), Daniel A. Jimenez (Iowa State University)	A Cloud-Scale Acceleration Architecture, Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Anepudi (Microsoft), Jeremy Lowers (Microsoft Research), Michael Haslam (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joon-Young Kim (Microsoft Research), Daniel Lu (Microsoft Research), Jodi Maswengu (Microsoft Research), Kain Dvorchov (Microsoft Research), Michael Papernschall (Microsoft Research), Lin Wu (Microsoft Research), Sivaram Lanka (Microsoft), Derek Chou (Microsoft), Qing Burger (Microsoft Research)			Low-Cost Soft Error Resilience with Unified Data Verification and Free-Generated Repair for Acoustic Sensor Based Detection , Qingxi Lu (Virginia Tech), Changhui Jung (Virginia Tech), Dongyong Lee (Virginia Tech), Jeevitha Suman (Oak Ridge National Lab)	Content-based Congestion Management in Large-Scale Networks , Gyeongmin Kim (KAIST), Cheongun Kim (KAIST), Hyun Jeong (KAIST), Mika Parker (Intel), John Kim (KAIST), Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques, Dominic Dittmann (Ohio University), Brian Jordan (Ohio University), Avinash Kati (Ohio University), Ahmed Lam (George Washington University)		
Register Sharing for Equal Quality Prediction, Arthur Peres (IBM), Fernando A. Endo (NIRA), Andre Saenz (NIRA)	A Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvrajit Sarmah (Massachusetts Institute of Technology), Mahesh Abraham (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)			Lazy Release Consistency for GPUs , Jeevitha Aklav (University of Illinois at Urbana-Champaign), Ameer Hameed (University of Wisconsin - Madison, AMD Research), Man S. Coor (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research)	Reducing Data Movement Energy via On-chip Data Clustering and Encoding , Shibo Wang (University of Rochester), Engin Ipek (University of Rochester)		
Register Sharing for Quality Prediction, Arthur Peres (IBM), Fernando A. Endo (NIRA), Andre Saenz (NIRA)	Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvrajit Sarmah (Massachusetts Institute of Technology), Mahesh Abraham (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)			10:30-12:10	Break		
12:00-14:00	Lunch			10:30-12:10	Break		
14:00-15:40	Poster session			12:10-14:10	Award Lunch (including Bob Hoo Award, best of time)		
15:40-16:00	Break			Session 4a: Multicore	Session 4b: Security		
Session 2a: GPU	Session 2b: Neural Networks			Racer: TSO Consistency via Race Detection , Alberto Iles (Universidad de Murcia), Stefano Karavas (Uppsala University)	Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors , Mikhail Kasdagli (University of Texas at Austin), Vijay Janapa Reddy (University of Texas at Austin), Mingliu Tian (University of Texas at Austin)		
MIMD Synchronization on SIMD Architectures, Ahmad Alilawati (The University of British Columbia), Tarik A. Aamodi (The University of British Columbia)	From High-Level Deep Neural Models to FPGA, Herdik Sharma (Georgia Institute of Technology), Jongsik Park (Georgia Institute of Technology), Oyea Mahajan (Georgia Institute of Technology), Emmanuel Anato (Georgia Institute of Technology), Joon Young Kim (Georgia Institute of Technology), Chawki Shaq (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)			Exploiting Semantic Commutativity in Hardware Speculation , Guozuo Zhang (MIT CSAIL), Virginia Chu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)	Poinovity: Safe Speculation for Secure Memory , Iyanna Subergenti Lehman (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University)		
Efficient Kernel Synthesis for Performance Portable Programming, Li-Wen Chang (University of Illinois at Urbana-Champaign), Izat El Hag (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Hawaii America Research Lab), Juan Gomez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	VDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Hwu (NVIDIA), Natalia Gervashin (NVIDIA), Jason Clemons (NVIDIA), Arden Zilberstein (NVIDIA), Stephen Wu, Jacobus (NVIDIA)			14:10-15:20	CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems, Cheechen Chau (Georgia Tech), Aamer Jaleel (NVIDIA), Mamoudou K. Qurashi (Georgia Tech)		
KLIP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism, Izat El Hag (University of Illinois at Urbana-Champaign), Juan Gomez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dejan Misovic (Pheylett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	Combining K: An Accelerator for Sparse Neural Networks, Shih Zhong (Institute of Computing Technology, CAS), Ziding Du (Institute of Computing Technology, CAS), Lei Zhang (Institute of Chinese Academy of Sciences), Huaying Lan (Institute of Computing Technology, CAS), Shaohui Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Hengshu Tang (Institute of Computing Technology, CAS), Ting Liu (Institute of Computing Technology, CAS)			CSD: Mitigating the NUMA Bottleneck via Coherent DRAM Caches , Cheng-Kiaih Huang (University of Edinburgh), Sakshil Kumar (University of Edinburgh), Marco Eber (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)	ReplyConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay , Mengyao Yan (University of Illinois at Urbana-Champaign), Joseph Terrellan (University of Illinois at Urbana-Champaign)		
Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPUs, Naifeng Jing (Shanghai Jiao Tong University), Jianfeng Wang (Shanghai Jiao Tong University), Tengfeng Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Jun Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyan Liang (Shanghai Jiao Tong University), Zoruzi A. Holistic Approach to Resource Virtualization in GPUs, Tanvika Vepakumar (Samsung Mobile University), Saunil Shah (Samsung Mobile University), Gennadiy Petrov (Samsung Mobile University), Semra Khan (Samsung Mobile University), Ashish Shah (Samsung Mobile University), Adwait Deshpande (Samsung Mobile University), William and Maria, Philip S. Gibbons (Samsung Mobile University), Ching-Hsiang Lin (Samsung Mobile University), GBAP: Minimizing Energy for GPU Applications, Hengrui Hu (University of Illinois at Urbana-Champaign), Performance Requirements, Muhammad Asif (University of Illinois at Urbana-Champaign), Henry Hoffmann (University of Illinois at Urbana-Champaign)				15:30-16:00	Break		
18:00-20:00	Business meeting			15:30-16:00	Break		
				Session 5a: Approximate Computing	Session 5b: Accelerators 1		
				Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation , Aniranh Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shiv-Dhak Lun (University of Michigan, Ann Arbor), Muneesh Khan (Uppsala University), Md E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingtao Yang (University of Michigan, Ann Arbor), Alex C. Hirsh (University of Michigan, Ann Arbor), Virginia Chu (University of Michigan, Ann Arbor), Arunesh Sherfan (Simon Fraser University), Srinathkumar Karmali (Simon Fraser University), Apala Guha (Simon Fraser University), Avinash Shrivastava (Simon Fraser University)	HARE: Hardware Accelerator for Regular Expressions , Weibiao Gogale (University of Michigan), Anuresh Kati (University of Michigan), Michael J. Lafferty (University of Michigan), Lora O'Rourke (University of Wisconsin-Madison), Adam Werthner (University of Wisconsin-Madison)		

$$\left(\frac{20 \text{ min}}{\text{paper}} \right) \left(\frac{61 \text{ papers}}{\text{MICRO}} \right) \left(\frac{54/2 + 7 \text{ papers}}{\text{MICRO}} \right)$$

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Amer Jaleel

16:00-20:00 Reception	
October 17, 2016 (Monday)	
7:00-8:00	Breakfast
8:00-8:20	Opening remarks
8:20-9:20	Keynote 1: Internet of Things, Mobile and Edge: Technology and Policy Margaret Martonosi (Microsoft)
9:20-10:00	Lightning Session I
10:00-10:20	Break
Session 1a: Microarchitecture	Session 1b: Cloud & Storage
Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. Binambardan Panda (NVIDIA Research), Anshu Saxena (NVIDIA Research)	SABRe: Atomic Object Reads for In-Memory Rack-Scale Computing. Alexander Dagit (EPFL), Dimitri Uskoproff (EPFL), Marko Novakovic (EPFL), Eduard Bugren (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)
Perception Learning for Reuse Prediction. Huiwen Liang (Intel ASIM University), Zhe Wang (Intel Labs), Daniel A. Jiménez (Intel ASIM University)	A Cloud-Scale Acceleration Architecture. Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angepal (Microsoft), Jeremy Losses (Microsoft Research), Michael Haslam (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joo-Young Kim (Microsoft Research), Daniel Lu (Microsoft Research), Jodi Masberg (Microsoft Research), Kain Dytchew (Microsoft Research), Michael Pappaschall (Microsoft Research), Lin Woods (Microsoft Research), Sivaram Lanka (Microsoft), Derek Chau (Microsoft), Qing Burger (Microsoft Research)
pTack: A Smart Prefetching Scheme for OS Intensive Applications. Prathmesh Kulkarni (Intel Delhi), Sumesh R. Sarangi (Intel Delhi)	Forward Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. Yang Hu, and Tao Li (University of Florida)
Register Sharing for Equality Prediction. Arthur Perais (INRIA), Fernando A. Endo (INRIA), André Sarmas (INIA)	Bridging the I/O Performance Gap for Big Data Workloads: A New NVDIMM-based Approach. Benjie Chen (The Hong Kong Polytechnic University), Yao Li (NDS/University of Florida)
Data-Centric Execution of Speculative Parallel Programs. Mark C. Jeffrey (Massachusetts Institute of Technology), Suvraj Subramaniam (Massachusetts Institute of Technology), Mahesh Abeyethera (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)	NeSoC: Self-Virtualizing Nested State Controller. Yonatan Gutfreund (Technion), Yoav Elovic (Technion)
12:00-14:00 Lunch	
14:00-15:40 Poster session	
15:40-16:00 Break	
Session 2a: GPU	Session 2b: Neural Networks
MIMD Synchronization on SIMD Architectures. Ahmad Alkhatib (The University of British Columbia), Tarik M. Aamodi (The University of British Columbia)	From High-Level Deep Neural Models to FPGA. Harsh Sharma (Georgia Institute of Technology), Jorgoe Park (Georgia Institute of Technology), Doye Mahajan (Georgia Institute of Technology), Emmanuel Anato (Georgia Institute of Technology), Joon Myung Kim (Georgia Institute of Technology), Chawika Shao (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)
Efficient Kernel Synthesis for Performance Portable Programming. Li-Wen Chang (University of Illinois at Urbana-Champaign), Izat El Hag (University of Illinois at Urbana-Champaign), Christopher Rodriguez (Pensacola America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	VDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Hwu (NVIDIA), Natalia Gimmlshen (NVIDIA), Jason Clarmont (NVIDIA), Arvin Zulige (NVIDIA), Stephen Wu, Kevlin (NVIDIA)
KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism. Izat El Hag (University of Illinois at Urbana-Champaign), Juan Gómez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dejan Mojicic (Pheylett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)	Combiner-X: An Accelerator for Sparse Neural Networks. Shih Zhang (Institute of Computing Technology, CAS), Ziding Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huaying Liu (Institute of Computing Technology, CAS), Shaoh Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Henshi Yan (Institute of Computing Technology, CAS), Ting Lu (Institute of Computing Technology, CAS)
Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPGPUs. Naifeng Jing (Shanghai Jiao Tong University), Jianfei Wang (Shanghai Jiao Tong University), Fengfeng Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Li Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyan Liang (Shanghai Jiao Tong University), Zoruc A. Holistic Approach to Resource Virtualization in GPUs. Tereza Vojtkova (Samsung Mobile), Sven Bolk (Georgia Institute of Technology), Gerald Schick (Samsung Mobile University), Simeon Huet (Samsung Mobile University), Simeon Huet (Samsung Mobile University), Arghy Banerjee (Georgia Institute of Technology), Genguo Chen (Georgia Institute of Technology), Adnan Azeem (Georgia Institute of Technology), William and Maria Philip S. Labiano (Samsung Mobile University), Chao Li (Shanghai Jiao Tong University), Shih-Chieh Liu (Shanghai Jiao Tong University)	
GPAPS: Mitigating Energy for GPU Applications. Harsh Sharma (Georgia Institute of Technology), Suvraj Subramaniam (Georgia Institute of Technology), Joel Emer (University of Illinois at Urbana-Champaign), Binambardan Panda (NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)	
16:00-20:00 Business meeting	

7:00-8:30 Breakfast	
8:30-9:30 Keynote 1: Low Power (CPU, IoT, Mobile to Wearable & IoT) Ulisses Ko (Redhat)	
9:30-10:10 Lightning Session II	
10:10-10:20 Break	
Session 3a: Compilation & Memory	Session 3b: Interconnects
Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Armesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)	OISCAR: Orchestrating TTT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Iw Zhan (UCSC), Omur Kayran (AMD Research), Gabriel H. Loh (AMD Research), Chite R. Das (PISA), Yuan Xie (UCSB)
CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jalury (University of Michigan), Daniel Rong (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjie Tang (University of Michigan), Jason Mars (University of Michigan)	A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Jai Zhan (UCSB), Sri Raghav (UCSB), Jihun Zhao (UCSC), Ki Daein (HP Labs), Paolo Farabochi (HP Labs), Yungang Wang (Huawei), Yuan Xie (UCSB)
Low-Cost Soft Error Resilience with Unified Data Verification and Free-Reduced Recovery for Atomic Sense Based Detection. Qingxiu Lu (Virginia Tech), Changhua Jung (Virginia Tech), Dongyong Lee (Virginia Tech), Dewshri Imman (Oak Ridge National Lab)	Content-based Congestion Management in Large-Scale Networks. Ganggang Kim (KAIST), Cheongun Kim (KAIST), Jiyun Jeong (KAIST), Mika Parker (Intel), John Kim (KAIST)
Lazy Release Consistency for GPUs. Jeevanth Akhup (University of Illinois at Urbana-Champaign, AMD Research), Man S. Cox (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research)	Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Dominic D'Amico (Ohio University), Brian Boman (Ohio University), Avirach Koth (Ohio University), Ahmed Elmaghrabi (George Washington University)
Improving Energy Efficiency of DRAM by Exploiting Half Page Row Access. Heeyoung Heo (Stanford University), Andrian Padman (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)	Reducing Data Movement Energy via Online Data Clustering and Encoding. Shihui Wang (University of Rochester), Engin Ipek (University of Rochester)
12:10-14:10 Award Lunch (including Bob Hwu Award, best of time)	
Session 4a: Multicore	Session 4b: Security
Racer: TSO Consistency via Race Detection. Alberto Iles (Universidad de Murcia), Stefanos Karabas (Alibaba University), Exploiting Semantic Commutativity in Hardware Speculation. Guozhen Zhang (MIT CSAIL), Virginia Chu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)	Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors. Mikhail Kasdagli (University of Texas at Austin), Vijay Janapa Reddy (University of Texas at Austin), Minh Toan (University of Texas at Austin)
CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. Cheachen Chau (Georgia Tech), Amer Jalael (NVIDIA), Mamduddin K. Qureshi (Georgia Tech)	Polynosity: Safe Speculation for Secure Memory. Yemana Subergelji Lehman (Duke University), Andrew D. Hillston (Duke University), Benjamin C. Lee (Duke University)
CSD: Mitigating the NUMA Bottleneck via Coherent DRAM Caches. Cheng-Chieh Huang (University of Edinburgh), Gwanik Kumaer (University of Edinburgh), Marco Elovic (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)	ReplyConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay. Mengyao Yan (University of Illinois at Urbana-Champaign), Imran Shafiq (University of Illinois at Urbana-Champaign), Joseph Torrellas (University of Illinois at Urbana-Champaign)
15:30-16:00 Break	
Session 5a: Approximate Computing	Session 5b: Accelerators 1
Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. Aniranth Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shiv-Dheek Lun (University of Michigan, Ann Arbor), Muneeb Khan (Sipplake University), Md E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Mark H. Horowitz (University of Michigan, Ann Arbor)	HARE: Hardware Accelerator for Regular Expressions. Weibiao Gogale (University of Michigan), Aaweshh Koli (University of Michigan), Michael J. Lafferty (University of Michigan), Leena D'Antonio (University of Wisconsin-Madison), Mark Horowitz (University of Michigan)
AggPrognostic: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application. Arundhanai Senthil Kumar (University of Illinois at Urbana-Champaign), Arundhanai Senthil Kumar (University of Illinois at Urbana-Champaign), Yuhang Xie (University of Illinois at Urbana-Champaign), Arundhanai Senthil Kumar (University of Illinois at Urbana-Champaign)	The Microarchitecture of a Real-time Robot Motion Planning Accelerator. Sam Singh (University of Michigan), Aniranth Jain (University of Michigan), George J. Goussard (University of Michigan)
16:00-17:00	

7:00-8:00 Breakfast	
Session 6a: Accelerators 2	
An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Irena Isardova Amrnbadi (Universidad Politécnica de Catalunya (UPC)), Albert Sangra (Universidad Politécnica de Catalunya (UPC)), Jose-María Armas (Universidad Politécnica de Catalunya (UPC)), Antonio González (Universidad Politécnica de Catalunya (UPC))	Session 6b: Mobile & Power Mgmt
Co-Designing Accelerators and SoC Interfaces using gem5-10. Harvard University, Vijayarathnam Srinivasan (IBM), Gu-Lin Wei (Harvard University), David Brooks (Harvard University)	A Patch Memory System for Image Processing and Computer Vision. Jason Clarmont (NVIDIA), Chih-Chih Cheng (NVIDIA), Lun Fosso (NVIDIA), Daniel Johnson (NVIDIA), Steve Kackler (NVIDIA)
CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Aniranth Shriram (Simon Fraser University), Srinaksh Kumar (Simon Fraser University), Apala Gulur (Simon Fraser University), Anirudh Shrivastava (Simon Fraser University)	Evaluating Programmable Architectures for Imaging and Vision Applications. Arsen Vesilj (Stanford), Nikhil Shekhar (Stanford), Andrius Pedram (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)
Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Moghaddam (UTRAC), Young Moon Son (KAIST), Jung Ho Ahn (KAIST), Nam Sung Kim (KAIST)	Defining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users. Kaige Fan (University of Houston), Xinping Zhang (University of Houston), Jengweia Yan (University of Houston), Xin Fu (University of Houston)
8:00-9:40	
Session 7: Best Paper Candidates	
Graphiconado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. Lee Jun Hwi (Pinceton University), Liu Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nandhar Sathish (Parallel Computing Lab, Intel Corporation)	Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks. Dimitrios Skarakostas (KAIST), Benji Thomas (Intel), Aditya Agrawal (Nvidia), Mohan Qin (KAIST), Robert Pinnau-Podgorski (KAIST), Uyen H. Karpuzcu (UMN), Radu Ionescu (CSU), Nam Sung Kim (KAIST), Joseph Torrellas (KAIST)
Improving Bank-Level Parallelism for Irregular Applications. Xulong Jiang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedapalli (Penn State), Agadesh Kotov (Penn State)	Throttle: Processor Power Management in the Temperature Inversion Region. Yichun Zu (University of Texas at Austin), Mike Huang (AMD Research), Indran Paul (AMD Research), Vijay Janapa Reddy (University of Texas at Austin)
9:40-10:00 Break	
Delegated Perist Ordering. Aaweshh Koli (University of Michigan), Jeff Hosen (Intel/Cornell Computing), Stephen Dieckhoff (AIM), Ali Saedi (AIM), Steven Halley (Intel/Cornell Computing), Shang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Chen (University of Michigan)	
Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emanations. Nadim Sabahatkhah (Georgia Tech), Alexia Nazer (Georgia Tech), Alokesh Zarg (Georgia Tech), Jinchun Kim (Texas A&M University), Seth H. Pugsley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Narsimha Reddy (Texas A&M University), Chris Willerson (Intel Labs), Zhihan Chen (Intel Labs)	
Path Consistent Based Lookahead Prefetching. Inchun Kim (Texas A&M University), Seth H. Pugsley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Narsimha Reddy (Texas A&M University), Chris Willerson (Intel Labs), Zhihan Chen (Intel Labs)	
Continuous Runahead: Transparent Hardware Accelerator for Memory Intensive Workloads. Mihai Hershman (UT Austin), Omur Kulkali (CMU), Nile N. Patel (UT Austin)	
Session 8: Conference Closing and Best Paper Award	
12:30-18:15 Conference Outing (includes sack lunch)	

$$\left(\frac{20 \text{ min}}{\text{paper}}\right) \left(\frac{61 \text{ papers}}{\text{MICRO}}\right) \left(\frac{54/2 + 7 \text{ papers}}{\text{MICRO}}\right) \left(\frac{6 \text{ kilojoules}}{\text{min}}\right) =$$

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Amer Jalael

October 16, 2016 (Sunday)	
18:00-20:00 Reception	
October 17, 2016 (Monday)	
7:00-8:00 Breakfast	
8:00-8:20 Opening remarks	
8:20-9:20	 keynote 1: Internet of Things, Mobile and Edge Technology and Policy Margaret Martonosi (University of Washington)
9:20-10:00 Lightning Session I	
10:00-10:20 Break	
Session 1a: Microarchitecture	Session 1b: Cloud & Storage SABRe: Atomic Object Reads for In-Memory Rack-Scale Computing. Alexander Daglis (EPFL), Dimitri Logothopoulos (EPFL), Miroslav Novakovic (EPFL), Eduard Bugren (EPFL), Bekab Jafari (EPFL), Boris Grosz (University of Michigan) A Cloud-Scale Acceleration Architecture. Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Anappai (Microsoft Research), Jeremy Ismael (Microsoft Research), Michael Haselman (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joon-Yeong Kim (Microsoft Research), Daniel Lo (Microsoft Research), Jodi Messerges (Microsoft Research), Kain Dutchavorn (Microsoft Research), Michael Pappaschall (Microsoft Research), Lisa Woods (Microsoft Research), Sivasari Lenka (Microsoft), Derek Chao (Microsoft), Qingpu Burger (Microsoft Research) NVIDIA Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. Yang Hu, and Tao Li (University of Florida) Bridging the I/O Performance Gap for Big Data Workloads: A New NVMMIO-based Approach. Benjie Chen (The Hong Kong Polytechnic University), Tao Li (USF University of Florida) NeSC: Self-Virtualizing Nested Storage Controller. Yonatan Gottman (Technion), Yoav Eason (Technion)
Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. Binambardan Panda (NVIDIA Research), Andre Seneac (NVIDIA Research) Perception Learning for Reuse Prediction. Uthra Iyer (Iowa State University), Zhe Wang (Intel Labs), Daniel A. Jimenez (Iowa State University) pTack: A Smart Prefetching Scheduler for OS Intensive Applications. Prathmesh Kulkarni (EIT Delhi), Sumit R. Sanzgiri (EIT Delhi) Register Sharing for Equality Prediction. Arthur Heras (NVIDIA Research), A. Endo (NVIDIA), Andre Seneac (NVIDIA) Data-Centric Execution of Speculative Parallel Programs. Mark C. Jeffrey (Massachusetts Institute of Technology), Suvrajit Subramanian (Massachusetts Institute of Technology), Mahesh Aravamudan (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)	
12:00-14:00 Lunch	
14:00-15:40 Poster session	
15:40-16:00 Break	
Session 2a: GPU	Session 2b: Neural Networks From High-Level Deep Neural Models to FPGAs. Harsh Sharma (Georgia Institute of Technology), Jorgo Park (Georgia Institute of Technology), Doyu Mahajan (Georgia Institute of Technology), Ermanno Azzato (Georgia Institute of Technology), Joonyoung Kim (Georgia Institute of Technology), Chenshi Shao (Georgia Institute of Technology), Anil Mohite (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology) VDRN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Huh (NVIDIA), Natalia Gromshina (NVIDIA), Jason Clarno (NVIDIA), Arden Zillicoff (NVIDIA), Stephen Wu, Kevin (NVIDIA) Stripes: Bit-Serial Deep Neural Network Computing. Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Taylor Hetherington (University of British Columbia), Jojo Aamodt (University of British Columbia), Andrew Moschovos (University of Toronto) Combinor-X: An Accelerator for Sparse Neural Networks. Shih-Zheng Institute of Computing Technology, CAS, Ziding Gu (Institute of Computing Technology, CAS), Li Zhang (University of Chinese Academy of Sciences), Huaying Liu (Institute of Computing Technology, CAS), Shaohui Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Hengshu Han (Institute of Computing Technology, CAS), Yong-Li Song (Institute of Computing Technology, CAS) MEMBRANE: Reconfigurable Accelerators for Deep Neural Networks. Shengbo Li (University of Toronto), Mehrdad Karimi (University of Toronto), Yanzhao Zhang (University of Toronto), Gregor G. Roth (University of Toronto), Naveed Anjum (University of Toronto), Gaurav Mehta (University of Toronto), David Kohrt (University of Toronto), Billian and Marc, Philip S. Gibbons (University of Toronto), Yipeng Chen (University of Toronto), Junyuan Wang (University of Toronto), Yipeng Chen (University of Toronto), Michael J. Schmitz (University of Toronto), Yipeng Chen (University of Toronto) GEARS: Mitigating Energy for GPU Applications. Yipeng Chen (University of Toronto), Junyuan Wang (University of Toronto), Yipeng Chen (University of Toronto), Michael J. Schmitz (University of Toronto), Yipeng Chen (University of Toronto)
18:00-20:00 Business meeting	

October 16, 2016 (Sunday)	
7:00-8:30 Breakfast	
8:30-9:30	 keynote 1: Low Power (EPFL) / Keynote 1: Mobile to Wearable & IoT (Intel) / Intel x86 Media Lab
9:30-10:10 Lightning Session II	
10:10-10:20 Break	
Session 3a: Compilation & Memory	Session 3b: Information
Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Anirban Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjia Lang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor) CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jekuty (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjia Lang (University of Michigan), Jason Mars (University of Michigan) Low-Cost Soft Error Resilience with Unified Data Verification and Free-Genated Recovery for Acoustic Sensor-Based Detection. Qingpu Lu (Virginia Tech), Changshu Jiang (Virginia Tech), Dongyuan Lee (Virginia Tech), Dewei Siman (Duke University National Lab) Lazy Release Consistency for GPUs. Jovannah Akrop (University of Illinois at Urbana-Champaign), Ahmad Razaifar, Man S. Cox (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research) Improving Energy Efficiency of DRAM by Exploiting Half-Page Row Access. Haoyue He (Stanford University), Andriyan Padman (Stanford University), Stephen Richardson (Stanford University), Shaoh Kwiatkowsky (Technion), Mark Horowitz (Stanford University)	
10:30-12:10	OSCAR: Orchestrating STT-DRAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Iw Zhan (UCSB), Onur Kayran (AMD Research), Gabriel H. Lee (AMD Research), Chih-R. Dow (PISA), Yuan Xie (UCSB) A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Jai Zhan (UCSB), Ji-Il Jeon (UCSB), Jinho Zhan (UCSB), Ki Deon (HPL Labs), Paolo Farabochi (HP Labs), Yuanqing Wang (Hawaii), Yuan Xie (UCSB) Content-based Congestion Management in Large-Scale Networks. Guojun Chen (SKIT), Changuan Kim (KAIST), Ayun Jeong (KAIST), Mika Parker (Intel), John Kim (KAIST) Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Dominic DiMarzio (Ohio University), Brian Jordan (Ohio University), Avinash Kati (Ohio University), Ahmed Lassi (George Washington University) Reducing Data Movement Energy via Online Data Clustering and Encoding. Shihui Wang (University of Rochester), Engin Imekci (University of Rochester)
12:10-14:10	AWARD Lunch: (including Bob Haw. Award, best of time) Session 4a: Multicore Racer: TSO Consistency via Race Detection. Alberto Iles (Universidad de Murcia), Stefano Keates (Alibaba University) Exploiting Semantic Commutativity in Hardware Acceleration. Guozuo Zhang (MIT CSAIL), Virginia Chu (MIT CSAIL), Daniel Sanchez (MIT CSAIL) CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. Cheechee Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Manmohan K. Qureshi (Georgia Tech) CSIT: Mitigating the NUMA Bottleneck via Coherent DRAM Caches. Cheng-Chieh Huang (University of Edinburgh), Sahesh Kumar (University of Edinburgh), Marco Iliuc (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)
15:30-16:00 Break	
Session 5a: Approximate Computing	Session 5b: Accelerators 1
Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. Anirban Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shih-Chieh Lin (University of Michigan, Ann Arbor), Muneeb Khan (Uppsala University), Md. E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingjia Lang (University of Michigan, Ann Arbor), Mehrdad Karimi (University of Toronto)	Session 5b: Accelerators 2 An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Raza Yazdan Arshad (University of Politecnico de Catalunya (UPC)), Albert Sangra (Universitat Politecnica de Catalunya (UPC)), Jose-María Arnaez (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC)) Co-Designing Accelerators and SoC Interfaces using gem5+Aladdin. Yashu Soptia Shao (Harvard University), Sam Li (UCSD) (Harvard University), Vijayalakshmi Srivastava (IBM), Gu-Yuan Wei (Harvard University), David Brooks (Harvard University) CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Arshad Shaheen (Simon Fraser University), Srinathkumar Kumar (Simon Fraser University), Apala Gulha (Simon Fraser University), Arunvith Shriraman (Simon Fraser University) Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Moghadam (UCL), Young Moon Son (SNU), Jung Ho Ahn (KAIST), Nam Sung Kim (KAIST)
16:00-17:00	Session 5c: Security Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors. Michael Kastagdi (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Minh Tien (University of Texas at Austin) Policy-Safe Speculation for Secure Memory. Imane Sberghini Lehmann (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University) Replay/Confusion: Detecting Cache-based Cover Channel Attacks Using Record and Replay. Mengyao Yan (University of Illinois at Urbana-Champaign), Joseph Torrellas (University of Illinois at Urbana-Champaign) Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. Uday Subramanian (SASTRI, Birmingham), Givory Porozemsy (SASTRI, Birmingham), Naveel Abu-Ghazwan (UK, Riverside) HARE: Hardware Accelerator for Regular Expressions. Vishnu Gupta (University of Michigan), Ananthesh Kollu (University of Michigan), Michael J. Laflamme (University of Michigan), Lora D'Arcangelo (University of Wisconsin-Madison)
18:00-20:00	The Microsubstitution of a Real-time Robot Planning Accelerator. Samy Bekkari (University of Wisconsin-Madison), Samy Bekkari (University of Wisconsin-Madison)

October 19, 2016 (Wednesday)	
7:00-8:00 Breakfast	
Session 6a: Accelerators 2	Session 6b: Mobile & Power Mgmt
An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Raza Yazdan Arshad (University of Politecnico de Catalunya (UPC)), Albert Sangra (Universitat Politecnica de Catalunya (UPC)), Jose-María Arnaez (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC)) Co-Designing Accelerators and SoC Interfaces using gem5+Aladdin. Yashu Soptia Shao (Harvard University), Sam Li (UCSD) (Harvard University), Vijayalakshmi Srivastava (IBM), Gu-Yuan Wei (Harvard University), David Brooks (Harvard University) CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Arshad Shaheen (Simon Fraser University), Srinathkumar Kumar (Simon Fraser University), Apala Gulha (Simon Fraser University), Arunvith Shriraman (Simon Fraser University) Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Moghadam (UCL), Young Moon Son (SNU), Jung Ho Ahn (KAIST), Nam Sung Kim (KAIST)	A Patch Memory System for Image Processing and Computer Vision. Jason Clarno (NVIDIA), Chih-Chih Chen (NVIDIA), Lun Foss (NVIDIA), Daniel Johnson (NVIDIA), Steve Kackler (NVIDIA) Evaluating Programmable Architectures for Imaging and Vision Applications. Arden Vesilky (Stanford), Toshihiro Stanford, Andriyan Padman (Stanford), Stephen Richardson (Stanford), Shaoh Kwiatkowsky (Technion), Mark Horowitz (Stanford) Defining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users. Kanyu Fan (University of Houston), Xinyang Zhang (University of Houston), Jingyuan Fan (University of Houston), Xin Fu (University of Houston) Switch: Opportunistically Reassigning Power Allocation between Processor and Memory in 2D Stacks. Dimitrios Skafaratos (SUNY), Benj Thomas (Intel), Aditya Agrawal (Nvidia), Shihon Qin (SUNY), Robert Pinnau-Podgórski (SUNY), Ulyse R. Kapuscus (UMN), Radu Ionescu (SUNY), Nam Sung Kim (SUNY), Joseph Torrellas (SUNY) Throttle: Processor Power Management in the Temperature Inversion Region. Yichou Zu (University of Texas at Austin), We Huang (AMD Research), Indran Paul (AMD Research), Vijay Janapa Reddi (University of Texas at Austin)
8:00-9:40	Session 7: Best Paper Candidates Graphiconado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. Iaw Jun Hien (Pinceton University), Xia Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nadeebur Saikh (Parallel Computing Lab, Intel Corporation) Improving Bank-Level Parallelism for Irregular Applications. Kulong Jiang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapati (Penn State), Agadesh Kotov (Penn State) Delegated Perist Ordering. Aashesh Koli (University of Michigan), Jeff Rosen (Intel/Oracle Computing), Stephen Duesterhold (AIM), Ali Saedi (AIM), Steven Halley (Intel/Oracle Computing), Shang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Chen (University of Michigan) Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emanations. Nadar Sahlabakhsh (Georgia Tech), Alexrea Naeem (Georgia Tech), Alekx Zepi (Georgia Tech), Milos Pruckovic (Georgia Tech) Patch Confidence based Lookahead Prefetching. Junehun Kim (Iowa State University), Seth H. Pugsley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Narasimha Reddy (Texas A&M University), Chris Wilkinson (Intel Labs), Zhenhan Cheshi (Intel Labs) Continuous Runahead: Transparent Hardware Acceleration for Memory Intensive Workloads. Mirad Hashemi (UT Austin), Onur Mutlu (CMU), Yale N. Patt (UT Austin)
9:40-10:00 Break	
10:00-12:00	Session 8: Conference Closing and Best Paper Award 12:30-18:15 Conference Outing (includes sack lunch)

$$\left(\frac{20 \text{ min}}{\text{paper}}\right) \left(\frac{61 \text{ papers}}{\text{MICRO}}\right) \left(\frac{54/2 + 7 \text{ papers}}{\text{MICRO}}\right) \left(\frac{6 \text{ kilojoules}}{\text{min}}\right) =$$

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Aamer Jaleel

<p>10:00-20:00 Reception</p> <p>7:00-8:00 Breakfast</p> <p>8:00-9:00 Opening Remarks</p> <p>9:00-9:20</p> <p>9:20-10:00 Lightning Session 1</p> <p>10:00-10:20 Break</p> <p>Session 5a: Microarchitecture</p> <p>Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. <i>Shrawanendra Pandey (MSU), Arvind Senani (MSU), Arvind Senani (MSU)</i></p> <p>Perception Learning for Swave Prediction. <i>Chao Tian (Tsinghua University), Wei Wang (Tsinghua), Daniel A. Jiménez (MSU, AMD)</i></p> <p>gTask: A Smart Prefetching Scheme for OS Intensive Applications. <i>Prathmesh Salunkar (UIUC), Suresh R. Suresh (UIUC)</i></p> <p>Register Sharing for Equality Prediction. <i>Arifur Haque (MSU), Fernando A. Sindr (MSU), Arvind Senani (MSU)</i></p> <p>Data-Centric Execution of Speculative Parallel Programs. <i>Mark L. Atiles (Massachusetts Institute of Technology), Souvik Saha (Massachusetts Institute of Technology), Mahan Mohtashemi (Massachusetts Institute of Technology), and Omar Shalwan (Massachusetts Institute of Technology and NVIDIA), Daniel Gardner (Massachusetts Institute of Technology)</i></p> <p>12:00-14:00 Lunch</p> <p>14:00-14:10 Poster session</p> <p>14:10-14:30 Break</p> <p>Session 5a: GPU</p> <p>MIMD Synchronization on SMT Architectures. <i>Minmin Wu (The University of British Columbia), Yu-Mei Kwang (The University of British Columbia)</i></p> <p>Efficient Kernel Synthesis for Performance Portable Programming. <i>Li-Wen Chang (University of Illinois at Urbana-Champaign), Liat H. Hag (University of Illinois at Urbana-Champaign), Christophe Rodriguez (Riviera America Research Lab), Luis Gomez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)</i></p> <p>KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism. <i>Yuan (The University of Illinois at Urbana-Champaign), Luis Gomez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dipan Majumdar (Purdue-Richard Lab), Wen-mei Hwu (University of Illinois at Urbana-Champaign)</i></p> <p>Cache-Embedded Register File: An Integrated On-Chip Memory Architecture for High Performance GPUs. <i>Yanfeng Jing (Shanghai Jiao Tong University), Jianli Wang (Shanghai Jiao Tong University), Tingting Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Jing (Shanghai Jiao Tong University), Chen Li (Shanghai Jiao Tong University), Xiaoyan Jiang (Shanghai Jiao Tong University), Zhenyu Li (Shanghai Jiao Tong University)</i></p> <p>Zorus: A Holistic Approach to Resource Virtualization in GPUs. <i>Natalia Vasyukina (Georgia Institute of Technology), Saurabh Shah (Georgia Institute of Technology), Gernot Heiser (Georgia Institute of Technology), Suresh Mani (Georgia Institute of Technology), Arvind Senani (Georgia Institute of Technology), Arvind Senani (Georgia Institute of Technology), William and Maria, Philip S. Gibbons (Georgia Institute of Technology), Chao Tian (Georgia Institute of Technology), GAPE: Minimizing Energy for GPU Applications. <i>Performance Requirements, Muhammad Ali (University of Houston), Henry Hoffmann (University of Houston)</i></i></p>	<p>October 16, 2016 (Monday)</p> <p>October 17, 2016 (Monday)</p> <p>Session 5a: Cloud & Storage</p> <p>SARee: Atomic Object Reads for In-Memory Rack-Scale Computing. <i>Adityan Dasgupta (DRI), Shreshth Vasudeva (DRI), Ranjan Ramesh (DRI), Edward Suresh (DRI), Gokul Laksh (DRI), Ravi Sridhar (University of Edinburgh)</i></p> <p>A Cloud-Scale Acceleration Architecture. <i>Ashim M. Laskhal (Microsoft Research), Tom Y. Hoang (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Anupam (Microsoft Research), Jeremy Young (Microsoft Research), Michael Heuleman (Microsoft Research), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft Research), Kowal Nak (Microsoft Research), David M. Kaelin-Lang (Microsoft Research), Todd Mowery (Microsoft Research), Sabin Chatterjee (Microsoft Research), Michael Papernost (Microsoft Research), Lou Wood (Microsoft Research), Stefan Lanka (Microsoft Research), Derek Chiu (Microsoft Research), Roger (Microsoft Research)</i></p> <p>Towards Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. <i>Yong Hu, and Lei Li (University of Toronto)</i></p> <p>Bridging the I/O Performance Gap for Big Data Workloads: A New RDMA-based Approach. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>Session 5b: Neural Networks</p> <p>From High-Level Deep Neural Models to FPGA. <i>Huili Shao (Georgia Institute of Technology), Yongge Park (Georgia Institute of Technology), Deepak Rajeev (Georgia Institute of Technology), Komal Anand (Georgia Institute of Technology), Joon Seung Kim (Georgia Institute of Technology), Chanku Kim (Georgia Institute of Technology), And Malika Mittal (Carnegie Mellon), Hui (Carnegie Mellon)</i></p> <p>VDRN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. <i>Minmin Wu (University of Illinois at Urbana-Champaign), Arvind Senani (MSU), Arvind Senani (MSU), Arvind Senani (MSU)</i></p> <p>Stripes: Bit-Serial Deep Neural Network Computation. <i>Yuan (The University of Illinois at Urbana-Champaign), Jiyang Huang (University of Toronto), Jiyang Huang (University of Toronto), Saurabh Nath (University of British Columbia), Jay Ramesh (University of British Columbia), Arvind Senani (University of British Columbia), Arvind Senani (University of British Columbia)</i></p> <p>Combiner-B: An Accelerator for Sparse Neural Network Computation. <i>Yuan (The University of Illinois at Urbana-Champaign), Jiyang Huang (University of Toronto), Jiyang Huang (University of Toronto), Saurabh Nath (University of British Columbia), Jay Ramesh (University of British Columbia), Arvind Senani (University of British Columbia), Arvind Senani (University of British Columbia)</i></p> <p>Session 5c: Approximate Computing</p> <p>Combiner Leads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p> <p>Session 5d: Accelerators 1</p> <p>SARe: Hardware Accelerator for Regular Expressions. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p>	<p>October 18, 2016 (Tuesday)</p> <p>7:00-8:00 Breakfast</p> <p>8:30-9:30</p> <p>9:30-10:30 Lightning Session II</p> <p>10:10-10:30 Break</p> <p>Session 5b: Compilation & Memory</p> <p>Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p> <p>CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. <i>Stephen Jafari (University of Michigan), Daniel Singh (University of Michigan), Nathan Harada (University of Michigan), Michael Lauer (University of Michigan), Lingjie Tang (University of Michigan), Arvind Senani (University of Michigan), Arvind Senani (University of Michigan)</i></p> <p>A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. <i>Chao Tian (UIUC), Arvind Senani (UIUC), Arvind Senani (UIUC), Arvind Senani (UIUC), Arvind Senani (UIUC), Arvind Senani (UIUC)</i></p> <p>Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection. <i>Gregory Liu (Virginia Tech), Chao Tian (Virginia Tech), Dongyuan Luo (Virginia Tech), Chao Tian (Virginia Tech), Chao Tian (Virginia Tech)</i></p> <p>Lazy Release Consistency for GPUs. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p> <p>Improving Energy Efficiency of DRAM by Exploiting Half-Page Row Access. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>Racer: TSD Consistency via Race Detection. <i>Alberto Ros (Universidad de Murcia), Stefano Garcia (Alibaba Inc), Exploiting Semantic Commutativity in Hardware Speculation. <i>Lingjie Tang (UIUC), Lingjie Tang (UIUC), Lingjie Tang (UIUC), Lingjie Tang (UIUC)</i></i></p> <p>CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. <i>Chao Tian (Virginia Tech), Aamer Jaleel (MSU), Arvind Senani (MSU)</i></p> <p>CSD: Mitigating the NUMA Bottleneck via Coherent DRAM Caches. <i>Chao Tian (Virginia Tech), Lingjie Tang (Virginia Tech), Lingjie Tang (Virginia Tech), Lingjie Tang (Virginia Tech)</i></p> <p>Reply/Confusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>Session 5a: Accelerators 2</p> <p>An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p> <p>Co-Designing Accelerators and SoC Interfaces using gem5-Addin. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>CHARMS: Van-Neuron Accelerators to Leverage Food Instruction Claims. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p> <p>Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. <i>Lee Yun (Princeton University), Lee Yun (Princeton University), Lee Yun (Princeton University), Lee Yun (Princeton University)</i></p> <p>Improving Energy Efficiency of GPU Applications. <i>Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign), Arvind Senani (University of Illinois at Urbana-Champaign)</i></p> <p>Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emissions. <i>Natalie Vasyukina (Georgia Tech), Arvind Senani (Georgia Tech), Arvind Senani (Georgia Tech), Arvind Senani (Georgia Tech)</i></p> <p>Path Confidence based Lookahead Prefetching. <i>Yusuf Chan (The Hong Kong Polytechnic University), Qi Zhou (The Hong Kong Polytechnic University), Yali Qiu (University of Illinois)</i></p> <p>Session 5b: Conference Closing and Best Paper Award</p> <p>12:30-12:15 Conference Closing (includes best lunch)</p>
--	---	---

4x energy of Starfleet phaser rifle!

$$\left(\frac{20 \text{ min}}{\text{paper}}\right) \left(\frac{61 \text{ papers}}{\text{MICRO}}\right) \left(\frac{54/2 + 7 \text{ papers}}{\text{MICRO}}\right) \left(\frac{6 \text{ kilojoules}}{\text{min}}\right) =$$

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Ronger and Aamer Jaleel

October 16, 2016 (Sunday)	
8:30-20:00	Registration
October 17, 2016 (Monday)	
7:00-8:00	Breakfast
8:00-8:20	Opening Remarks
8:20-9:20	
9:20-10:20	Lightning Session 1
10:30-10:20	Break
Session 5a: Microarchitecture	<p>Session 5a: Cloud & Storage</p> <p>SAFER: Atomic Object Reads for In-Memory Rack-Scale Computing. Aditya Dasgupta (DRI), Shreshth Upadhyay (DRI), Ranbir Neelgandhi (DRI), Edward Suresh (DRI), Babak Jalali (DRI), Benoit Bonhomme (Edinburgh)</p> <p>A Cloud-Scale Acceleration Architecture. Adrian M. Luchinat (Microsoft Research), Yan Li (Google), Microsoft Research, Andrew Putnam (Microsoft Research), Hari Anandaraj (Microsoft), Jeremy Owens (Microsoft Research), Michael Heuleman (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Kunal Kulkarni (Microsoft), Bao-Ngung Kim (Microsoft Research), Daniel Li (Microsoft Research), Todd Mowbray (Microsoft Research), Sathya Dhanasekaran (Microsoft Research), Michael Papernost (Microsoft Research), Lou Woods (Microsoft Research), Stefan Laska (Microsoft), Derek Chen (Microsoft), Sheng-Rong Jiang (Microsoft Research)</p> <p>Towards Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. Yong Hu, and Lei Li (University of Toronto)</p> <p>Bridging the I/O Performance Gap for Big Data Workloads: A New NVMM-Based Approach. Weidong Chen (The Hong Kong Polytechnic University), Qi Wang (The Hong Kong Polytechnic University), Yali Guo (Microsoft), Neel S. Gupte (Virtualizing Next-Generation Storage), Sathya Dhanasekaran (Microsoft Research)</p>
12:00-14:00	Lunch
14:00-14:40	Poster session
14:40-16:00	Break
Session 5a: GPU	<p>Session 5b: Neural Networks</p> <p>From High-Level Deep Neural Models to FPGA. Huihui Shao (Georgia Institute of Technology), Joseph Park (Georgia Institute of Technology), Emmanuel Anzures (Georgia Institute of Technology), Jun Young Kim (Georgia Institute of Technology), Chankyu Shin (Georgia Institute of Technology), Aidan Healy (Copenhagen), David Kaelin-Lang (Georgia Institute of Technology)</p> <p>VDRN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Huh (NVIDIA), Natalie Goodenough (NVIDIA), Jason Collins (NVIDIA), Andre Clavier (NVIDIA), Stephen Hinton (NVIDIA)</p> <p>Stripes: Bit-Serial Deep Neural Network Compiling. Patrick Ahl (University of Toronto), Jorge Albericio (University of Toronto), Stefan Hoffmann (University of British Columbia), Ben Karmali (University of British Columbia), Andrew Michaloski (University of British Columbia)</p> <p>Combinator-B: An Accelerator for Sparse Neural Networks. Sheng Zhong (Institute of Computing Technology, CAS), Jintao Xu (Institute of Computing Technology, CAS), Lei Zhang (Institute of Chinese Academy of Sciences), Ruiyuan Lan (Institute of Computing Technology, CAS), Shuai Lu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Jianshi Chen (Institute of Computing Technology, CAS), Yong Chen (Institute of Computing Technology, CAS)</p> <p>NeuFRAMS: Neural Network Transformation and Co-design Under Neuro-inspired Hardware Constraints. Yi (Yongqiang Chen), Weidong Chen (Tsinghua Univ.), Changchun Li (USTC), Peng-Li (USTC), Yuhang Song (Tsinghua Univ.), Peng Lu (Tsinghua Univ.), Yuan He (USTC), Weidong Chen (Tsinghua Univ.)</p> <p>Flow-Level CNVR: Accelerators, Systems, and Design. Joshua San Miguel (University of Toronto), Jorge Albericio (University of Toronto), Stefan Hoffmann (University of Toronto), Aamer Jaleel (USTC)</p>
16:00-18:00	Session 5a: GPU
18:00-20:00	Business meeting

October 16, 2016 (Sunday)	
7:00-8:00	Breakfast
8:30-9:30	
9:30-10:30	Lightning Session II
10:30-10:20	Break
Session 5a: Compilation & Memory	<p>Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Anirban Basu (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingya Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)</p> <p>CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jang (University of Michigan), Daniel Singh (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingya Tang (University of Michigan), Jason Mars (University of Michigan), Jesse Mars (University of Michigan)</p> <p>Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection. Qingxiu Liu (Virginia Tech), Chuanjun Luo (Virginia Tech), Dongmei Liu (Virginia Tech), Weidong Chen (Rutgers)</p> <p>Lazy Release Consistency for GPUs. Jonathan Alap (University of Toronto, Canada), Yuhang Chen (AMD Research), Madan S. Gupte (University of Wisconsin - Madison, AMD Research), Sathya Dhanasekaran (Microsoft Research), David A. Wood (University of Wisconsin - Madison, AMD Research)</p> <p>Improving Energy Efficiency of DRAM by Exploiting Half-Page Row Access. Heung Ju Han (University of Anbar), Andrew Song (Georgia Institute of Technology), Sathya Dhanasekaran (Microsoft Research), Yuhang Chen (AMD Research), Madan S. Gupte (University of Wisconsin - Madison, AMD Research)</p>
10:30-12:10	Session 5a: Compilation & Memory
14:10-15:10	Session 5a: Compiler
15:30-16:00	Break
Session 5a: Approximate Computing	<p>Concise Loads and Stores: The Case for an Anysymmetric Compute-Memory Architecture for Approximation. Anirban Basu (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Sathya Dhanasekaran (Microsoft Research), Madan S. Gupte (University of Wisconsin - Madison, AMD Research), Madan S. Gupte (University of Wisconsin - Madison, AMD Research), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingya Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)</p> <p>Approxifyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency. Radha Venkatesh (University of Illinois at Urbana-Champaign), Abhishek Maheshwari (University of Illinois at Urbana-Champaign), Vivek Kumar Sanyal (Microsoft), Sarvesh V. Advani (University of Illinois at Urbana-Champaign)</p> <p>The Bunker Cache for Spatio-Value Approximation. Joshua San Miguel (University of Toronto), Jorge Albericio (University of Toronto), Sathya Dhanasekaran (Microsoft Research), Aamer Jaleel (USTC)</p>
17:00-21:00	Banquet

October 16, 2016 (Sunday)	
7:00-8:00	Breakfast
8:30-9:30	
9:30-10:30	Lightning Session II
10:30-10:20	Break
Session 5a: Accelerators 2	<p>Session 5b: Hardware</p> <p>OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Yu Chen (USTC), Chao Xuebin (AMD Research), Subhojit Chakrabarti (AMD Research), Chao He (USTC), Yuan He (USTC)</p> <p>A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Xu Chen (USTC), Xu Zhang (USTC), Anirban Basu (USTC), Ai Xiong (USTC), Felix Leuschke (USTC), Yungang Wang (USTC), Yuan He (USTC)</p> <p>Contention-Based Congestion Management in Large-Scale Networks. Jungho Kim (KAIST), Changmin Kim (KAIST), Donghyun Kim (KAIST), Minjae Park (KAIST), Joonhyun Kim (KAIST), Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Gokarna Datta (Ohio University), Travis Stricker (Ohio University), Anirban Basu (Ohio University), Arvind Krishnamoorti (University of Washington)</p> <p>Reducing Data Movement Energy via Online Data Clustering and Encoding. Shihang Wang (University of Rochester), Jingbo Park (University of Rochester)</p> <p>Quantifying and Improving the Efficiency of Hardware-Based Multi-Malware Detectors. Michael Kang (University of Illinois at Austin), Vijay Janapa Reddy (University of Texas at Austin), Mohit Tiwari (University of Texas at Austin)</p> <p>Priority-Scale Specialization for Secure Memory. Venka Subramanian (Ohio State University), Andrew D. Pillemer (Ohio State University), Benjamin C. Lu (Ohio State University)</p> <p>Reply/Confusion: Detecting Cache-Based Covert Channel Attacks Using Record and Replay. Shengyu Liu (University of Illinois at Urbana-Champaign), Jorge Sanjivan (University of Illinois at Urbana-Champaign)</p> <p>Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. Gokarna Datta (Ohio State University), Gokarna Datta (Ohio State University), Neal Abu-Obadiah (UC Riverside)</p>
8:00-9:40	Session 5a: Accelerators 2
9:40-10:00	Break
10:00-11:00	Session 5a: Accelerators 2
11:00-12:10	Session 5a: Accelerators 2
12:30-13:30	Conference Closing and Best Paper Award

Spatio-value similarity...

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Aamer Jaleel

16:00-20:00	Reception	October 16, 2016 (Sunday)
7:00-8:00	Breakfast	October 17, 2016 (Monday)
8:00-8:20	Opening remarks	
8:20-9:20		Keynote I: Internet of Things, Hologram and Hyper-Technology and Policy Margaret Martonosi (University of Maryland)
9:20-10:00	Lightning Session I	
10:00-10:20	Break	
Session 1a: Microarchitecture	Session 1b: Cloud & Storage	
<p>Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches. Binabandhan Panda (NVIDIA Research), Anshu Saxena (NVIDIA Research), Anshu Saxena (NVIDIA Research).</p> <p>Perception Learning for Reuse Prediction. Uthra Iyeran (Intel ASUJ University), Zhe Wang (Intel Labs), Daniel A. Jimenez (Intel ASUJ University).</p> <p>pTack: A Smart Prefetching Scheme for OS Intensive Applications. Prathmesh Kulkarni (IIIT Delhi), Suresh R. Sarangi (IIIT Delhi).</p> <p>Register Sharing for Equality Prediction. Arthur Perias (NVIDIA), Fernando A. Endo (NVIDIA), Andre Semenc (NVIDIA).</p> <p>Data-Centric Execution of Speculative Parallel Programs. Mark C. Jeffrey (Massachusetts Institute of Technology), Suviray Subramanian (Massachusetts Institute of Technology), Mahesh Abeydeya (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology).</p>	<p>Session 1c: Cloud & Storage</p> <p>SABRe: Atomic Object Reads for In-Memory Rack-Scale Computing. Alexander Daglis (EPFL), Dimitri Uzunoglu (EPFL), Marko Novakovic (EPFL), Eduard Bugren (EPFL), Bekir Fakih (EPFL), Boris Grosz (University of Virginia).</p> <p>A Cloud-Scale Acceleration Architecture. Adrian M. Caulfield (Microsoft Research), Eric X. Cheng (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Anepudi (Microsoft), Jeremy Losses (Microsoft Research), Michael Healyman (Microsoft), Stephen Hall (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joon-Yang Kim (Microsoft Research), Daniel Lo (Microsoft Research), Jodi Maswerg (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papernast (Microsoft Research), Lin Woods (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chou (Microsoft), Doug Burger (Microsoft Research).</p> <p>Towards Efficient Server Architecture for Virtualized Network Function Deployment: Implications and Implementations. Yang Hu, and Tao Li (University of Florida).</p> <p>Reaching the I/O Performance Gap for Big Data Workloads: A New NVMMIO-based Approach. Benjie Chen (The Hong Kong Polytechnic University), Jao Li (NDS (University of Florida)).</p> <p>NeSC: Self-Virtualizing Nested Storage Controller. Yonatan Gutfreund (Technion), Yoav Eason (Technion).</p>	
12:00-14:00	Lunch	
14:00-15:40	Poster session	
15:40-16:00	Break	
Session 2a: GPU	Session 2b: Neural Networks	
<p>MIMD Synchronization on SIMD Architectures. Ahmad Alilawati (The University of British Columbia), Tarik A. Amoodt (The University of British Columbia).</p> <p>Efficient Kernel Synthesis for Performance Portable Programming. Li-Wen Chang (University of Illinois at Urbana-Champaign), Lize Li (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Pew Research Center), Christopher Rodrigues (Pew Research Center), Juan Gomez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign).</p> <p>KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism. Lize Li (University of Illinois at Urbana-Champaign), Juan Gomez-Luna (University of Cordoba), Cheng Li (University of Illinois at Urbana-Champaign), Li-Wen Chang (University of Illinois at Urbana-Champaign), Dejan Mijovic (Pewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign).</p> <p>Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPGPUs. Naifeng Jing (Shanghai Jiao Tong University), Jianfei Wang (Shanghai Jiao Tong University), Fengling Fan (Shanghai Jiao Tong University), Wenkang Yu (Shanghai Jiao Tong University), Liang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyan Liang (Shanghai Jiao Tong University).</p> <p>Zorus: A Holistic Approach to Resource Utilization in GPUs. Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gerard Pflueger (Carnegie Mellon University), Semra Khan (University of Virginia), Akshith Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Chiu-Ming (Carnegie Mellon University).</p> <p>GRAPE: Minimizing Energy for GPU Applications with Performance Requirements. Muhammad Husain Sanjib (Curtis University), Henry Hoffmann (University of Illinois at Urbana-Champaign).</p>	<p>From High-Level Deep Neural Models to FPGA. Herdik Sharma (Georgia Institute of Technology), Jorge Park (Georgia Institute of Technology), Emmanuel Amato (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chawika Saha (Georgia Institute of Technology), Aadi Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology).</p> <p>VDRN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design. Minsoo Hwu (NVIDIA), Natalia Gromkova (NVIDIA), Jason Clemens (NVIDIA), Arden Zilber (NVIDIA), Stephen Wu (Intel), NVIDIA).</p> <p>Synapse Bi-Serial Deep Neural Network Computing. Patrick Judd (University of Toronto), Jorge Albancico (University of Toronto), Taylor Hetherington (University of British Columbia), Joe Amoedo (University of British Columbia), Andrew Moshovos (University of Toronto).</p> <p>Combustion-X: An Accelerator for Sparse Neural Networks. Shih-Zheng Institute of Computing Technology, CAS, Ziding Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huaying Lan (Institute of Computing Technology, CAS), Shaoh Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Henshi Chen (Institute of Computing Technology, CAS), Yunfeng Chen (Institute of Computing Technology, CAS).</p> <p>NEURAMS: Neural Network Transformation and Co-design under Neuro-inspired Hardware Constraints. Yu Ji (Tsinghua Univ.), Youfeng Zhang (Tsinghua Univ.), Shuangchen Li (UCSB), Ping Chi (UCSB), Chifang Jiang (Tsinghua Univ.), Peng Du (Tsinghua Univ.), Yuan Xu (UCSB), WenGuo Chen (Tsinghua Univ.).</p> <p>Fixed-Layer CNN Accelerators. Manoj Alwan (Stony Brook University), Han Chen (Stony Brook University), Michael Fedrman (Stony Brook University), Peter Milder (Stony Brook University).</p>	
16:00-18:00	Cache-Emulated Register File: An Integrated On-Chip Memory Architecture for High Performance GPGPUs.	
18:00-20:00	Business meeting	

7:00-8:00	Breakfast	October 18, 2016 (Tuesday)
8:00-9:30		Keynote I: Low Power (CPU, GPU) Mobile to Wearable & IoT Urooj K. Malik (Media Labs)
9:30-10:10	Lightning Session II	
10:10-10:30	Break	
Session 3a: Compilation & Memory	Session 3b: Interconnects	
<p>Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting. Armesh Iyer (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor).</p> <p>CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning. Stephen Jekuty (University of Michigan), Daniel Rings (University of Michigan), Nathan Barata (University of Michigan), Michael Laurenzano (University of Michigan), Lingjie Tang (University of Michigan), Jason Mars (University of Michigan).</p> <p>Low-Cost Soft Error Resilience with Unified Data Verification and Free-Generated Recovery for Acoustic Sensor Based Detection. Qingyu Lu (Virginia Tech), Changshu Jung (Virginia Tech), Dongyong Lee (Virginia Tech), Dewesh Iman (Oak Ridge National Lab).</p> <p>Lazy Release Consistency for GPUs. Johnathan Akrop (University of Illinois at Urbana-Champaign, AMD Research), Man S. Cho (University of Wisconsin - Madison, AMD Research), Bradford M. Beckmann (AMD Research), David A. Wood (University of Wisconsin - Madison, AMD Research).</p> <p>Improving Energy Efficiency of DRAM by Exploiting Half Page Row Access. Hoang Ha (Stanford University), Arslan Ibrahim (Stanford University), Saikat Halder (Stanford University), Shaoh Kwiatkowsky (Technion), Mark Horowitz (Stanford University).</p>	<p>OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures. Iw Zhan (UCSB), Omer Kayran (AMD Research), Gabriel H. Loh (AMD Research), Chih-R. Dow (PSU), Yuan Xie (UCSB).</p> <p>A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers. Jia Zhuo (UCSB), Sir Akshay (UCSB), Jinhui Zhao (UCSB), Ai Davis (HP Labs), Paolo Farabochi (HP Labs), Yungang Wang (Hawaii), Yuan Xie (UCSB).</p> <p>Content-based Network Partitioning in Large-Scale Networks. Geongho Kim (KAIST), Changyeon Kim (KAIST), Ayeon Jeong (KAIST), Mika Parker (Intel), John Kim (KAIST).</p> <p>Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques. Dominic DiMatteo (Ohio University), Brian Bonader (Ohio University), Avinash Kati (Ohio University), Ahmed El-Ghazal (George Washington University).</p> <p>Reducing Data Movement Energy via Online Data Clustering and Encoding. Shihong Wang (University of Rochester), Engin Ipek (University of Rochester).</p>	
10:30-12:10		
12:10-14:10	Award Lunch (including Bob Hwu Award, best of time)	
Session 4a: Multicore	Session 4b: Security	
<p>Racer: TSO Consistency via Race Detection. Alberto Iles (Universidad de Murcia), Stefano Karavas (Uppsala University).</p> <p>Exploiting Semantic Commutativity in Hardware Speculation. Guozhen Zhang (MIT CSAIL), Mahan Kuo (MIT CSAIL), Daniel Sanchez (MIT CSAIL).</p> <p>CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems. Cheechen Chau (Georgia Tech), Aamer Jalil (NVIDIA), Masumuddin K. Qureshi (Georgia Tech).</p> <p>CS1: Mitigating the NUMA Bottleneck via Coherent DRAM Caches. Cheng-Chieh Huang (University of Edinburgh), Saketh Kumar (University of Edinburgh), Marco Iliu (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh).</p>	<p>Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors. Michael Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Mohit Tiwari (University of Iowa, Austin).</p> <p>Poinisnoy: Safe Speculation for Secure Memory. Imane Suberguel Lehmann (Duke University), Andrew D. Hilt (Duke University), Benjamin C. Lee (Duke University).</p> <p>ReplyConfusion: Detecting Cache-based Cover Channel Attacks Using Record and Replay. Mengyao Yan (University of Illinois at Urbana-Champaign), Joseph Torrellas (University of Illinois at Urbana-Champaign).</p> <p>Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR. Urvay Vohra (Stanford University), Dmitry Potomayev (SIUW Birmingham), Naveel Abu-Ghazwan (UK Riverside).</p>	
14:10-15:10		
15:30-16:00	Break	
Session 5a: Approximate Computing	Session 5b: Accelerators 1	
<p>Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation. Aniranth Jain (University of Michigan, Ann Arbor), Parker Hill (University of Michigan, Ann Arbor), Shih-Chieh Lin (University of Michigan, Ann Arbor), Muneeb Khan (Uppsala University), Md E. Haque (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Scott Mahlke (University of Michigan, Ann Arbor), Lingjie Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor).</p> <p>Approxlyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency. Radha Venkatesan (University of Illinois at Urbana-Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana-Champaign), Gita Kumar Sanjay Hari (Nvidia), Santia V. Adve (University of Illinois at Urbana-Champaign).</p> <p>Approxlyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency. Radha Venkatesan (University of Illinois at Urbana-Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana-Champaign), Gita Kumar Sanjay Hari (Nvidia), Santia V. Adve (University of Illinois at Urbana-Champaign).</p> <p>The Bunker Cache for Spatio-Value Approximation. Joshua San-Miguel (University of Toronto), Jorge Albancico (University of Toronto), Natalie Enright Jerger (University of Toronto), Aamer Jalil (NVIDIA).</p>	<p>HARE: Hardware Accelerator for Regular Expressions. Weibiao Gogole (University of Michigan), Aarushesh Koli (University of Michigan), Michael J. Lafferty (University of Michigan), Lena D'Amore (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan).</p> <p>The Microarchitecture of a Real-time Robot Motion Planning Accelerator. Sean Murray (Duke University), Will Floyd-Jones (Duke University), Ying Qi (Duke University), George Korndorff (Duke University), Daniel J. Sornoff (Duke University), Efficient Data Supply for Hardware Accelerators with Prefetching and Access/Execute Decoupling. Joo Chen (Cornell University), G. Edward Suh (Cornell University).</p>	
16:00-17:00		
17:00-21:00	Benquet	

7:00-8:00	Breakfast	October 19, 2016 (Wednesday)
Session 6a: Accelerators 2	Session 6b: Mobile & Power Mgmt	
<p>An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition. Ritesh Vadara Amrabad (Universitat Politècnica de Catalunya (UPC)), Albert Sarguch (Universitat Politècnica de Catalunya (UPC)), Jose-María Armas (Universitat Politècnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politècnica de Catalunya (UPC)).</p> <p>Co-Designing Accelerators and SoC Interfaces using gem5-3D. Yuhang Sun (Harvard University), Sam (Lijun) Zhang (Harvard University), Vijayalakshmi Srivastava (IBM), Gu-Yueun Wei (Harvard University), David Brooks (Harvard University).</p> <p>CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains. Arslan Shafiq (Simon Fraser University), Srinivas Kumar (Simon Fraser University), Apala Guha (Simon Fraser University), Aravindh Shriraman (Simon Fraser University).</p> <p>Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems. Hadi Asghar-Moghaddam (UCL), Young-Hoon Son (SNU), Jung Ho Ahn (SNU), Nam-Sung Kim (UCL).</p>	<p>Evaluating Programmable Architectures for Imaging and Vision Applications. Arslan Venkyi (Stanford), Sanku Ghosh (Stanford), Arslan Venkyi (Stanford), Stephen Richardson (Stanford), Shaoh Kwiatkowsky (Technion), Mark Horowitz (Stanford).</p> <p>Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users. Keyue Fan (University of Houston), Xinyang Zhang (University of Houston), Jengwee Tan (University of Houston), Xin Fu (University of Houston).</p> <p>Switch: Opportunistically Reassigning Power Allocation between Processor and Memory in 2D Stacks. Dimitrios Skafaridis (UCL), Benj Thomas (Intel), Aditya Agrawal (Nvidia), Mihail Qin (UCL), Robert Plonka-Podgorski (UCL), Uwe R. Karpuzcu (ARM), Radu Ionescu (DSU), Nam-Sung Kim (UCL), Joseph Torrellas (UCL).</p> <p>Thalita: Processor Power Management in the Temperature Inversion Region. Yichun Zu (University of Texas at Austin), Wei Huang (AMD Research), Indran Paul (AMD Research), Vijay Janapa Reddi (University of Texas at Austin).</p>	
8:00-9:40		
9:40-10:00	Break	
Session 7: Best Paper Candidates		
<p>Graphicionado: A High-Performance and Energy-Efficient Accelerator for Graph Analytics. Lee Jun-Hwi (Piscataway University), Lina Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nandharath Sarith (Parallel Computing Lab, Intel Corporation), Margaret Martonosi (Piscataway University).</p> <p>Improving Bank-Level Parallelism for Irregular Applications. Xulong Jiang (Penn State), Mahmut Kandemir (Penn State), Venkatesh Yedlapati (Penn State), Agadesh Kotov (Penn State).</p> <p>Delegated Perist Order. Aarushesh Koli (University of Michigan), Jeff Hosen (Intel/Oracle Computing), Stephen Dieckhoff (IBM), Ali Saedi (ARM), Steven Halley (Intel/Oracle Computing), Shuang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan).</p> <p>Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emissions. Nader Shehatahkhah (Georgia Tech), Alexrea Nazari (Georgia Tech), Aleksei Zepi (Georgia Tech), Michel Preucloze (Georgia Tech).</p> <p>Path Confidence based Lookahead Prefetching. Jinchun Kim (Texas A&M University), Seth H. Pugsley (Intel Labs), Paul V. Gratz (Texas A&M University), A. L. Naresanba Reddy (Texas A&M University), Chris Willerson (Intel Labs), Anshu Chhabra (Intel Labs), Urvay Vohra (UCL), Nave N. Puri (UT Austin).</p>		
10:00-12:00		
Session 8: Conference Closing and Best Paper Award		
12:30-18:15	Conference Outing (includes sack lunch)	

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation
 Session 5a, Tuesday 4:40pm
 Joshua San Miguel, Jorge Albancico, Nataie Enright Jerger and Aamer Jalil

October 16, 2016 (Sunday)		October 17, 2016 (Monday)		October 18, 2016 (Tuesday)		October 19, 2016 (Wednesday)	
18:00-20:00	Reception	October 17, 2016 (Monday)		7:00-8:00	Breakfast	7:00-8:00	Breakfast
7:00-8:00	Breakfast	October 17, 2016 (Monday)		8:00-9:30	Breakfast	8:00-9:30	Session 6a: Accelerators 2
8:00-8:20	Opening remarks	October 17, 2016 (Monday)		9:30-10:10	Lightning Session II	8:00-9:40	Session 6a: Accelerators 2
8:20-9:20	Lightning Session I	October 17, 2016 (Monday)		10:10-10:30	Break	8:00-9:40	Session 6a: Accelerators 2
10:00-10:20	Break	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2
10:20-12:00	Session 1a: Microarchitecture	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2
12:00-14:00	Lunch	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2
14:00-15:40	Poster session	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2
15:40-16:00	Break	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2
16:00-18:00	Session 2a: GPU	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2
18:00-20:00	Business meeting	October 17, 2016 (Monday)		10:30-12:10	Session 3a: Compilation & Memory	8:00-9:40	Session 6a: Accelerators 2

Session 2b: Neural Networks

From High-Level Deep Neural Models to FPGA: Harsh Kohli (Georgia Institute of Technology), Jorge Park (Georgia Institute of Technology), Emmanuel Anato (Georgia Institute of Technology), Joonhyun Kim (Georgia Institute of Technology), Chankai Shao (Georgia Institute of Technology), Ashi Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

VDFN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design: Minsoo Ryu (NVIDIA), Natalia Gromkova (NVIDIA), Jason Clemens (NVIDIA), Arden Zilber (NVIDIA), Songhan Wu (NVIDIA)

StyxNet: Bit-Serial Deep Neural Network Computing: Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Taylor Hetherington (University of British Columbia), Yifan Aernold (University of British Columbia), Andrew Moshovos (University of Toronto)

Combustion-X: An Accelerator for Sparse Neural Networks: Shih Zhang (Institute of Computing Technology, CAS), Ziding Yu (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huaying Lan (Institute of Computing Technology, CAS), Shaoh Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Hengshu Chen (Institute of Computing Technology, CAS), Yunfeng Chen (Institute of Computing Technology, CAS)

NEUTRANS: Neural Network Transformation and Co-design under Neuro-morphic Hardware Constraints: Yu Li (Tsinghua Univ.), Youfeng Zhang (Tsinghua Univ.), Shuangchen Li (UCSB), Jing Chi (UCSB), Chifang Jiang (Tsinghua Univ.), Peng Du (Tsinghua Univ.), Yuan Xu (UCSB), Wenqiang Chen (Tsinghua Univ.)

Fluxed-Layer CNN Accelerators: Manoj Ahasani (Stony Brook University), Han Chen (Stony Brook University), Michael Jederman (Stony Brook University), Peter Milder (Stony Brook University)

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation
 Session 5a, Tuesday 4:40pm
 Joshua San Miguel, Jorge Albericio, Natalia Enright Jerger and Amer Jaleel

October 16, 2016 (Sunday)		October 17, 2016 (Monday)		October 18, 2016 (Tuesday)		October 19, 2016 (Wednesday)	
7:00-8:00	Reception	7:00-8:00	Breakfast	7:00-8:00	Breakfast	7:00-8:00	Breakfast
8:00-9:00	Opening remarks	8:00-9:00	Keynote 1: Lionel Dreyer (EPFL) , Tommy Martin (University of Toronto)	8:00-9:00	Keynote 2: Lionel Dreyer (EPFL) , Tommy Martin (University of Toronto)	8:00-9:00	Session 6a: Accelerators 2
9:20-9:30	Lightning Session I	9:20-9:30	Lightning Session I	9:30-10:10	Lightning Session II	9:30-10:10	Lightning Session II
10:00-10:20	Break	10:00-10:20	Break	10:10-10:30	Break	10:10-10:30	Break
10:30-12:00	Session 5a: Microarchitecture	10:30-12:00	Session 5a: Microarchitecture	10:30-12:00	Session 5a: Microarchitecture	10:30-12:00	Session 5a: Microarchitecture
12:00-14:00	Lunch	12:00-14:00	Lunch	12:00-14:00	Lunch	12:00-14:00	Lunch
14:00-15:40	Poster session	14:00-15:40	Poster session	14:00-15:40	Poster session	14:00-15:40	Poster session
15:40-16:00	Break	15:40-16:00	Break	15:40-16:00	Break	15:40-16:00	Break
16:00-18:00	Session 5a: GPU	16:00-18:00	Session 5a: GPU	16:00-18:00	Session 5a: GPU	16:00-18:00	Session 5a: GPU
18:00-20:00	Business meeting	18:00-20:00	Business meeting	18:00-20:00	Business meeting	18:00-20:00	Business meeting

Session 5a: GPU

From High-Level Deep Neural Models to FPGA, Herdik Sharma (Georgia Institute of Technology), Jorge Park (Georgia Institute of Technology), Emmanuel Anato (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chawki Shah (Georgia Institute of Technology), Aad Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

VDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Ryu (University of California, Berkeley), Stephen W. Keckler (NVIDIA), Srinivas Aravamudan (Google), Patrick Juett (University of Toronto), Jorge Albericio (University of Toronto), Taylor Hetherington (University of British Columbia), Anand Rajaraman (University of British Columbia)

Combination-K: An Accelerator for Sparse Neural Networks, Shih-Zhen Lin (Institute of Computing Technology, CAS), Ziding Yu (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huaying Lan (Institute of Computing Technology, CAS), Shaoh Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Hengshu Chen (Institute of Computing Technology, CAS), Yung-Chieh Lin (Institute of Computing Technology, CAS)

NEURTRAMS: Neural Network Transformation and Co-design under Neuroomorphic Hardware Constraints, Yu Li (Tsinghua Univ.), Youfeng Zhang (Tsinghua Univ.), Shuangchen Li (UCSB), Jing Chi (UCSB), Chifang Jiang (Tsinghua Univ.), Peng Du (Tsinghua Univ.), Yuan Xu (UCSB), Wenqiang Chen (Tsinghua Univ.)

Fixed-Layer CNN Accelerators, Manoj Ahasani (Stony Brook University), Han Chen (Stony Brook University), Michael Jederman (Stony Brook University), Peter Milder (Stony Brook University)

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation
 Session 5a, Tuesday 4:40pm
 Joshua San Miguel, Jorge Albericio, Nataïe Enright Jerger and Amer Jaleel

October 16, 2016 (Sunday)		October 16, 2016 (Monday)		October 17, 2016 (Tuesday)		October 18, 2016 (Wednesday)	
18:00-20:00 Reception		7:00-8:00 Breakfast		7:00-8:00 Breakfast		7:00-8:00 Breakfast	
7:00-8:00 Breakfast		8:00-9:30 Breakfast		8:00-9:30 Breakfast		8:00-9:30 Breakfast	
8:00-8:20 Opening Remarks		9:30-10:10 Lighting Session II		9:30-10:10 Lighting Session II		9:30-10:10 Lighting Session II	
8:20-9:00		10:10-10:30 Break		10:10-10:30 Break		10:10-10:30 Break	
9:20-10:00 Lighting Session I		10:30-11:00 Session 3a: Compilation & Memory		10:30-11:00 Session 3a: Compilation & Memory		10:30-11:00 Session 3a: Compilation & Memory	
10:00-10:20 Break		11:00-11:30 Session 3b: Security		11:00-11:30 Session 3b: Security		11:00-11:30 Session 3b: Security	
10:20-12:00 Session 1a: Microarchitecture		12:10-14:10 Award Lunch (including Bob Haward, lead of Intel)		12:10-14:10 Award Lunch (including Bob Haward, lead of Intel)		12:10-14:10 Award Lunch (including Bob Haward, lead of Intel)	
12:00-14:00 Lunch		14:10-15:10 Session 3c: Accelerators 1		14:10-15:10 Session 3c: Accelerators 1		14:10-15:10 Session 3c: Accelerators 1	
14:00-15:40 Poster session		15:10-16:00 Break		15:10-16:00 Break		15:10-16:00 Break	
15:40-18:00 Break		16:00-17:00 Session 3d: Accelerators 2		16:00-17:00 Session 3d: Accelerators 2		16:00-17:00 Session 3d: Accelerators 2	
18:00-20:00 Business meeting		17:00-21:00 Banquet		17:00-21:00 Banquet		17:00-21:00 Banquet	

Last-level cache

Session 5a: Neural Networks

From High-Level Deep Neural Models to FPGA, Harkit, Intel Labs (University of Toronto), George J. Gordon (Georgia Institute of Technology), George Katagiri (University of Texas at Austin), Emmanuel Anasto (Georgia Institute of Technology), Jun Young Kim (Georgia Institute of Technology), Jiankai Shi (Georgia Institute of Technology), Aidan Mallya (Intel Corporation), Rajkumar Venkatesh (Georgia Institute of Technology)

VDFN: Virtualized Deep Neural Networks for Scalable Memory-Efficient Neural Network Pruning, Srinivas Aravamudan (Google), Stephen Chong (Google), Arvind Krishnamoorti (University of Toronto), Jorge Albericio (University of Toronto), Sanku Han (Heriot-Watt University of British Columbia)

Bit-Sized Deep Neural Network Computing, Daniel Feldman (University of Toronto), Jorge Albericio (University of Toronto), Sanku Han (Heriot-Watt University of British Columbia)

Combining It As Accelerator for Sparse Neural Networks, Shih Zhang (Institute of Computing Technology, CAS), Jintao Xu (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Ruiyuan Liu (Institute of Computing Technology, CAS), Shuai Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Jianshi Chen (Institute of Computing Technology, CAS), Yong Chen (Institute of Computing Technology, CAS)

NEURAMS: Neural Network Transformation and Co-design Under Neuroorphic Hardware Constraints, Y. J. Peng (MIT), Weidong Zhang (Tsinghua Univ.), Shuang-Jin Li (UCSB), Yong-Lin Li (UCSB), Chihang Song (Tsinghua Univ.), Peng Gu (Tsinghua Univ.), Yuen Kw (UCSB), Wang-Gang Chen (Tsinghua Univ.)

Fused-Layer CPUs Accelerators, Minye Albert (Tsinghua University), Han Chen (Tsinghua University), Michael Feldman (Stanford University), Peter Miller (Stanford University)

The Bunker Cache for Spatio-Value Approximation
 Session 5a, Tuesday 4:40pm
 Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Amer Jaleel

October 16, 2016 (Sunday)	October 17, 2016 (Monday)	October 18, 2016 (Tuesday)	October 19, 2016 (Wednesday)
7:00-8:00 Breakfast	7:00-8:00 Breakfast	7:00-8:00 Breakfast	7:00-8:00 Breakfast
8:00-8:20 Opening remarks	8:00-8:20 Opening remarks	8:00-9:30 Break	8:00-9:40 Break
8:20-9:20	8:20-9:20	9:30-10:10 Lighting Session II	9:40-10:00 Break
9:20-10:00 Lighting Session I	9:20-10:00 Lighting Session I	10:10-10:30 Break	10:00-12:00 Break
10:00-10:20 Break	10:00-10:20 Break	10:30-12:10	10:00-12:00
Session 1a: Microarchitecture	Session 1a: Microarchitecture	Session 2a: CPU	Session 2a: CPU
Session 1b: Storage	Session 1b: Storage	Session 2b: GPU	Session 2b: GPU
Session 1c: Memory	Session 1c: Memory	Session 2c: Accelerators	Session 2c: Accelerators
Session 1d: Security	Session 1d: Security	Session 2d: Security	Session 2d: Security
Session 1e: Networking	Session 1e: Networking	Session 2e: Accelerators	Session 2e: Accelerators
Session 1f: Emerging Topics	Session 1f: Emerging Topics	Session 2f: Emerging Topics	Session 2f: Emerging Topics
12:00-14:00 Lunch	12:00-14:00 Lunch	12:10-14:10	12:10-14:10
14:00-15:40 Poster session	14:00-15:40 Poster session	14:10-15:10	14:10-15:10
15:40-16:00 Break	15:40-16:00 Break	15:10-16:00 Break	15:10-16:00 Break
16:00-18:00	16:00-18:00	16:00-17:00	16:00-17:00
18:00-20:00 Business meeting	18:00-20:00 Business meeting	17:00-21:00 Banquet	17:00-21:00 Banquet

Last-level cache

From High-Level Deep Neural Models to FPGA, Harkit, Georgia Institute of Technology, Joseph Park (Georgia Institute of Technology), Deepak Mahajan (Georgia Institute of Technology), Joon Young Kim (Georgia Institute of Technology), Chanki Sha (Georgia Institute of Technology), Aid Mikhie (Intel Corporation), Rajkumar Venkatesan (Georgia Institute of Technology)

VDFN: Virtualized Deep Neural Networks for Scalable Memory-Efficient Inference, Stephen Ho, Aamer Jaleel, Virginia Tech University, Jorgo Albericio, Jorge Albericio (University of Toronto), Jorgo Albericio (University of Toronto), Aamer Jaleel (University of British Columbia), Jorgo Albericio (University of Toronto)

Combining it As an Accelerator for Sparse Neural Networks, Peng Zhang (Institute of Computing Technology, CAS), Jintao Xu (Institute of Computing Technology, CAS), Lei Zhang (Institute of Chinese Academy of Sciences), Ruiyong Lan (Institute of Computing Technology, CAS), Shuai Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Jiansi Chen (Institute of Computing Technology, CAS), Yong Chen (Institute of Computing Technology, CAS)

NEURFAMS: Neural Network Transformation and Co-design Under Neurographic Hardware Constraints, Y. F. Cheng (Intel), Nathan Chen (Intel), Y. Li (Intel), Shuang Jin (UCSD), Jing Liu (UCSD), Chihang Song (Stanford), Peng Du, Peng Du, Yonghua Luo, Yuan Ke (UCSD), Wangjiao Chen (Tsinghua Univ.)

Fluxed-Layer CPUs Accelerators, Minso Ahn (Sungkyunkwan University), Han Chen (Sungkyunkwan University), Michael Weidman (Sungkyunkwan University), Peter Miller (Sungkyunkwan University)

1.58x runtime, 1.72x dynamic, 1.65x leakage savings!

The Bunker Cache for Spatio-Value Approximation
 Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Aamer Jaleel