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: History and Hype, Technology and Policy
Margaret Martonosi (Princeton)

9:20-10:00 Lightning Session I

10:00-10:20 Break

Session 1a: Microarchitecture

Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches, Biswabandan Panda (INRIA Rennes), André Seznec (INRIA Rennes)

Perceptron Learning for Reuse Prediction, Elvira Teran (Texas A&M University), Zhe Wang (Intel Labs), Daniel A. Jiménez (Texas A&M University)

pTask: A Smart Prefetching Scheme for OS Intensive Applications, Prathmesh Kallurkar (IIT Delhi), Smruti R. Sarangi (IIT Delhi)

Register Sharing for Equality Prediction, Arthur Perais (INRIA), Fernando A. Endo (INRIA), André Seznec (INRIA)

Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvinay Subramanian (Massachusetts Institute of Technology), Maleen Abeydeera (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)

10:20-12:00

Session 1b: Cloud & Storage

SABRes: Atomic Object Reads for In-Memory Rack-Scale Computing, Alexandros Daglis (EPFL), Dmitri Ustiugov (EPFL), Stanko Novaković (EPFL), Edouard Bugnion (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)

A Cloud-Scale Acceleration Architecture, Adrian M. Caulfield (Microsoft Research), Eric S. Chung (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angepat (Microsoft), Jeremy Fowers (Microsoft Research), Michael Haselman (Microsoft), Stephen Heil (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joo-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Todd Massengill (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papamichael (Microsoft Research), Lisa Woods (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chiou (Microsoft), Doug Burger (Microsoft Research)

Towards 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 NVDIMM-based Approach, Renhai Chen (The Hong Kong Polytechnic University), Zili Shao (The Hong Kong Polytechnic University), Tao Li (NSI /University of Florida)

NeSC: Self-Virtualizing Nested Storage Controller, Yonatan Gottesman (Technion), Yoav Etsion (Technion)

12:00-14:00 Lunch

14:00-15:40 Poster session

15:40-16:00 Break

Session 2a: GPU

MIMD Synchronization on SIMT Architectures, Ahmed El-Mahdy (The University of British Columbia), Ibrahim M. Aamodt (The University of British Columbia)

Efficient Kernel Synthesis for Performance Portable Programming, Li-Wen Chang (University of Illinois at Urbana-Champaign), Izzat El Hajj (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Huawei America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism, Izzat El Hajj (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 Milojicic (Hewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

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), Li Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyao Liang (Shanghai Jiao Tong University)

Zorua: A Holistic Approach to Resource Virtualization in GPUs, Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gennady Pekhimenko (Carnegie Mellon University), Samira Khan (University of Virginia), Ashish Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Onur Mutlu (Carnegie Mellon University)

GRAPE: Minimizing Energy for GPU Applications with Performance Requirements, Muhammad Husni Santriaji (Surya University), Henry Hoffmann (University of Chicago)

Session 2b: Neural Networks

From High-Level Deep Neural Models to FPGAs, Hardik Sharma (Georgia Institute of Technology), Jongse Park (Georgia Institute of Technology), Divya Mahajan (Georgia Institute of Technology), Emmanuel Amaro (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chenkai Shao (Georgia Institute of Technology), Asit Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

vDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Rhu (NVIDIA), Natalia Gimelshein (NVIDIA), Jason Clemons (NVIDIA), Arslan Zulfiqar (NVIDIA), Stephen W. Keckler (NVIDIA)

Stripes: Bit-Serial Deep Neural Network Computing, Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Tayler Hetherington (University of British Columbia), Tor Aamodt (University of British Columbia), Andreas Moshovos (University of Toronto)

Cambricon-X: An Accelerator for Sparse Neural Networks, Shijin Zhang (Institute of Computing Technology, CAS), Zidong Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huiying Lan (Institute of Computing Technology, CAS), Shaoli Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Tianshi Chen (Institute of Computing Technology, CAS), Yunji Chen (Institute of Computing Technology, CAS)

NEUTRAMS: Neural Network Transformation and Co-design under Neuromorphic Hardware Constraints, Yu Ji (Tsinghua Univ.), Youhui Zhang (Tsinghua Univ.), ShuangChen Li (UCSB), Ping Chi (UCSB), Cihang Jiang (Tsinghua Univ.), Peng Qu (Tsinghua Univ.), Yuan Xie (UCSB), WenGuang Chen (Tsinghua Univ.)

Fused-Layer CNN Accelerators, Manoj Alwani (Stony Brook University), Han Chen (Stony Brook University), Michael Ferdman (Stony Brook University), Peter Milder (Stony Brook University)

18:00-20:00 Business meeting

October 18, 2016 (Tuesday)

7:00-8:30 Breakfast

8:30-9:30 Keynote II: Low Power GPU: From Mobile to Wearable & IoT
Umang Ko (Mediatek)

9:30-10:10 Lightning Session II

10:10-10:30 Break

Session 3a: Compilation & Memory

Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting, Animesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning, Stephen Zekany (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjia Tang (University of Michigan), Jason Mars (University of Michigan)

Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection, Qingrui Liu (Virginia Tech), Changhee Jung (Virginia Tech), Dongyoon Lee (Virginia Tech), Devesh Tiwari (Oak Ridge National Lab)

Lazy Release Consistency for GPUs, Johnathan Alsop (University of Illinois at Urbana-Champaign, AMD Research), Marc S. Orr (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, Hoonjae Ha (Stanford University), Ardavan Pedram (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)

10:30-12:10

Session 3b: Interconnect

OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures, Jia Zhan (UCSB), Onur Kayiran (AMD Research), Gabriel H. Loh (AMD Research), Chita R. Das (PSU), Yuan Xie (UCSB)

A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers, Jia Zhan (UCSB), Itir Akgun (UCSB), Jishen Zhao (UCSC), Al Davis (HP Labs), Paolo Faraboschi (HP Labs), Yuangang Wang (Huawei), Yuan Xie (UCSB)

Contention-based Congestion Management in Large-Scale Networks, Gwangsun Kim (KAIST), Changhyun Kim (KAIST), Jiyun Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)

Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques, Dominic DiTomaso (Ohio University), Travis Boraten (Ohio University), Avinash Kodi (Ohio University), Ahmed Louri (George Washington University)

Reducing Data Movement Energy via Online Data Clustering and Encoding, Shibo Wang (University of Rochester), Engin Ipek (University of Rochester)

12:10-14:10 Award Lunch: (including Bob Rau Award, Test of Time)

Session 4a: Multicore

Racer: TSO Consistency via Race Detection, Alberto Ros (Universidad de Murcia), Stefanos Kaxiras (Uppsala Universitet)

Exploiting Semantic Commutativity in Hardware Speculation, Guowei Zhang (MIT CSAIL), Virginia Chiu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)

CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems, Chiachen Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Moinuddin K. Qureshi (Georgia Tech)

14:10-15:30

CSI: Mitigating the NUMA Bottleneck via Coherent DRAM Caches, Cheng-Chieh Huang (University of Edinburgh), Rakesh Kumar (University of Edinburgh), Marco Elver (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)

Session 4b: Security

Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors, Mikhail Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Mohit Tiwari (University of Texas at Austin)

PoisonIvy: Safe Speculation for Secure Memory, Tamara Silbergleit Lehman (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University)

ReplayConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay, Mengjia Yan (University of Illinois at Urbana Champaign), Yasser Shalabi (University of Illinois at Urbana Champaign), Josep Torrellas (University of Illinois at Urbana Champaign)

Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR, Dmitry Evtyushkin (SUNY Binghamton), Dmitry Ponomarev (SUNY Binghamton), Nael Abu-Ghazaleh (UC Riverside)

15:30-16:00 Break

Session 5a: Approximate Computing

Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation, Animesh 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 Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

16:00-17:00

Approxilyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency, Radha Venkatagiri (University of Illinois at Urbana Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana Champaign), Siva Kumar Sastry Hari (Nvidia), Sarita 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), Natalie Enright Jerger (University of Toronto), Aamer Jaleel (NVIDIA)

Session 5b: Accelerators 1

HARE: Hardware Accelerator for Regular Expressions, Vaibhav Gogte (University of Michigan), Aasheesh Kolli (University of Michigan), Michael J. Cafarella (University of Michigan), Loris D'Antoni (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan)

The Microarchitecture of a Real-time Robot Motion Planning Accelerator, Sean Murray (Duke University), Will Floyd-Jones (Duke University), Ying Qi (Duke University), George Konidaris (Duke University), Daniel J. Sorin (Duke University)

Efficient Data Supply for Hardware Accelerators with Prefetching and Access/Execute Decoupling, Tao Chen (Cornell University), G. Edward Suh (Cornell University)

17:00-21:00 Banquet

October 19, 2016 (Wednesday)

7:00-8:00 Breakfast

Session 6a: Accelerators 2

An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition, Reza Yazdani Aminabadi (Universitat Politecnica de Catalunya (UPC)), Albert Segura (Universitat Politecnica de Catalunya (UPC)), Jose-Maria Arnau (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC))

Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin, Yakun Sophia Shao (Harvard University), Sam (Likun) Xi (Harvard University), Vijayalakshmi Srinivasan (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains, Armin Sheriflan (Simon Fraser University), Snehasish Kumar (Simon Fraser University), Apala Guha (Simon Fraser University), Arrvindh Shriraman (Simon Fraser University)

Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems, Hadi Asghari-Moghaddam (UIUC), Young Hoon Son (SNU), Jung Ho Ahn (SNU), Nam Sung Kim (UIUC)

8:00-9:40

Session 6b: Mobile & Power Mgmt

A Patch Memory System For Image Processing and Computer Vision, Jason Clemons (NVIDIA), Chih Chi Cheng (NVIDIA), Iuri Frosio (NVIDIA), Daniel Johnson (NVIDIA), Steve Keckler (NVIDIA)

Evaluating Programmable Architectures for Imaging and Vision Applications, Artem Vasilyev (Stanford), Nikhil Bhagdikar (Stanford), Ardavan Pedram (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)

Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users, Kaige Yan (University of Houston), Xingyao Zhang (University of Houston), Jingweija Tan (University of Houston), Xin Fu (University of Houston)

Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks, Dimitrios Skarlatos (UIUC), Reny Thomas (Intel), Aditya Agrawal (Nvidia), Shibin Qin (UIUC), Robert Pilawa-Podgurski (UIUC), Ulya R. Karpuzcu (UMN), Radu Teodorescu (OSU), Nam Sung Kim (UIUC), Josep Torrellas (UIUC)

Ti-states: Processor Power Management in the Temperature Inversion Region, Yazhou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indrani Paul (AMD Research), Vijay Janapa 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, Tae Jun Ham (Princeton University), Lisa Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nadathur Satish (Parallel Computing Lab, Intel Corporation), Margaret Martonosi (Princeton University)

Improving Bank-Level Parallelism for Irregular Applications, Xulong Tang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapalli (Penn State), Jagadish Kotra (Penn State)

Delegated Persist Ordering, Aasheesh Kolli (University of Michigan), Jeff Rosen (Snowflake Computing), Stephan Diestelhorst (ARM), Ali Saidi (ARM), Steven Pelley (Snowflake Computing), Sihang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan)

10:00-12:00

Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emanations, Nader Sehatbakhsh (Georgia Tech), Alireza Nazari (Georgia Tech), Alenka Zajic (Georgia Tech), Milos Prvulovic (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 Wilkerson (Intel Labs), Zeshan Chishti (Intel Labs)

Continuous Runahead: Transparent Hardware Acceleration for Memory Intensive Workloads, Milad Hashemi (UT Austin), Onur Mutlu (CMU), Yale N. Patt (UT Austin)

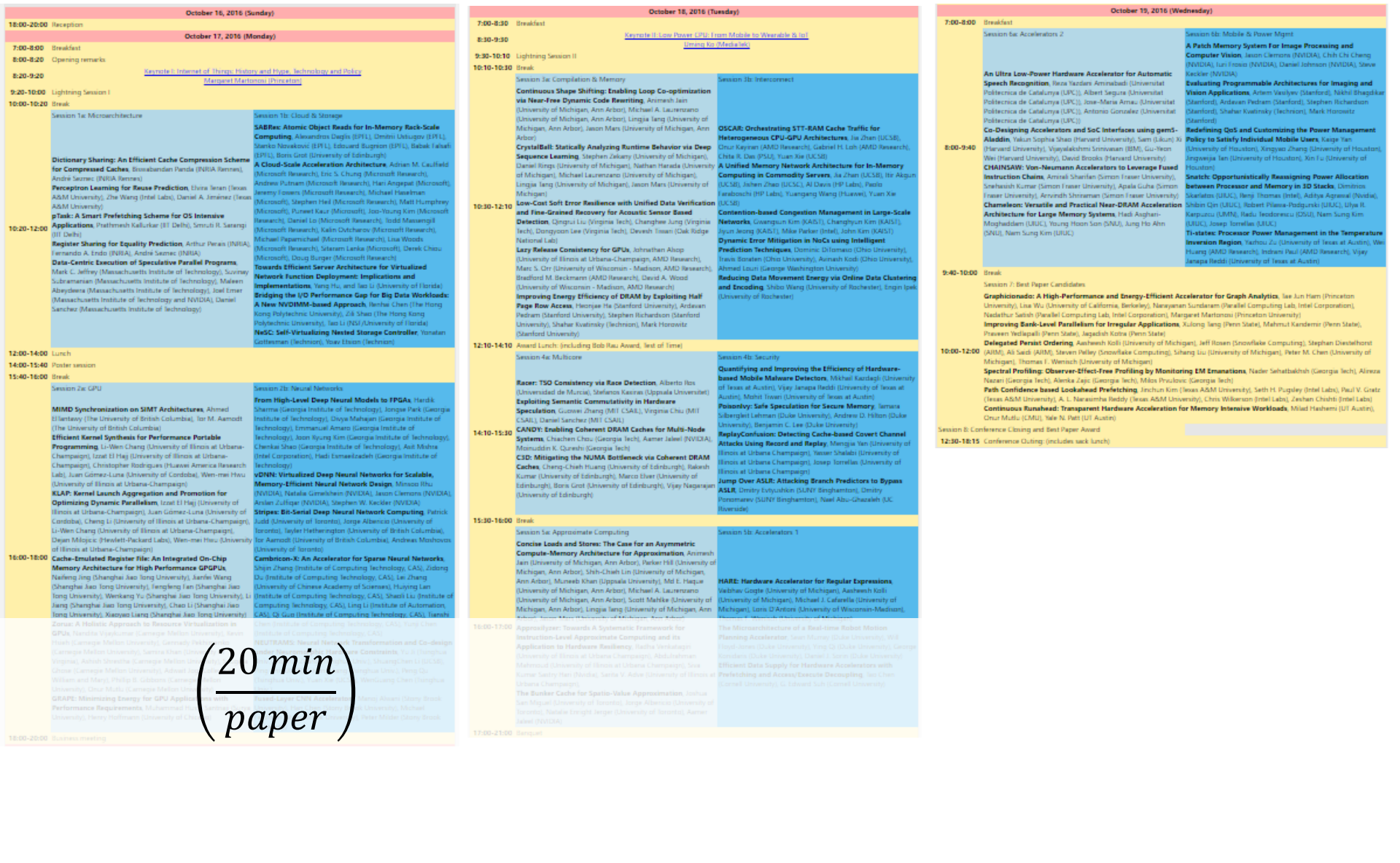
Session 8: Conference Closing and Best Paper Award

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, Natalie Enright Jerger and Aamer Jaleel

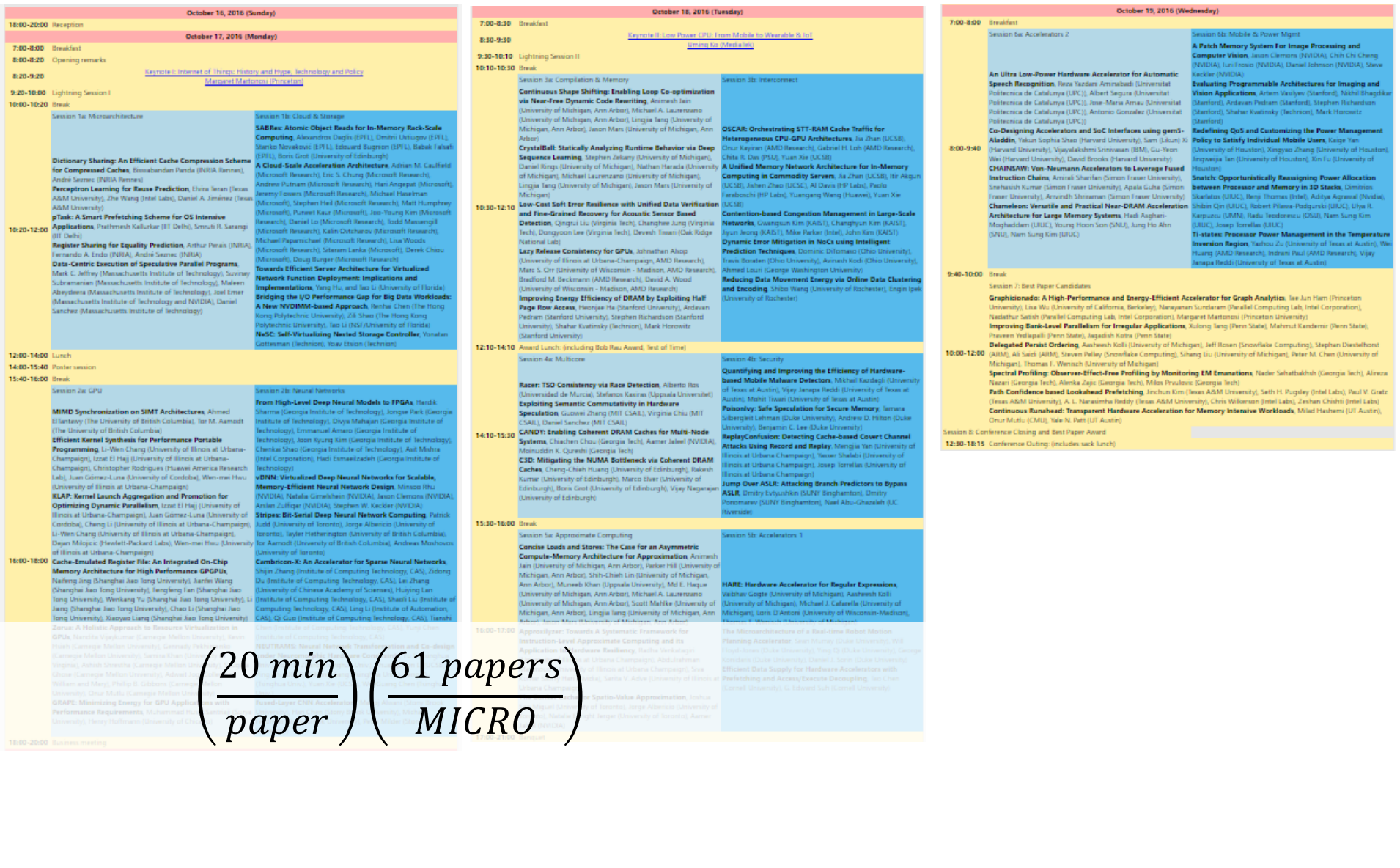


$$\left(\frac{20\ min}{paper}\right)$$

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: History and Hype, Technology and Policy
Margaret Martonosi (Princeton)

9:20-10:00 Lightning Session I

10:00-10:20 Break

Session 1a: Microarchitecture

Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches, Biswabandan Panda (INRIA Rennes), André Seznec (INRIA Rennes)

Perceptron Learning for Reuse Prediction, Elvira Teran (Texas A&M University), Zhe Wang (Intel Labs), Daniel A. Jiménez (Texas A&M University)

pTask: A Smart Prefetching Scheme for OS Intensive Applications, Prathmesh Kallurkar (IIT Delhi), Smruti R. Sarangi (IIT Delhi)

Register Sharing for Equality Prediction, Arthur Perais (INRIA), Fernando A. Endo (INRIA), André Seznec (INRIA)

Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvinay Subramanian (Massachusetts Institute of Technology), Maleen Abeydeera (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)

10:20-12:00

Session 1b: Cloud & Storage

SABRes: Atomic Object Reads for In-Memory Rack-Scale Computing, Alexandros Daglis (EPFL), Dmitri Ustiugov (EPFL), Stanko Novaković (EPFL), Edouard Bugnion (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)

A Cloud-Scale Acceleration Architecture, Adrian M. Caulfield (Microsoft Research), Eric S. Chung (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angepat (Microsoft), Jeremy Fowers (Microsoft Research), Michael Haselman (Microsoft), Stephen Heil (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joo-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Todd Massengill (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papamichael (Microsoft Research), Lisa Woods (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chiou (Microsoft), Doug Burger (Microsoft Research)

Towards 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 NVDIMM-based Approach, Renhai Chen (The Hong Kong Polytechnic University), Zili Shao (The Hong Kong Polytechnic University), Tao Li (NSI /University of Florida)

NeSC: Self-Virtualizing Nested Storage Controller, Yonatan Gottesman (Technion), Yoav Etsion (Technion)

12:00-14:00 Lunch

14:00-15:40 Poster session

15:40-16:00 Break

Session 2a: GPU

MIMD Synchronization on SIMT Architectures, Ahmed ElTantawy (The University of British Columbia), Tor M. Aamodt (The University of British Columbia)

Efficient Kernel Synthesis for Performance Portable Programming, Li-Wen Chang (University of Illinois at Urbana-Champaign), Izzat El Hajj (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Huawei America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism, Izzat El Hajj (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 Milojicic (Hewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

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), Li Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyao Liang (Shanghai Jiao Tong University)

Session 2b: Neural Networks

From High-Level Deep Neural Models to FPGAs, Hardik Sharma (Georgia Institute of Technology), Jongse Park (Georgia Institute of Technology), Divya Mahajan (Georgia Institute of Technology), Emmanuel Amaro (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chenkai Shao (Georgia Institute of Technology), Asit Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

vDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Rhu (NVIDIA), Natalia Gimelshein (NVIDIA), Jason Clemons (NVIDIA), Arslan Zulfiqar (NVIDIA), Stephen W. Keckler (NVIDIA)

Stripes: Bit-Serial Deep Neural Network Computing, Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Tayler Hetherington (University of British Columbia), Tor Aamodt (University of British Columbia), Andreas Moshovos (University of Toronto)

Cambricon-X: An Accelerator for Sparse Neural Networks, Shijin Zhang (Institute of Computing Technology, CAS), Zidong Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huiying Lan (Institute of Computing Technology, CAS), Shaoli Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Tianshi Chen (Institute of Computing Technology, CAS), Yunji Chen (Institute of Computing Technology, CAS)

Zorua: A Holistic Approach to Resource Virtualization in GPUs, Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gennady Pekhimenko (Carnegie Mellon University), Samira Khan (Carnegie Mellon University), Ashish Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Onur Mutlu (Carnegie Mellon University)

GRAPE: Minimizing Energy for GPU Applications with Performance Requirements, Mohammad Hedayati (University of Rochester), Henry Hoffmann (University of Chicago), Michael L. Scott (University of Rochester)

18:00-20:00 Business meeting

October 18, 2016 (Tuesday)

7:00-8:30 Breakfast

8:30-9:30 Keynote II: Low Power GPU: From Mobile to Wearable & IoT
Ulrang Ko (Mediatek)

9:30-10:10 Lightning Session II

10:10-10:30 Break

Session 3a: Compilation & Memory

Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting, Animesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning, Stephen Zekany (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjia Tang (University of Michigan), Jason Mars (University of Michigan)

Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection, Qingrui Liu (Virginia Tech), Changhee Jung (Virginia Tech), Dongyoon Lee (Virginia Tech), Devesh Tiwari (Oak Ridge National Lab)

Lazy Release Consistency for GPUs, Johnathan Alsop (University of Illinois at Urbana-Champaign, AMD Research), Marc S. Orr (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, Hoonjae Ha (Stanford University), Ardavan Pedram (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)

10:30-12:10

Session 3b: Interconnect

OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures, Jia Zhan (UCSB), Onur Kayiran (AMD Research), Gabriel H. Loh (AMD Research), Chita R. Das (PSU), Yuan Xie (UCSB)

A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers, Jia Zhan (UCSB), Itir Akgun (UCSB), Jishen Zhao (UCSC), Al Davis (HP Labs), Paolo Faraboschi (HP Labs), Yuangang Wang (Huawei), Yuan Xie (UCSB)

Contention-based Congestion Management in Large-Scale Networks, Gwangsun Kim (KAIST), Changhyun Kim (KAIST), Jiyun Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)

Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques, Dominic DiTomaso (Ohio University), Travis Boraten (Ohio University), Avinash Kodi (Ohio University), Ahmed Louri (George Washington University)

Reducing Data Movement Energy via Online Data Clustering and Encoding, Shibo Wang (University of Rochester), Engin Ipek (University of Rochester)

12:10-14:10 Award Lunch: (including Bob Rau Award, Test of Time)

Session 4a: Multicore

Racer: TSO Consistency via Race Detection, Alberto Ros (Universidad de Murcia), Stefanos Kaxiras (Uppsala Universitet)

Exploiting Semantic Commutativity in Hardware Speculation, Guowei Zhang (MIT CSAIL), Virginia Chiu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)

CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems, Chiachen Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Moinuddin K. Qureshi (Georgia Tech)

CSD: Mitigating the NUMA Bottleneck via Coherent DRAM Caches, Cheng-Chieh Huang (University of Edinburgh), Rakesh Kumar (University of Edinburgh), Marco Elver (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)

14:10-15:30

Session 4b: Security

Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors, Mikhail Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Mohit Tiwari (University of Texas at Austin)

PoisonIvy: Safe Speculation for Secure Memory, Tamara Silbergleit Lehman (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University)

ReplayConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay, Mengjia Yan (University of Illinois at Urbana Champaign), Yasser Shalabi (University of Illinois at Urbana Champaign), Josep Torrellas (University of Illinois at Urbana Champaign)

Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR, Dmitry Evtyushkin (SUNY Binghamton), Dmitry Ponomarev (SUNY Binghamton), Nael Abu-Ghazaleh (UC Riverside)

15:30-16:00 Break

Session 5a: Approximate Computing

Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation, Animesh 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 Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

Session 5b: Accelerators 1

HARE: Hardware Accelerator for Regular Expressions, Vaibhav Gogte (University of Michigan), Aasheesh Kolli (University of Michigan), Michael J. Cafarella (University of Michigan), Loris D'Antoni (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan)

16:00-17:00

Approximate Storage in Solid-State Drives, Jacob Nelson Sartori (University of Minnesota), John Sloan (University of Minnesota), Keshab K. Parhi (University of Minnesota), Rakesh Kumar (University of Illinois at Urbana-Champaign)

ApproxEye: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency, Radha Venkatagiri (University of Illinois at Urbana-Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana-Champaign), Siva Kumar Sastry Hari (NVIDIA), Sarita V. Adve (University of Illinois at Urbana-Champaign)

Approximate Data Types for Safe and General Low-Power Computation, Adrian Sampson (University of Washington), Werner Dietl (University of Washington), Emily Fortuna (University of Washington), Danushen Gnanapragasam (University of Washington), Luis Ceze (University of Washington), Dan Grossman (University of Washington), Michael Ernst (University of Washington), Todd Anderson (Intel Labs)

Doppelgänger: A Cache for Approximate Computation, Joshua San Miguel (University of Wisconsin-Madison), Jorge Albericio (University of Toronto), Andreas Moshovos (University of Toronto), Natalie Enright Jerger (University of Toronto)

October 19, 2016 (Wednesday)

7:00-8:00 Breakfast

Session 6a: Accelerators 2

An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition, Reza Yazdani Aminabadi (Universitat Politecnica de Catalunya (UPC)), Albert Segura (Universitat Politecnica de Catalunya (UPC)), Jose-Maria Arnau (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC))

Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin, Yakun Sophia Shao (Harvard University), Sam (Likun) Xi (Harvard University), Vijayalakshmi Srinivasan (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains, Arrvindh Shriraman (Simon Fraser University), Sreehaish Kumar (Simon Fraser University), Apala Guha (Simon Fraser University), Arrvindh Shriraman (Simon Fraser University)

Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems, Hadi Asghari-Moghaddam (UIUC), Young Hoon Son (SNU), Jung Ho Ahn (SNU), Nam Sung Kim (UIUC)

8:00-9:40

Session 6b: Mobile & Power Mgmt

A Patch Memory System For Image Processing and Computer Vision, Jason Clemons (NVIDIA), Chih Chi Cheng (NVIDIA), Iuri Frosio (NVIDIA), Daniel Johnson (NVIDIA), Steve Keckler (NVIDIA)

Evaluating Programmable Architectures for Imaging and Vision Applications, Artem Vasilyev (Stanford), Nikhil Bhagdikar (Stanford), Ardavan Pedram (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)

Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users, Kaige Yan (University of Houston), Xingyao Zhang (University of Houston), Jingweija Tan (University of Houston), Xin Fu (University of Houston)

Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks, Dimitrios Skarlatos (UIUC), Reny Thomas (Intel), Aditya Agrawal (Nvidia), Shibin Qin (UIUC), Robert Pilawa-Podgurski (UIUC), Ulya R. Karpuzcu (UMN), Radu Teodorescu (OSU), Nam Sung Kim (UIUC), Josep Torrellas (UIUC)

Ti-states: Processor Power Management in the Temperature Inversion Region, Yazhou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indrani Paul (AMD Research), Vijay Janapa 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, Tae Jun Ham (Princeton University), Lisa Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nadathur Satish (Parallel Computing Lab, Intel Corporation), Margaret Martonosi (Princeton University)

Improving Bank-Level Parallelism for Irregular Applications, Xulong Tang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapalli (Penn State), Jagadish Kotra (Penn State)

Delegated Persist Ordering, Aasheesh Kolli (University of Michigan), Jeff Rosen (Snowflake Computing), Stephan Diestelhorst (ARM), Ali Saidi (ARM), Steven Pelley (Snowflake Computing), Sihang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan)

10:00-12:00

Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emanations, Nader Sehatbakhsh (Georgia Tech), Alireza Nazari (Georgia Tech), Alenka Zajic (Georgia Tech), Milos Prvulovic (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 Wilkerson (Intel Labs), Zeshan Chishti (Intel Labs)

Continuous Runahead: Transparent Hardware Acceleration for Memory Intensive Workloads, Milad Hashemi (UT Austin), Onur Mutlu (CMU), Yale N. Patt (UT Austin)

Session 8: Conference Closing and Best Paper Award

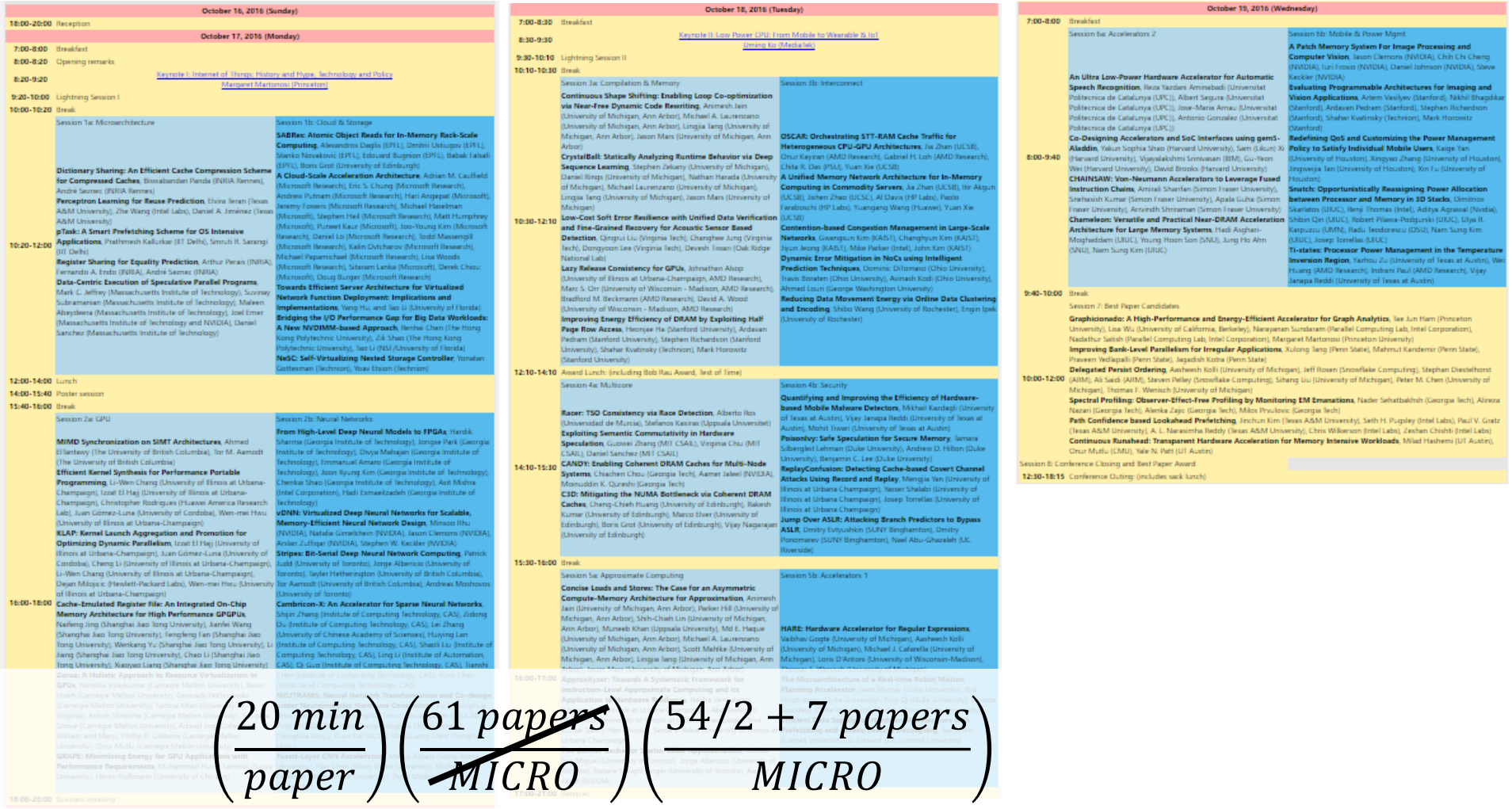
12:30-18:15 Conference Outing: (includes sack lunch)

$$\left(\frac{20\ min}{paper}\right)\left(\frac{61\ papers}{MICRO}\right)$$

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

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

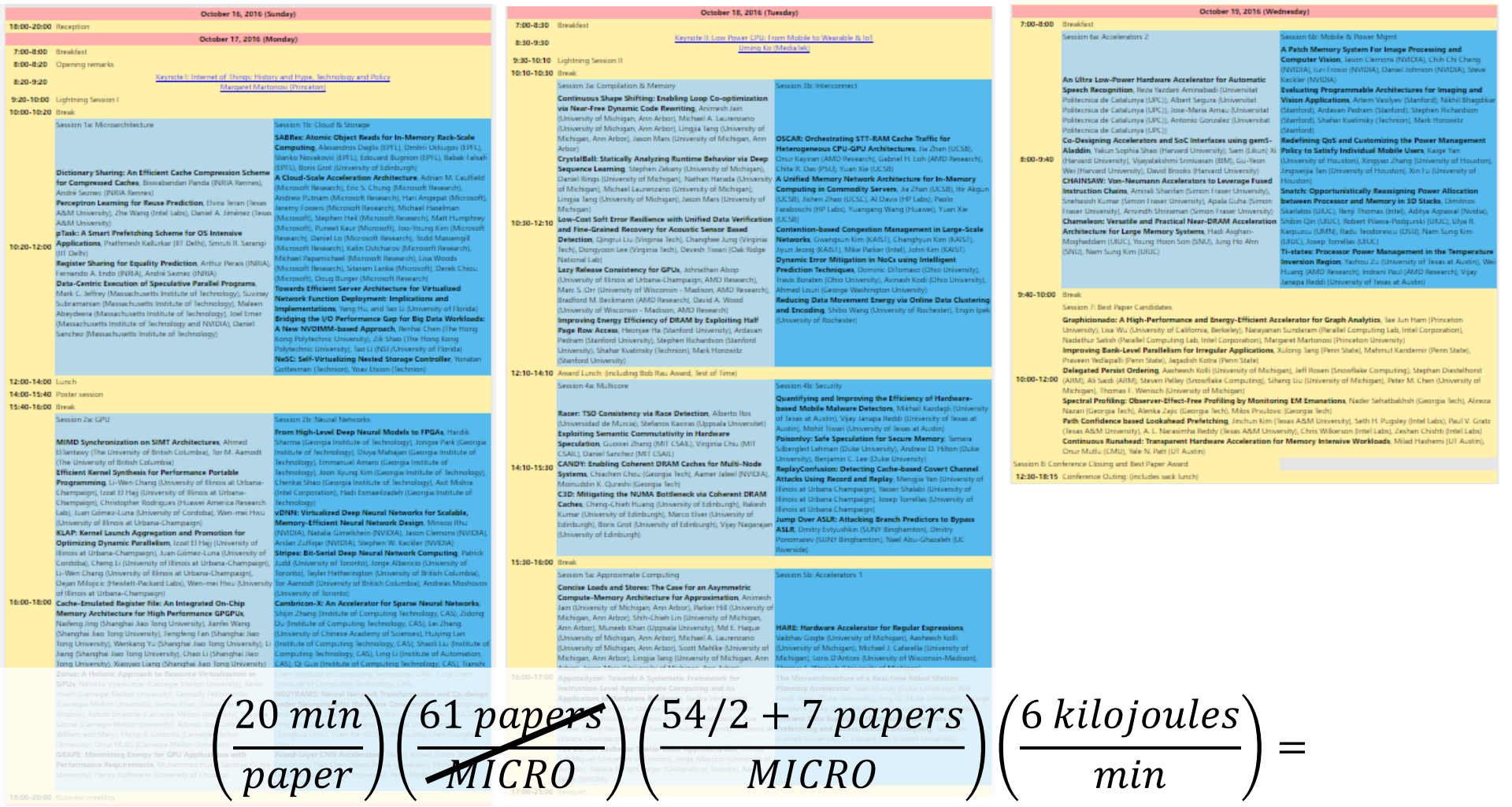


$$\left(\frac{20\ min}{paper}\right)\left(\frac{61\ papers}{MICRO}\right)\left(\frac{54/2+7\ papers}{MICRO}\right)$$

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

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

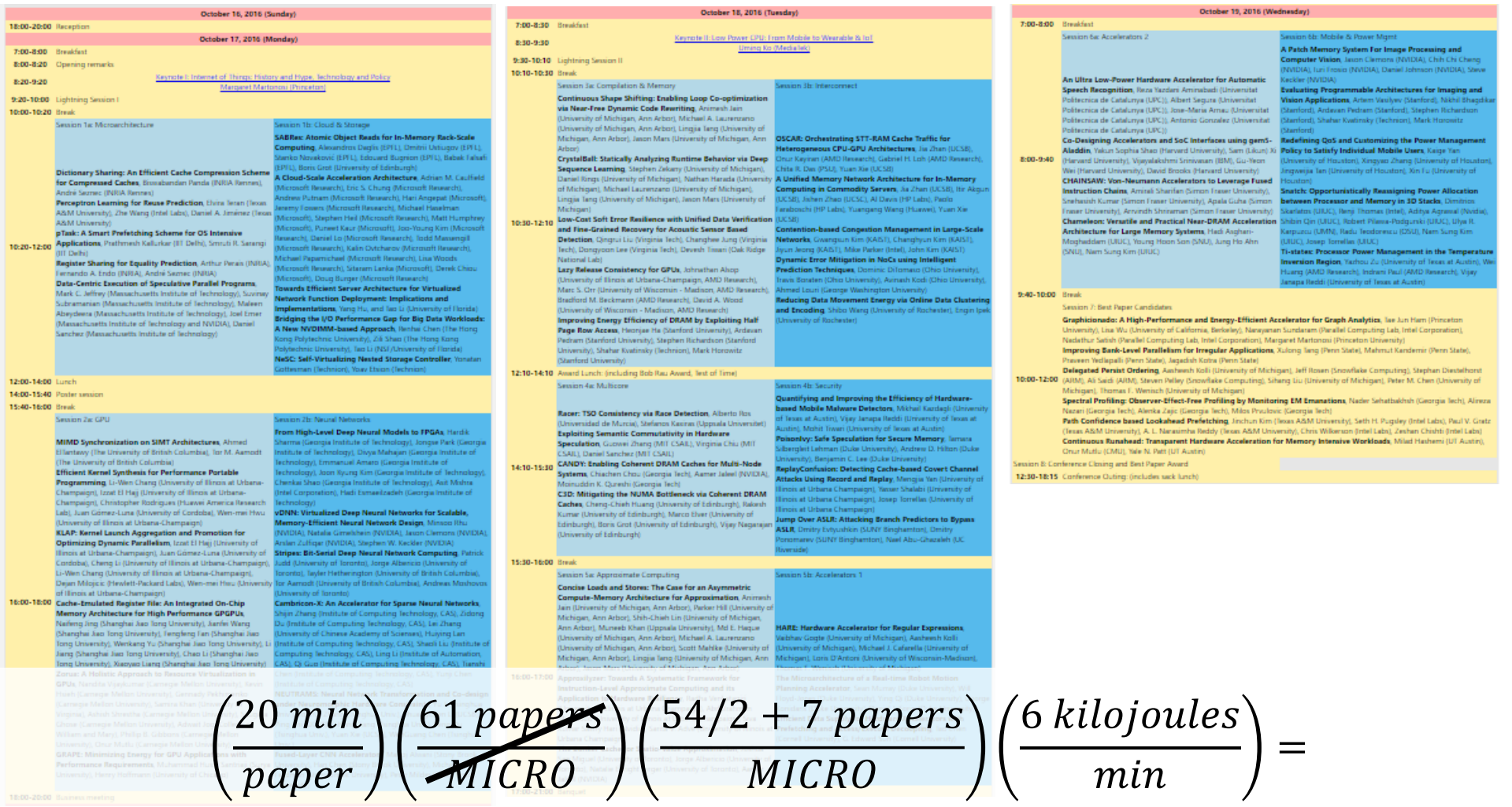


$$\left(\frac{20\ min}{paper}\right)\left(\frac{61\ papers}{\cancel{MICRO}}\right)\left(\frac{54/2+7\ papers}{MICRO}\right)\left(\frac{6\ kilojoules}{min}\right)=$$

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

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



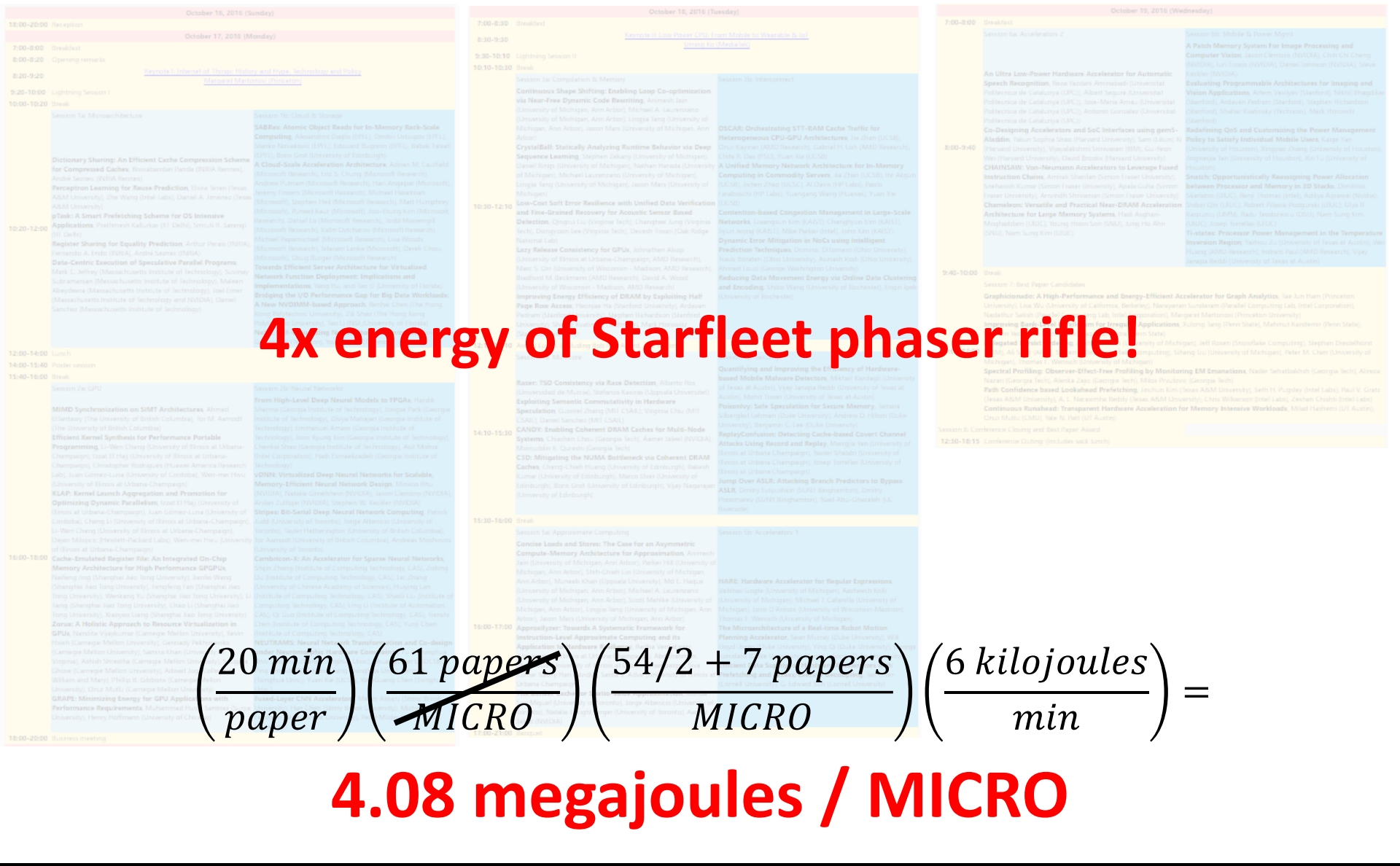
$$\left(\frac{20\ min}{paper}\right)\left(\frac{61\ papers}{\cancel{MICRO}}\right)\left(\frac{54/2+7\ papers}{MICRO}\right)\left(\frac{6\ kilojoules}{min}\right)=$$

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



4x energy of Starfleet phaser rifle!

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

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

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

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: History and Hype, Technology and Policy
Margaret Martonosi (Princeton)

9:20-10:00 Lightning Session I

10:00-10:20 Break

Session 1a: Microarchitecture

Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches, Biswabandan Panda (INRIA Rennes), André Seznec (INRIA Rennes)

Perceptron Learning for Reuse Prediction, Elvira Teran (Texas A&M University), Zhe Wang (Intel Labs), Daniel A. Jiménez (Texas A&M University)

pTask: A Smart Prefetching Scheme for OS Intensive Applications, Prathmesh Kallurkar (IIT Delhi), Smruti R. Sarangi (IIT Delhi)

Register Sharing for Equality Prediction, Arthur Perais (INRIA), Fernando A. Endo (INRIA), André Seznec (INRIA)

Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvinay Subramanian (Massachusetts Institute of Technology), Maleen Abeydeera (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)

Session 1b: Cloud & Storage

SABRes: Atomic Object Reads for In-Memory Rack-Scale Computing, Alexandros Daglis (EPFL), Dmitri Ustiugov (EPFL), Stanko Novaković (EPFL), Edouard Bugnion (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)

A Cloud-Scale Acceleration Architecture, Adrian M. Caulfield (Microsoft Research), Eric S. Chung (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angepat (Microsoft), Jeremy Fowers (Microsoft Research), Michael Haselman (Microsoft), Stephen Heil (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joo-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Todd Massengill (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papamichael (Microsoft Research), Lisa Woods (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chiou (Microsoft), Doug Burger (Microsoft Research)

Towards 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 NVDIMM-based Approach, Renhai Chen (The Hong Kong Polytechnic University), Zili Shao (The Hong Kong Polytechnic University), Tao Li (NSI /University of Florida)

NeSC: Self-Virtualizing Nested Storage Controller, Yonatan Gottesman (Technion), Yoav Etsion (Technion)

10:20-12:00

12:00-14:00 Lunch

14:00-15:40 Poster session

15:40-16:00 Break

Session 2a: GPU

MIMD Synchronization on SIMT Architectures, Ahmed ElTantawy (The University of British Columbia), Tor M. Aamodt (The University of British Columbia)

Efficient Kernel Synthesis for Performance Portable Programming, Li-Wen Chang (University of Illinois at Urbana-Champaign), Izzat El Hajj (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Huawei America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism, Izzat El Hajj (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 Milojicic (Hewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

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), Xiaoyao Liang (Shanghai Jiao Tong University)

Zorua: A Holistic Approach to Resource Virtualization in GPUs, Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gennady Pekhimenko (Carnegie Mellon University), Samira Khan (University of Virginia), Ashish Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Onur Mutlu (Carnegie Mellon University)

GRAPE: Minimizing Energy for GPU Applications with Performance Requirements, Muhammad Husni Santriaji (Surya University), Henry Hoffmann (University of Chicago)

Session 2b: Neural Networks

From High-Level Deep Neural Models to FPGAs, Hardik Sharma (Georgia Institute of Technology), Jongse Park (Georgia Institute of Technology), Divya Mahajan (Georgia Institute of Technology), Emmanuel Amaro (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chenkai Shao (Georgia Institute of Technology), Asit Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

vDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Rhu (NVIDIA), Natalia Gimelshein (NVIDIA), Jason Clemons (NVIDIA), Arslan Zulfiqar (NVIDIA), Stephen W. Keckler (NVIDIA)

Stripes: Bit-Serial Deep Neural Network Computing, Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Tayler Hetherington (University of British Columbia), Tor Aamodt (University of British Columbia), Andreas Moshovos (University of Toronto)

Cambricon-X: An Accelerator for Sparse Neural Networks, Shijin Zhang (Institute of Computing Technology, CAS), Zidong Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huiying Lan (Institute of Computing Technology, CAS), Shaoli Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Tianshi Chen (Institute of Computing Technology, CAS), Yunji Chen (Institute of Computing Technology, CAS)

NEUTRAMS: Neural Network Transformation and Co-design under Neuromorphic Hardware Constraints, Yu Ji (Tsinghua Univ.), Youhui Zhang (Tsinghua Univ.), ShuangChen Li (UCSB), Ping Chi (UCSB), Cihang Jiang (Tsinghua Univ.), Peng Qu (Tsinghua Univ.), Yuan Xie (UCSB), WenGuang Chen (Tsinghua Univ.)

Fused-Layer CNN Accelerators, Manoj Alwani (Stony Brook University), Han Chen (Stony Brook University), Michael Ferdman (Stony Brook University), Peter Milder (Stony Brook University)

16:00-18:00

18:00-20:00 Business meeting

October 18, 2016 (Tuesday)

7:00-8:30 Breakfast

8:30-9:30 Keynote II: Low Power GPU: From Mobile to Wearable & IoT
Umang Ko (Mediatek)

9:30-10:10 Lightning Session II

10:10-10:30 Break

Session 3a: Compilation & Memory

Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting, Animesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning, Stephen Zekany (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjia Tang (University of Michigan), Jason Mars (University of Michigan)

Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection, Qingrui Liu (Virginia Tech), Changhee Jung (Virginia Tech), Dongyoon Lee (Virginia Tech), Devesh Tiwari (Oak Ridge National Lab)

Lazy Release Consistency for GPUs, Johnathan Alsop (University of Illinois at Urbana-Champaign, AMD Research), Marc S. Orr (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, Hoonjae Ha (Stanford University), Ardavan Pedram (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)

Session 3b: Interconnect

OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures, Jia Zhan (UCSB), Onur Kayiran (AMD Research), Gabriel H. Loh (AMD Research), Chita R. Das (PSU), Yuan Xie (UCSB)

A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers, Jia Zhan (UCSB), Itir Akgun (UCSB), Jishen Zhao (UCSC), Al Davis (HP Labs), Paolo Faraboschi (HP Labs), Yuanqiang Wang (Huawei), Yuan Xie (UCSB)

Contention-based Congestion Management in Large-Scale Networks, Gwangsun Kim (KAIST), Changhyun Kim (KAIST), Jiyun Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)

Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques, Dominic DiTomaso (Ohio University), Travis Boraten (Ohio University), Avinash Kodi (Ohio University), Ahmed Louri (George Washington University)

Reducing Data Movement Energy via Online Data Clustering and Encoding, Shibo Wang (University of Rochester), Engin Ipek (University of Rochester)

12:10-14:10 Award Lunch: (including Bob Rau Award, Test of Time)

Session 4a: Multicore

Racer: TSO Consistency via Race Detection, Alberto Ros (Universidad de Murcia), Stefanos Kaxiras (Uppsala Universitet)

Exploiting Semantic Commutativity in Hardware Speculation, Guowei Zhang (MIT CSAIL), Virginia Chiu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)

CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems, Chiachen Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Moinuddin K. Qureshi (Georgia Tech)

C3D: Mitigating the NUMA Bottleneck via Coherent DRAM Caches, Cheng-Chieh Huang (University of Edinburgh), Rakesh Kumar (University of Edinburgh), Marco Elver (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)

Session 4b: Security

Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors, Mikhail Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Mohit Tiwari (University of Texas at Austin)

PoisonIvy: Safe Speculation for Secure Memory, Tamara Silbergleit Lehman (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University)

ReplayConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay, Mengjia Yan (University of Illinois at Urbana Champaign), Yasser Shalabi (University of Illinois at Urbana Champaign), Josep Torrellas (University of Illinois at Urbana Champaign)

Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR, Dmitry Evtyushkin (SUNY Binghamton), Dmitry Ponomarev (SUNY Binghamton), Nael Abu-Ghazaleh (UC Riverside)

14:10-15:30

15:30-16:00 Break

Session 5a: Approximate Computing

Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation, Animesh 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 Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

Approxilyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency, Radha Venkatagiri (University of Illinois at Urbana Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana Champaign), Siva Kumar Sastry Hari (Nvidia), Sarita 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), Natalie Enright Jerger (University of Toronto), Aamer Jaleel (NVIDIA)

Session 5b: Accelerators 1

HARE: Hardware Accelerator for Regular Expressions, Vaibhav Gogte (University of Michigan), Aasheesh Kolli (University of Michigan), Michael J. Cafarella (University of Michigan), Loris D'Antoni (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan)

The Microarchitecture of a Real-time Robot Motion Planning Accelerator, Sean Murray (Duke University), Will Floyd-Jones (Duke University), Ying Qi (Duke University), George Konidaris (Duke University), Daniel J. Sorin (Duke University)

Efficient Data Supply for Hardware Accelerators with Prefetching and Access/Execute Decoupling, Tao Chen (Cornell University), G. Edward Suh (Cornell University)

16:00-17:00

17:00-21:00 Banquet

October 19, 2016 (Wednesday)

7:00-8:00 Breakfast

Session 6a: Accelerators 2

An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition, Reza Yazdani Aminabadi (Universitat Politecnica de Catalunya (UPC)), Albert Segura (Universitat Politecnica de Catalunya (UPC)), Jose-Maria Arnau (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC))

Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin, Yakun Sophia Shao (Harvard University), Sam (Likun) Xi (Harvard University), Vijayalakshmi Srinivasan (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains, Armin Sharifian (Simon Fraser University), Snehasish Kumar (Simon Fraser University), Apala Guha (Simon Fraser University), Arrvindh Shriraman (Simon Fraser University)

Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems, Hadi Asghari-Moghaddam (UIUC), Young Hoon Son (SNU), Jung Ho Ahn (SNU), Nam Sung Kim (UIUC)

Session 6b: Mobile & Power Mgmt

A Patch Memory System For Image Processing and Computer Vision, Jason Clemons (NVIDIA), Chih Chi Cheng (NVIDIA), Iuri Frosio (NVIDIA), Daniel Johnson (NVIDIA), Steve Keckler (NVIDIA)

Evaluating Programmable Architectures for Imaging and Vision Applications, Artem Vasilyev (Stanford), Nikhil Bhagdikar (Stanford), Ardavan Pedram (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)

Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users, Kaige Yan (University of Houston), Xingyao Zhang (University of Houston), Jingweija Tan (University of Houston), Xin Fu (University of Houston)

Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks, Dimitrios Skarlatos (UIUC), Reny Thomas (Intel), Aditya Agrawal (Nvidia), Shibin Qin (UIUC), Robert Pilawa-Podgurski (UIUC), Ulya R. Karpuzcu (UMN), Radu Teodorescu (OSU), Nam Sung Kim (UIUC), Josep Torrellas (UIUC)

Ti-states: Processor Power Management in the Temperature Inversion Region, Yazhou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indrani Paul (AMD Research), Vijay Janapa 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, Tae Jun Ham (Princeton University), Lisa Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nadathur Satish (Parallel Computing Lab, Intel Corporation), Margaret Martonosi (Princeton University)

Improving Bank-Level Parallelism for Irregular Applications, Xulong Tang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapalli (Penn State), Jagadish Kotra (Penn State)

Delegated Persist Ordering, Aasheesh Kolli (University of Michigan), Jeff Rosen (Snowflake Computing), Stephan Diestelhorst (ARM), Ali Saidi (ARM), Steven Pelley (Snowflake Computing), Sihang 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 Emanations, Nader Sehatbakhsh (Georgia Tech), Alireza Nazari (Georgia Tech), Alenka Zajic (Georgia Tech), Milos Prvulovic (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 Wilkerson (Intel Labs), Zeshan Chishti (Intel Labs)

Continuous Runahead: Transparent Hardware Acceleration for Memory Intensive Workloads, Milad Hashemi (UT Austin), Onur Mutlu (CMU), Yale N. Patt (UT Austin)

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 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: History and Hype, Technology and Policy
Margaret Martonosi (Princeton)

9:20-10:00 Lightning Session I

10:00-10:20 Break

Session 1a: Microarchitecture

Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches, Biswabandan Panda (INRIA Rennes), André Seznec (INRIA Rennes)

Perceptron Learning for Reuse Prediction, Elvira Teran (Texas A&M University), Zhe Wang (Intel Labs), Daniel A. Jiménez (Texas A&M University)

pTask: A Smart Prefetching Scheme for OS Intensive Applications, Prathmesh Kallurkar (IIT Delhi), Smruti R. Sarangi (IIT Delhi)

Register Sharing for Equality Prediction, Arthur Perais (INRIA), Fernando A. Endo (INRIA), André Seznec (INRIA)

Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvinay Subramanian (Massachusetts Institute of Technology), Maleen Abeydeera (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)

10:20-12:00

Session 1b: Cloud & Storage

SABRes: Atomic Object Reads for In-Memory Rack-Scale Computing, Alexandros Daglis (EPFL), Dmitri Ustiugov (EPFL), Stanko Novaković (EPFL), Edouard Bugnion (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)

A Cloud-Scale Acceleration Architecture, Adrian M. Caulfield (Microsoft Research), Eric S. Chung (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angepat (Microsoft), Jeremy Fowers (Microsoft Research), Michael Haselman (Microsoft), Stephen Heil (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joo-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Todd Massengill (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papamichael (Microsoft Research), Lisa Woods (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chiou (Microsoft), Doug Burger (Microsoft Research)

Towards 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 NVDIMM-based Approach, Renhai Chen (The Hong Kong Polytechnic University), Zili Shao (The Hong Kong Polytechnic University), Tao Li (NSI /University of Florida)

NeSC: Self-Virtualizing Nested Storage Controller, Yonatan Gottesman (Technion), Yoav Etsion (Technion)

12:00-14:00 Lunch

14:00-15:40 Poster session

15:40-16:00 Break

Session 2a: GPU

MIMD Synchronization on SIMT Architectures, Ahmed El-Shafiey (The University of British Columbia), Tor M. Aamodt (The University of British Columbia)

Efficient Kernel Synthesis for Performance Portable Programming, Li-Wen Chang (University of Illinois at Urbana-Champaign), Izzat El Hajj (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Huawei America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism, Izzat El Hajj (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 Milojicic (Hewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

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), Lingjun Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyao Liang (Shanghai Jiao Tong University)

Zorua: A Holistic Approach to Resource Virtualization in GPUs, Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gennady Pekhimenko (Carnegie Mellon University), Samira Khan (University of Virginia), Ashish Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Onur Mutlu (Carnegie Mellon University)

GRAPE: Minimizing Energy for GPU Applications with Performance Requirements, Muhammad Husni Santriaji (Sultan Agung Islamic University), Henry Hoffmann (University of Chicago)

Session 2b: Neural Networks

From High-Level Deep Neural Models to FPGAs, Hardik Sharma (Georgia Institute of Technology), Jongse Park (Georgia Institute of Technology), Divya Mahajan (Georgia Institute of Technology), Emmanuel Amaro (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chenkai Shao (Georgia Institute of Technology), Asit Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

vDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Rhu (NVIDIA), Natalia Gimelshein (NVIDIA), Jason Clemons (NVIDIA), Arslan Zulfiqar (NVIDIA), Stephen W. Keckler (NVIDIA)

Stripes: Bit-Serial Deep Neural Network Computing, Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Tayler Hetherington (University of British Columbia), Tor Aamodt (University of British Columbia), Andreas Moshovos (University of Toronto)

Cambricon-X: An Accelerator for Sparse Neural Networks, Shijin Zhang (Institute of Computing Technology, CAS), Zidong Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huiying Lan (Institute of Computing Technology, CAS), Shaoli Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Tianshi Chen (Institute of Computing Technology, CAS), Yunji Chen (Institute of Computing Technology, CAS)

NEUTRAMS: Neural Network Transformation and Co-design under Neuromorphic Hardware Constraints, Yu Ji (Tsinghua Univ.), Youhui Zhang (Tsinghua Univ.), ShuangChen Li (UCSB), Ping Chi (UCSB), Cihang Jiang (Tsinghua Univ.), Peng Qu (Tsinghua Univ.), Yuan Xie (UCSB), WenGuang Chen (Tsinghua Univ.)

Fused-Layer CNN Accelerators, Manoj Alwani (Stony Brook University), Han Chen (Stony Brook University), Michael Ferdman (Stony Brook University), Peter Milder (Stony Brook University)

18:00-20:00 Business meeting

October 18, 2016 (Tuesday)

7:00-8:30 Breakfast

8:30-9:30 Keynote II: Low Power GPU: From Mobile to Wearable & IoT
Ulrang Ko (Mediatek)

9:30-10:10 Lightning Session II

10:10-10:30 Break

Session 3a: Compilation & Memory

Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting, Animesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning, Stephen Zekany (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjia Tang (University of Michigan), Jason Mars (University of Michigan)

Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection, Qingrui Liu (Virginia Tech), Changhee Jung (Virginia Tech), Dongyoon Lee (Virginia Tech), Devesh Tiwari (Oak Ridge National Lab)

Lazy Release Consistency for GPUs, Johnathan Alsop (University of Illinois at Urbana-Champaign, AMD Research), Marc S. Orr (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, Hoonjae Ha (Stanford University), Ardavan Pedram (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)

Session 3b: Interconnect

OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures, Jia Zhan (UCSB), Onur Kayiran (AMD Research), Gabriel H. Loh (AMD Research), Chita R. Das (PSU), Yuan Xie (UCSB)

A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers, Jia Zhan (UCSB), Itir Akgun (UCSB), Jishen Zhao (UCSC), Al Davis (HP Labs), Paolo Faraboschi (HP Labs), Yuanqiang Wang (Huawei), Yuan Xie (UCSB)

Contention-based Congestion Management in Large-Scale Networks, Gwangsun Kim (KAIST), Changhyun Kim (KAIST), Jiyun Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)

Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques, Dominic DiTomaso (Ohio University), Travis Boraten (Ohio University), Avinash Kodi (Ohio University), Ahmed Louri (George Washington University)

Reducing Data Movement Energy via Online Data Clustering and Encoding, Shibo Wang (University of Rochester), Engin Ipek (University of Rochester)

12:10-14:10 Award Lunch: (including Bob Rau Award, Test of Time)

Session 4a: Multicore

Racer: TSO Consistency via Race Detection, Alberto Ros (Universidad de Murcia), Stefanos Kaxiras (Uppsala Universitet)

Exploiting Semantic Commutativity in Hardware Speculation, Guowei Zhang (MIT CSAIL), Virginia Chiu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)

CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems, Chiachen Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Moinuddin K. Qureshi (Georgia Tech)

CSI: Mitigating the NUMA Bottleneck via Coherent DRAM Caches, Cheng-Chieh Huang (University of Edinburgh), Rakesh Kumar (University of Edinburgh), Marco Elver (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)

Session 4b: Security

Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors, Mikhail Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Mohit Tiwari (University of Texas at Austin)

PoisonIvy: Safe Speculation for Secure Memory, Tamara Silbergleit Lehman (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University)

ReplayConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay, Mengjia Yan (University of Illinois at Urbana Champaign), Yasser Shalabi (University of Illinois at Urbana Champaign), Josep Torrellas (University of Illinois at Urbana Champaign)

Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR, Dmitry Evtyushkin (SUNY Binghamton), Dmitry Ponomarev (SUNY Binghamton), Nael Abu-Ghazaleh (UC Riverside)

14:10-15:30

15:30-16:00 Break

Session 5a: Approximate Computing

Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation, Animesh 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 Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

Approxilyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency, Radha Venkatagiri (University of Illinois at Urbana Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana Champaign), Siva Kumar Sastry Hari (Nvidia), Sarita 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), Natalie Enright Jerger (University of Toronto), Aamer Jaleel (NVIDIA)

Session 5b: Accelerators 1

HARE: Hardware Accelerator for Regular Expressions, Vaibhav Gogte (University of Michigan), Aasheesh Kolli (University of Michigan), Michael J. Cafarella (University of Michigan), Loris D'Antoni (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan)

The Microarchitecture of a Real-time Robot Motion Planning Accelerator, Sean Murray (Duke University), Will Floyd-Jones (Duke University), Ying Qi (Duke University), George Konidaris (Duke University), Daniel J. Sorin (Duke University)

Efficient Data Supply for Hardware Accelerators with Prefetching and Access/Execute Decoupling, Tao Chen (Cornell University), G. Edward Suh (Cornell University)

16:00-17:00

17:00-21:00 Banquet

October 19, 2016 (Wednesday)

7:00-8:00 Breakfast

Session 6a: Accelerators 2

An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition, Reza Yazdani Aminabadi (Universitat Politecnica de Catalunya (UPC)), Albert Segura (Universitat Politecnica de Catalunya (UPC)), Jose-Maria Arnau (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC))

Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin, Yakun Sophia Shao (Harvard University), Sam (Likun) Xi (Harvard University), Vijayalakshmi Srinivasan (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains, Armin Sharifian (Simon Fraser University), Snehasish Kumar (Simon Fraser University), Apala Guha (Simon Fraser University), Arrvindh Shriraman (Simon Fraser University)

Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems, Hadi Asghari-Moghaddam (UIUC), Young Hoon Son (SNU), Jung Ho Ahn (SNU), Nam Sung Kim (UIUC)

8:00-9:40

Session 6b: Mobile & Power Mgmt

A Patch Memory System For Image Processing and Computer Vision, Jason Clemons (NVIDIA), Chih Chi Cheng (NVIDIA), Iuri Frosio (NVIDIA), Daniel Johnson (NVIDIA), Steve Keckler (NVIDIA)

Evaluating Programmable Architectures for Imaging and Vision Applications, Artem Vasilyev (Stanford), Nikhil Bhagdikar (Stanford), Ardavan Pedram (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)

Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users, Kaige Yan (University of Houston), Xingyao Zhang (University of Houston), Jingweija Tan (University of Houston), Xin Fu (University of Houston)

Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks, Dimitrios Skarlatos (UIUC), Reny Thomas (Intel), Aditya Agrawal (Nvidia), Shibin Qin (UIUC), Robert Pilawa-Podgurski (UIUC), Ulya R. Karpuzcu (UMN), Radu Teodorescu (OSU), Nam Sung Kim (UIUC), Josep Torrellas (UIUC)

TI-states: Processor Power Management in the Temperature Inversion Region, Yazhou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indrani Paul (AMD Research), Vijay Janapa 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, Tae Jun Ham (Princeton University), Lisa Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nadathur Satish (Parallel Computing Lab, Intel Corporation), Margaret Martonosi (Princeton University)

Improving Bank-Level Parallelism for Irregular Applications, Xulong Tang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapalli (Penn State), Jagadish Kotra (Penn State)

Delegated Persist Ordering, Aasheesh Kolli (University of Michigan), Jeff Rosen (Snowflake Computing), Stephan Diestelhorst (ARM), Ali Saidi (ARM), Steven Pelley (Snowflake Computing), Sihang 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 Emanations, Nader Sehatbakhsh (Georgia Tech), Alireza Nazari (Georgia Tech), Alenka Zajic (Georgia Tech), Milos Prvulovic (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 Wilkerson (Intel Labs), Zeshan Chishti (Intel Labs)

Continuous Runahead: Transparent Hardware Acceleration for Memory Intensive Workloads, Milad Hashemi (UT Austin), Onur Mutlu (CMU), Yale N. Patt (UT Austin)

Session 8: Conference Closing and Best Paper Award

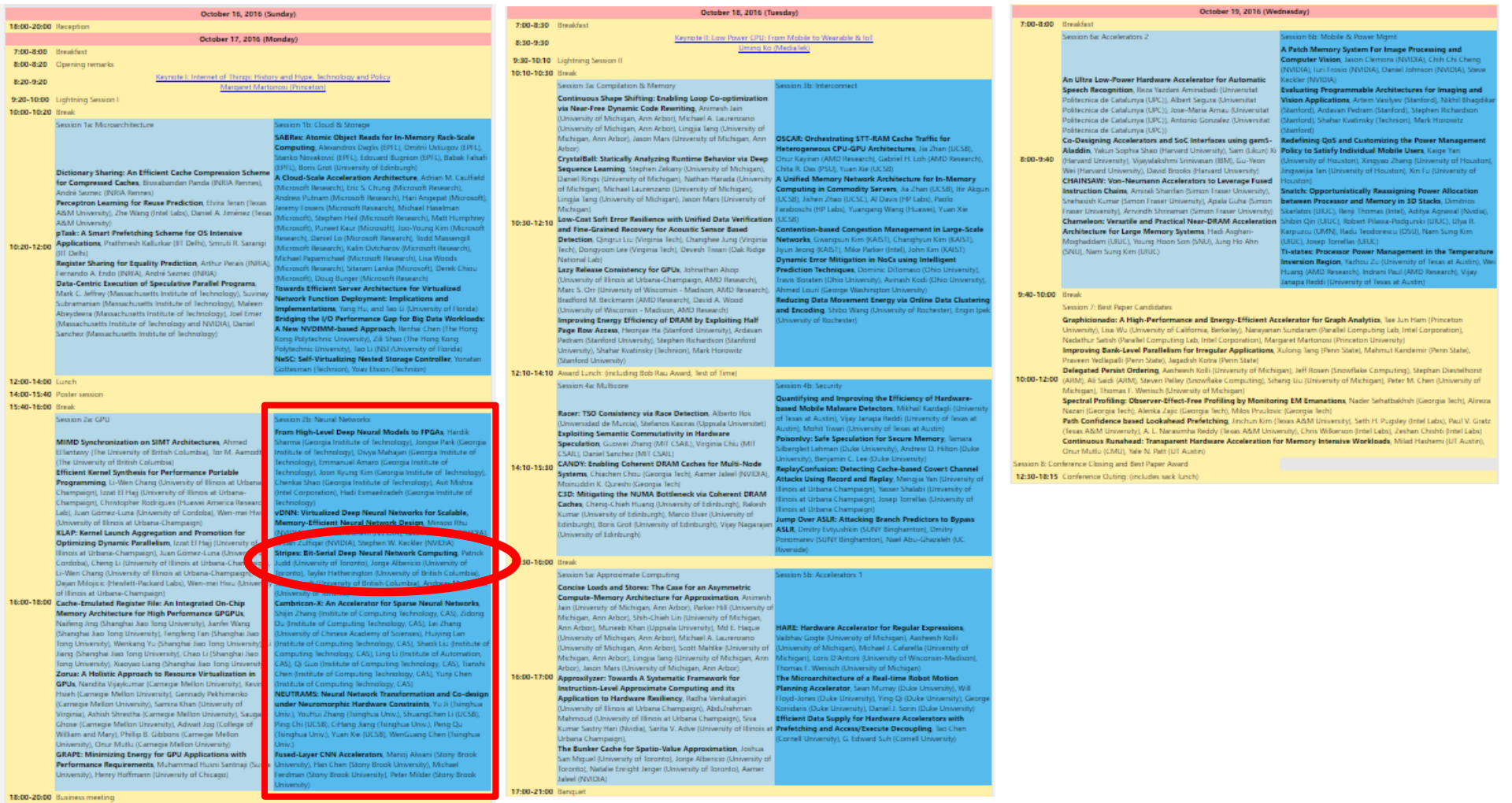
12:30-18:15 Conference Outing: (includes sack lunch)

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

Joshua San Miguel, Jorge 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: History and Hype, Technology and Policy
Margaret Martonosi (Princeton)

9:20-10:00 Lightning Session I

10:00-10:20 Break

Session 1a: Microarchitecture

Dictionary Sharing: An Efficient Cache Compression Scheme for Compressed Caches, Biswabandan Panda (INRIA Rennes), André Seznec (INRIA Rennes)

Perceptron Learning for Reuse Prediction, Elvira Teran (Texas A&M University), Zhe Wang (Intel Labs), Daniel A. Jiménez (Texas A&M University)

pTask: A Smart Prefetching Scheme for OS Intensive Applications, Prathmesh Kallurkar (IIT Delhi), Smruti R. Sarangi (IIT Delhi)

Register Sharing for Equality Prediction, Arthur Perais (INRIA), Fernando A. Endo (INRIA), André Seznec (INRIA)

Data-Centric Execution of Speculative Parallel Programs, Mark C. Jeffrey (Massachusetts Institute of Technology), Suvinay Subramanian (Massachusetts Institute of Technology), Maleen Abeydeera (Massachusetts Institute of Technology), Joel Emer (Massachusetts Institute of Technology and NVIDIA), Daniel Sanchez (Massachusetts Institute of Technology)

10:20-12:00

Session 1b: Cloud & Storage

SABRes: Atomic Object Reads for In-Memory Rack-Scale Computing, Alexandros Daglis (EPFL), Dmitri Ustiugov (EPFL), Stanko Novaković (EPFL), Edouard Bugnion (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)

A Cloud-Scale Acceleration Architecture, Adrian M. Caulfield (Microsoft Research), Eric S. Chung (Microsoft Research), Andrew Putnam (Microsoft Research), Hari Angepat (Microsoft), Jeremy Fowers (Microsoft Research), Michael Haselman (Microsoft), Stephen Heil (Microsoft Research), Matt Humphrey (Microsoft), Puneet Kaur (Microsoft), Joo-Young Kim (Microsoft Research), Daniel Lo (Microsoft Research), Todd Massengill (Microsoft Research), Kalin Ovtcharov (Microsoft Research), Michael Papamichael (Microsoft Research), Lisa Woods (Microsoft Research), Sitaram Lanka (Microsoft), Derek Chiou (Microsoft), Doug Burger (Microsoft Research)

Towards 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 NVDIMM-based Approach, Renhai Chen (The Hong Kong Polytechnic University), Zili Shao (The Hong Kong Polytechnic University), Tao Li (NSI /University of Florida)

NeSC: Self-Virtualizing Nested Storage Controller, Yonatan Gottesman (Technion), Yoav Etsion (Technion)

12:00-14:00 Lunch

14:00-15:40 Poster session

15:40-16:00 Break

Session 2a: GPU

MIMD Synchronization on SIMT Architectures, Ahmed El Hajjaj (The University of British Columbia), Tor M. Aamodt (The University of British Columbia)

Efficient Kernel Synthesis for Performance Portable Programming, Li-Wen Chang (University of Illinois at Urbana-Champaign), Izzat El Hajj (University of Illinois at Urbana-Champaign), Christopher Rodrigues (Huawei America Research Lab), Juan Gómez-Luna (University of Cordoba), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

KLAP: Kernel Launch Aggregation and Promotion for Optimizing Dynamic Parallelism, Izzat El Hajj (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 Milojicic (Hewlett-Packard Labs), Wen-mei Hwu (University of Illinois at Urbana-Champaign)

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), Lingjun Jiang (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Xiaoyao Liang (Shanghai Jiao Tong University)

Zorua: A Holistic Approach to Resource Virtualization in GPUs, Nandita Vijaykumar (Carnegie Mellon University), Kevin Hsieh (Carnegie Mellon University), Gennady Pekhimenko (Carnegie Mellon University), Samira Khan (University of Virginia), Ashish Shrestha (Carnegie Mellon University), Saugata Ghose (Carnegie Mellon University), Adwait Jog (College of William and Mary), Phillip B. Gibbons (Carnegie Mellon University), Onur Mutlu (Carnegie Mellon University)

GRAPE: Minimizing Energy for GPU Applications with Performance Requirements, Muhammad Husni Santriaji (Sultan Qaboos University), Henry Hoffmann (University of Chicago)

16:00-18:00

Session 2b: Neural Networks

From High-Level Deep Neural Models to FPGAs, Hardik Sharma (Georgia Institute of Technology), Jongse Park (Georgia Institute of Technology), Divya Mahajan (Georgia Institute of Technology), Emmanuel Amaro (Georgia Institute of Technology), Joon Kyung Kim (Georgia Institute of Technology), Chenkai Shao (Georgia Institute of Technology), Asit Mishra (Intel Corporation), Hadi Esmaeilzadeh (Georgia Institute of Technology)

vDNN: Virtualized Deep Neural Networks for Scalable, Memory-Efficient Neural Network Design, Minsoo Rhu (NVIDIA), Natalia Gimelshein (NVIDIA), Jason Clemons (NVIDIA), Arslan Zulfiqar (NVIDIA), Stephen W. Keckler (NVIDIA)

Stripes: Bit-Serial Deep Neural Network Computing, Patrick Judd (University of Toronto), Jorge Albericio (University of Toronto), Tayler Hetherington (University of British Columbia), Andreas Moshovos (University of British Columbia), Natalie Enright Jerger (University of Toronto)

Cambricon-X: An Accelerator for Sparse Neural Networks, Shijin Zhang (Institute of Computing Technology, CAS), Zidong Du (Institute of Computing Technology, CAS), Lei Zhang (University of Chinese Academy of Sciences), Huiying Lan (Institute of Computing Technology, CAS), Shaoli Liu (Institute of Computing Technology, CAS), Ling Li (Institute of Automation, CAS), Qi Guo (Institute of Computing Technology, CAS), Tianshi Chen (Institute of Computing Technology, CAS), Yunji Chen (Institute of Computing Technology, CAS)

NEUTRAMS: Neural Network Transformation and Co-design under Neuromorphic Hardware Constraints, Yu Ji (Tsinghua Univ.), Youhui Zhang (Tsinghua Univ.), ShuangChen Li (UCSB), Ping Chi (UCSB), Cihang Jiang (Tsinghua Univ.), Peng Qu (Tsinghua Univ.), Yuan Xie (UCSB), WenGuang Chen (Tsinghua Univ.)

Fused-Layer CNN Accelerators, Manoj Alwani (Stony Brook University), Han Chen (Stony Brook University), Michael Ferdman (Stony Brook University), Peter Milder (Stony Brook University)

18:00-20:00 Business meeting

October 18, 2016 (Tuesday)

7:00-8:30 Breakfast

8:30-9:30 Keynote II: Low Power GPU: From Mobile to Wearable & IoT
Ulrang Ko (Mediatek)

9:30-10:10 Lightning Session II

10:10-10:30 Break

Session 3a: Compilation & Memory

Continuous Shape Shifting: Enabling Loop Co-optimization via Near-Free Dynamic Code Rewriting, Animesh Jain (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

CrystalBall: Statically Analyzing Runtime Behavior via Deep Sequence Learning, Stephen Zekany (University of Michigan), Daniel Rings (University of Michigan), Nathan Harada (University of Michigan), Michael Laurenzano (University of Michigan), Lingjia Tang (University of Michigan), Jason Mars (University of Michigan)

Low-Cost Soft Error Resilience with Unified Data Verification and Fine-Grained Recovery for Acoustic Sensor Based Detection, Qingrui Liu (Virginia Tech), Changhee Jung (Virginia Tech), Dongyoon Lee (Virginia Tech), Devesh Tiwari (Oak Ridge National Lab)

Lazy Release Consistency for GPUs, Johnathan Alsop (University of Illinois at Urbana-Champaign, AMD Research), Marc S. Orr (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, Hoonjae Ha (Stanford University), Ardavan Pedram (Stanford University), Stephen Richardson (Stanford University), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford University)

10:30-12:10

Session 3b: Interconnect

OSCAR: Orchestrating STT-RAM Cache Traffic for Heterogeneous CPU-GPU Architectures, Jia Zhan (UCSB), Onur Kayiran (AMD Research), Gabriel H. Loh (AMD Research), Chita R. Das (PSU), Yuan Xie (UCSB)

A Unified Memory Network Architecture for In-Memory Computing in Commodity Servers, Jia Zhan (UCSB), Itir Akgun (UCSB), Jishen Zhao (UCSC), Al Davis (HP Labs), Paolo Faraboschi (HP Labs), Yuanqiang Wang (Huawei), Yuan Xie (UCSB)

Contention-based Congestion Management in Large-Scale Networks, Gwangsun Kim (KAIST), Changhyun Kim (KAIST), Jiyun Jeong (KAIST), Mike Parker (Intel), John Kim (KAIST)

Dynamic Error Mitigation in NoCs using Intelligent Prediction Techniques, Dominic DiTomaso (Ohio University), Travis Boraten (Ohio University), Avinash Kodi (Ohio University), Ahmed Louri (George Washington University)

Reducing Data Movement Energy via Online Data Clustering and Encoding, Shibo Wang (University of Rochester), Engin Ipek (University of Rochester)

12:10-14:10 Award Lunch: (including Bob Rau Award, Test of Time)

Session 4a: Multicore

Racer: TSO Consistency via Race Detection, Alberto Ros (Universidad de Murcia), Stefanos Kaxiras (Uppsala Universitet)

Exploiting Semantic Commutativity in Hardware Speculation, Guowei Zhang (MIT CSAIL), Virginia Chiu (MIT CSAIL), Daniel Sanchez (MIT CSAIL)

CANDY: Enabling Coherent DRAM Caches for Multi-Node Systems, Chiachen Chou (Georgia Tech), Aamer Jaleel (NVIDIA), Moinuddin K. Qureshi (Georgia Tech)

CSI: Mitigating the NUMA Bottleneck via Coherent DRAM Caches, Cheng-Chieh Huang (University of Edinburgh), Rakesh Kumar (University of Edinburgh), Marco Elver (University of Edinburgh), Boris Grot (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)

14:10-15:30

Session 4b: Security

Quantifying and Improving the Efficiency of Hardware-based Mobile Malware Detectors, Mikhail Kazdagli (University of Texas at Austin), Vijay Janapa Reddi (University of Texas at Austin), Mohit Tiwari (University of Texas at Austin)

PoisonIvy: Safe Speculation for Secure Memory, Tamara Silbergleit Lehman (Duke University), Andrew D. Hilton (Duke University), Benjamin C. Lee (Duke University)

ReplayConfusion: Detecting Cache-based Covert Channel Attacks Using Record and Replay, Mengjia Yan (University of Illinois at Urbana Champaign), Yasser Shalabi (University of Illinois at Urbana Champaign), Josep Torrellas (University of Illinois at Urbana Champaign)

Jump Over ASLR: Attacking Branch Predictors to Bypass ASLR, Dmitry Evtyushkin (SUNY Binghamton), Dmitry Ponomarev (SUNY Binghamton), Nael Abu-Ghazaleh (UC Riverside)

15:30-16:00 Break

Session 5a: Approximate Computing

Concise Loads and Stores: The Case for an Asymmetric Compute-Memory Architecture for Approximation, Animesh 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 Tang (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor)

Approxilyzer: Towards A Systematic Framework for Instruction-Level Approximate Computing and its Application to Hardware Resiliency, Radha Venkatagiri (University of Illinois at Urbana Champaign), Abdulrahman Mahmoud (University of Illinois at Urbana Champaign), Siva Kumar Sastry Hari (Nvidia), Sarita 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), Natalie Enright Jerger (University of Toronto), Aamer Jaleel (NVIDIA)

16:00-17:00

Session 5b: Accelerators 1

HARE: Hardware Accelerator for Regular Expressions, Vaibhav Gogte (University of Michigan), Aasheesh Kolli (University of Michigan), Michael J. Cafarella (University of Michigan), Loris D'Antoni (University of Wisconsin-Madison), Thomas F. Wenisch (University of Michigan)

The Microarchitecture of a Real-time Robot Motion Planning Accelerator, Sean Murray (Duke University), Will Floyd-Jones (Duke University), Ying Qi (Duke University), George Konidaris (Duke University), Daniel J. Sorin (Duke University)

Efficient Data Supply for Hardware Accelerators with Prefetching and Access/Execute Decoupling, Tao Chen (Cornell University), G. Edward Suh (Cornell University)

17:00-21:00 Banquet

October 19, 2016 (Wednesday)

7:00-8:00 Breakfast

Session 6a: Accelerators 2

An Ultra Low-Power Hardware Accelerator for Automatic Speech Recognition, Reza Yazdani Aminabadi (Universitat Politecnica de Catalunya (UPC)), Albert Segura (Universitat Politecnica de Catalunya (UPC)), Jose-Maria Arnau (Universitat Politecnica de Catalunya (UPC)), Antonio Gonzalez (Universitat Politecnica de Catalunya (UPC))

Co-Designing Accelerators and SoC Interfaces using gem5-Aladdin, Yakun Sophia Shao (Harvard University), Sam (Likun) Xi (Harvard University), Vijayalakshmi Srinivasan (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

CHAINSAW: Von-Neumann Accelerators to Leverage Fused Instruction Chains, Armin Sharifian (Simon Fraser University), Snehasish Kumar (Simon Fraser University), Apala Guha (Simon Fraser University), Arrvindh Shriraman (Simon Fraser University)

Chameleon: Versatile and Practical Near-DRAM Acceleration Architecture for Large Memory Systems, Hadi Asghari-Moghaddam (UIUC), Young Hoon Son (SNU), Jung Ho Ahn (SNU), Nam Sung Kim (UIUC)

8:00-9:40

Session 6b: Mobile & Power Mgmt

A Patch Memory System For Image Processing and Computer Vision, Jason Clemons (NVIDIA), Chih Chi Cheng (NVIDIA), Iuri Frosio (NVIDIA), Daniel Johnson (NVIDIA), Steve Keckler (NVIDIA)

Evaluating Programmable Architectures for Imaging and Vision Applications, Artem Vasilyev (Stanford), Nikhil Bhagdikar (Stanford), Ardavan Pedram (Stanford), Stephen Richardson (Stanford), Shahar Kvatinsky (Technion), Mark Horowitz (Stanford)

Redefining QoS and Customizing the Power Management Policy to Satisfy Individual Mobile Users, Kaige Yan (University of Houston), Xingyao Zhang (University of Houston), Jingweija Tan (University of Houston), Xin Fu (University of Houston)

Snatch: Opportunistically Reassigning Power Allocation between Processor and Memory in 3D Stacks, Dimitrios Skarlatos (UIUC), Reny Thomas (Intel), Aditya Agrawal (Nvidia), Shibin Qin (UIUC), Robert Pilawa-Podgurski (UIUC), Ulya R. Karpuzcu (UMN), Radu Teodorescu (OSU), Nam Sung Kim (UIUC), Josep Torrellas (UIUC)

TI-states: Processor Power Management in the Temperature Inversion Region, Yazhou Zu (University of Texas at Austin), Wei Huang (AMD Research), Indrani Paul (AMD Research), Vijay Janapa 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, Tae Jun Ham (Princeton University), Lisa Wu (University of California, Berkeley), Narayanan Sundaram (Parallel Computing Lab, Intel Corporation), Nadathur Satish (Parallel Computing Lab, Intel Corporation), Margaret Martonosi (Princeton University)

Improving Bank-Level Parallelism for Irregular Applications, Xulong Tang (Penn State), Mahmut Kandemir (Penn State), Praveen Yedlapalli (Penn State), Jagadish Kotra (Penn State)

Delegated Persist Ordering, Aasheesh Kolli (University of Michigan), Jeff Rosen (Snowflake Computing), Stephan Diestelhorst (ARM), Ali Saidi (ARM), Steven Pelley (Snowflake Computing), Sihang Liu (University of Michigan), Peter M. Chen (University of Michigan), Thomas F. Wenisch (University of Michigan)

10:00-12:00

Spectral Profiling: Observer-Effect-Free Profiling by Monitoring EM Emanations, Nader Sehatbakhsh (Georgia Tech), Alireza Nazari (Georgia Tech), Alenka Zajic (Georgia Tech), Milos Prvulovic (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 Wilkerson (Intel Labs), Zeshan Chishti (Intel Labs)

Continuous Runahead: Transparent Hardware Acceleration for Memory Intensive Workloads, Milad Hashemi (UT Austin), Onur Mutlu (CMU), Yale N. Patt (UT Austin)

Session 8: Conference Closing and Best Paper Award

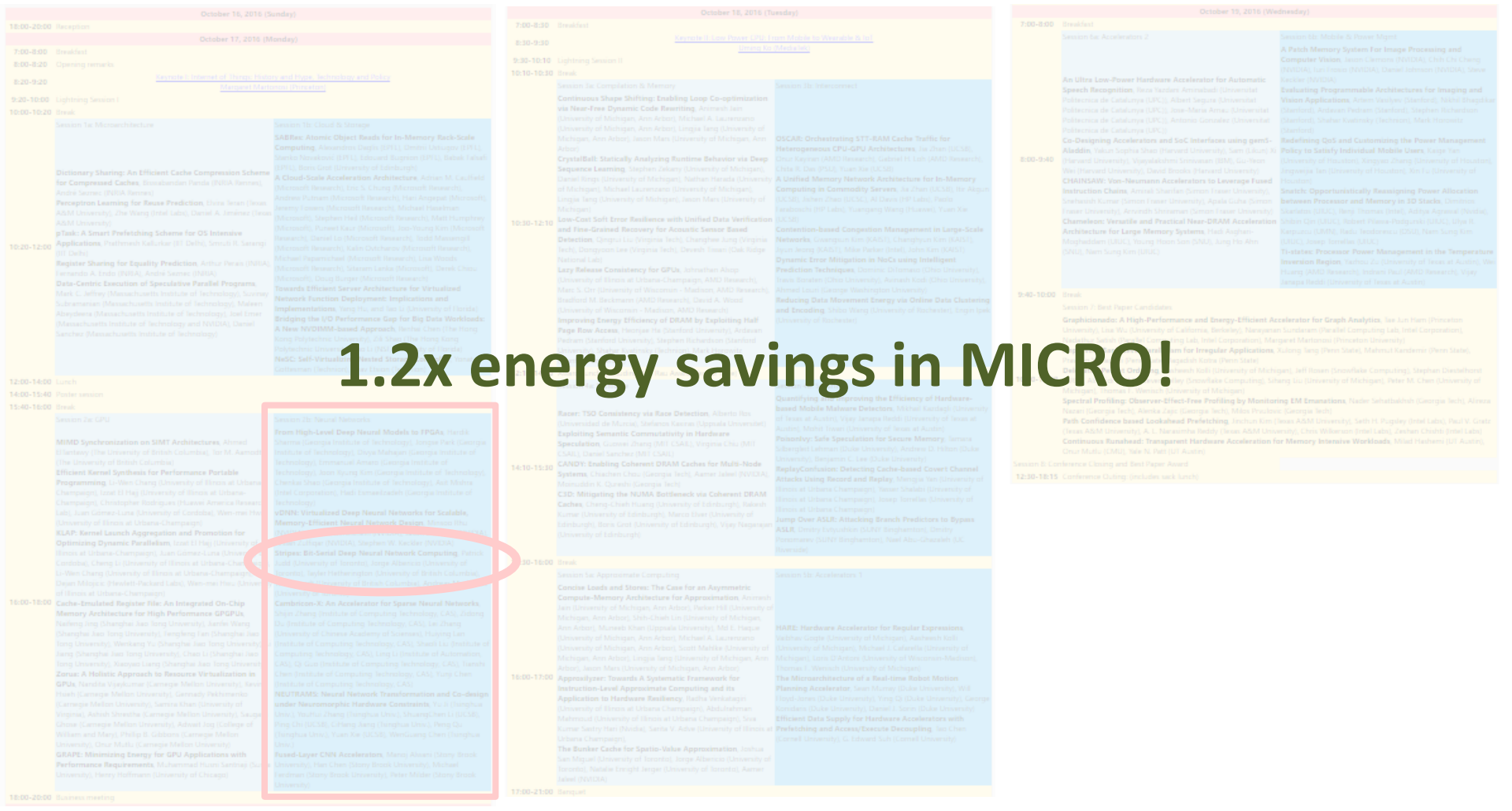
12:30-18:15 Conference Outing: (includes sack lunch)

4.08 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

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



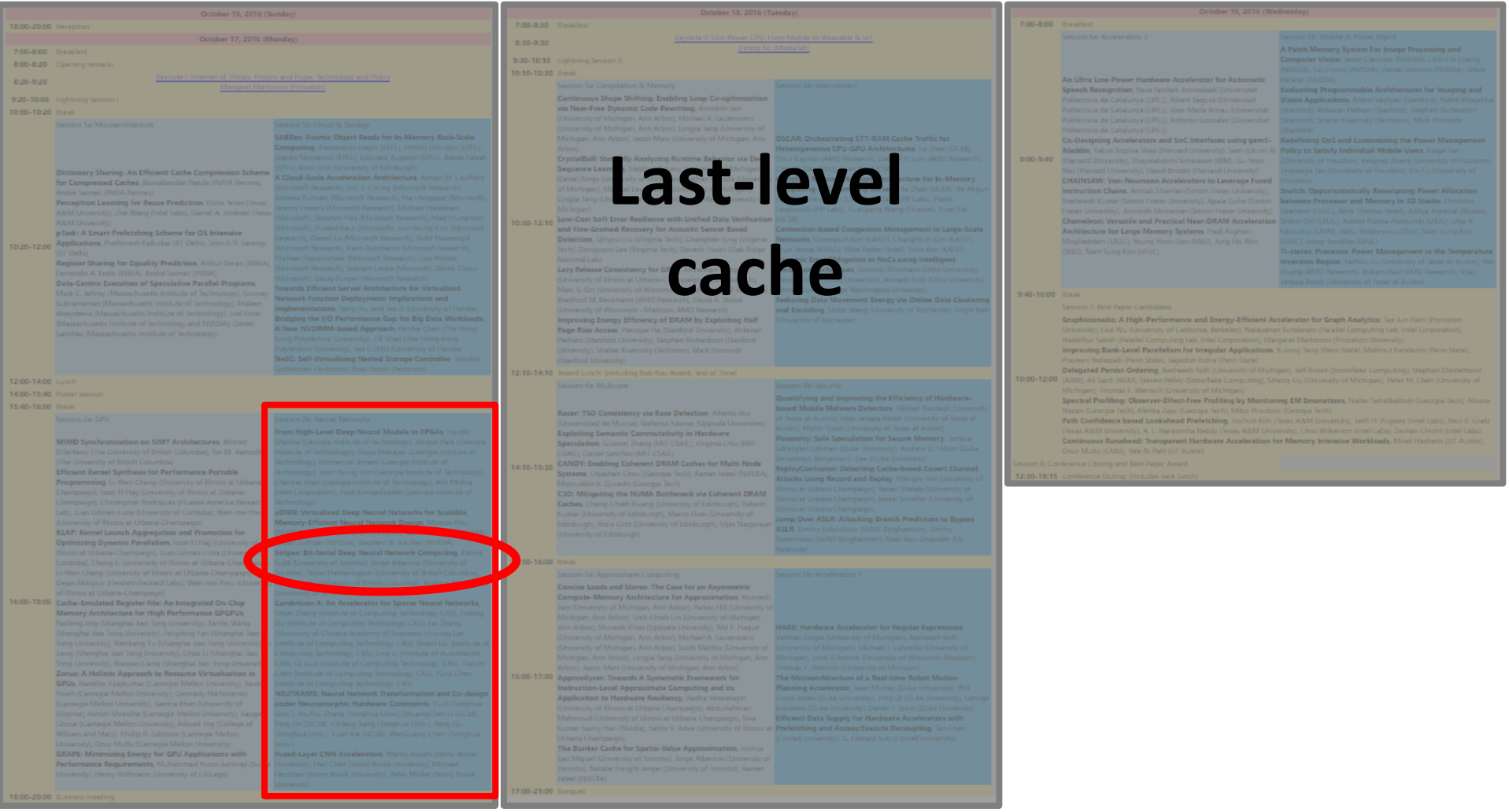
1.2x energy savings in MICRO!

3.48 megajoules / MICRO

The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

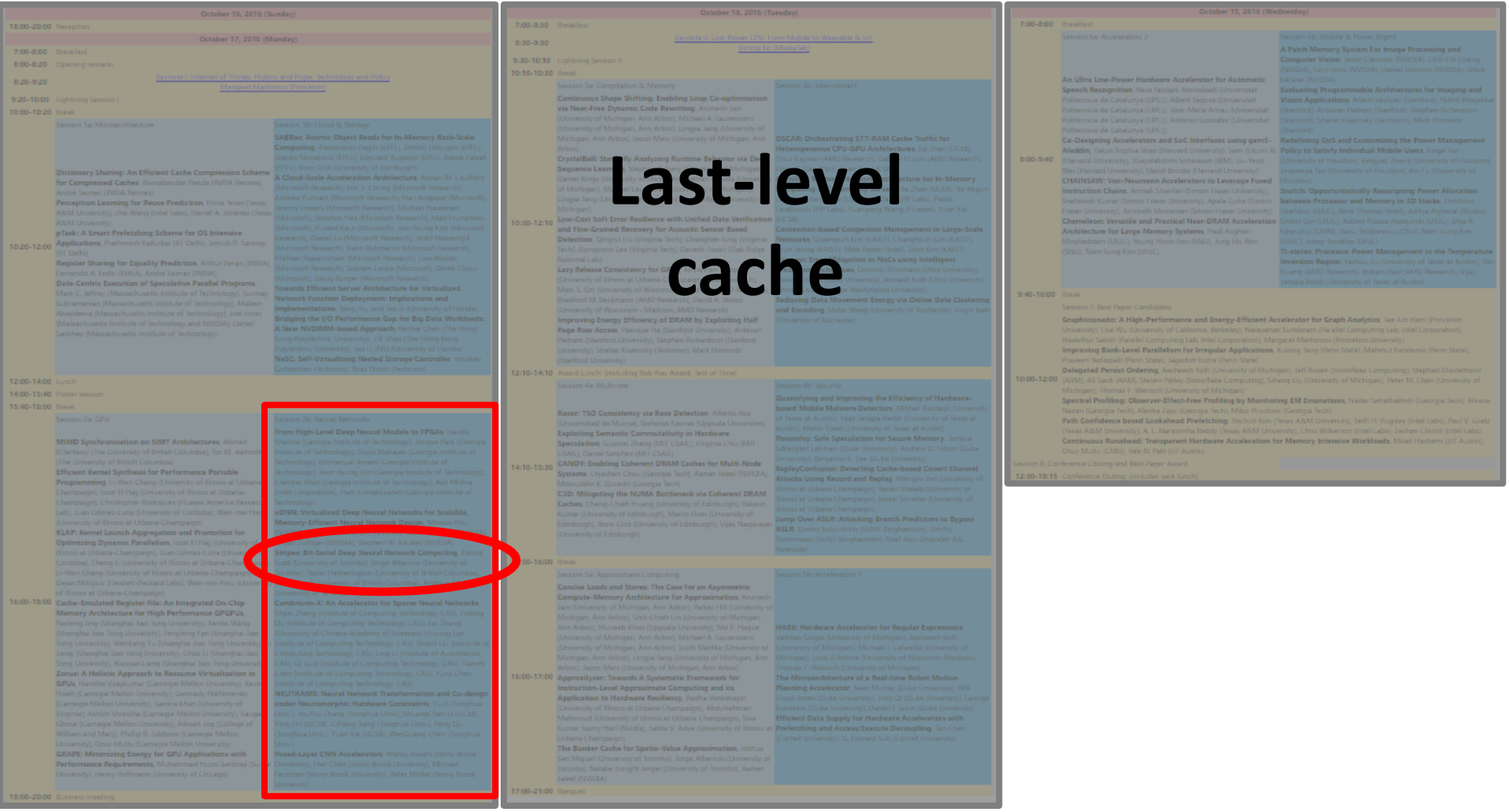
Joshua San Miguel, Jorge Albericio, Natalie Enright Jerger and Aamer Jaleel



The Bunker Cache for Spatio-Value Approximation

Session 5a, Tuesday 4:40pm

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



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