2015 IEEE 21st International Symposium on High Performance Computer Architecture

(HPCA 2015)

Burlingame, California, USA
7-11 February 2015

IEEE

IEEE Catalog Number: CFP15013-POD
ISBN: 978-1-4799-8931-7

Table of Contents

Message from the General Chair vii
Michael B. Taylor

Message from the Program Chair xi
Lieven Eeckhout

Exploring Architectural Heterogeneity in Intelligent Vision Systems 1
Nandhini Chandramoorthy (The Pennsylvania State University), Giuseppe Tagliavini (University of Bologna), Kevin Irick (The Pennsylvania State University), Antonio Pullini (ETH Zurich), Siddharth Advani (The Pennsylvania State University), Sulaiman Al Habsi (The Pennsylvania State University), Matthew Cotter (The Pennsylvania State University), John Sampson (The Pennsylvania State University), Vijaykrishnan Narayanan (The Pennsylvania State University) and Luca Benini (University of Bologna / ETH Zurich)

BeBoP: A Cost Effective Predictor Infrastructure for Superscalar Value Prediction 13
Arthur Perais, André Seznec (INRIA)

VSR Sort: A Novel Vectorised Sorting Algorithm and Architecture Extensions for Future Microprocessors 26
Timothy Hayes, Oscar Palomar, Osman Unsal, Adrian Cristal, Mateo Valero (Barcelona Supercomputing Center)

Increasing Multicore System Efficiency through Intelligent Bandwidth Shifting 39
Victor Jimenez (IBM Research), Alper Buyuktosunoglu (IBM Research), Pradip Bose (IBM Research), Francis P. O'Connell (IBM Systems and Technology Group), Francisco Cazorla (Barcelona Supercomputing Center), Mateo Valero (Barcelona Supercomputing Center)

Exploiting Compressed Block Size as an Indicator of Future Reuse 51
Gennady Pekhimenko, Tyler Huberty, Rui Cai, Onur Mutlu (Carnegie Mellon University), Phillip B. Gibbons, Michael A. Kozuch (Intel Labs), Todd C. Mowry (Carnegie Mellon University)

Talus: A Simple Way to Remove Cliffs in Cache Performance 64
Nathan Beckmann, Daniel Sanchez (MIT)

Coordinated Static and Dynamic Cache Bypassing for GPUs 76
Xiaolong Xie, Yun Liang (Peking University, China), Yu Wang (Tsinghua University, China), Guangyu Sun, Tao Wang (Peking University, China)

Priority-Based Cache Allocation for Throughput Processors 89
Dong Li (The University of Texas at Austin), Minsoo Rhu, Daniel R. Johnson (NVIDIA), Mike O'Connor (The University of Texas at Austin / NVIDIA), Mattan Erez (The University of Texas at Austin), Doug Burger (Microsoft), Donald S. Fussell (The University of Texas at Austin) and Stephen W. Keckler (The University of Texas at Austin) (NVIDIA)

Bamboo ECC: Strong, Safe, and Flexible Codes for Reliable Computer Memory 101
Jungrae Kim, Michael Sullivan, Mattan Erez (The University of Texas at Austin)

XChange: A Market-based Approach to Scalable Dynamic Multi-resource Allocation in Multicore Architectures 113
Xiaodong Wang, José Martínez (Cornell University)

Heterogeneous Memory Architectures: A HW/SW Approach for Mixing Die-stacked and Off-package Memories 126
Mitesh Meswani, Sergey Blagodurov, David Roberts, John Slice, Mike Ignatowski, Gabriel Loh (Advanced Micro Devices)

Event-based Scheduling for Energy-Efficient QoS (eQoS) in Mobile Web Applications 137
Yuhao Zhu, Matthew Halpern, Vijay Janapa Reddi (UT Austin)

Domain Knowledge Based Energy Management in Handhelds 150
Nachiappan Chidambaram Nachiappan, Praveen Yedlapalli (Pennsylvania State University), Niranjan Soundararajan (Intel), Anand Sivasubramaniam, Mahmut Kandemir (Pennsylvania State University) Ravi Iyer (Intel), Chita R. Das (Pennsylvania State University)

GPU Voltage Noise: Characterization and Hierarchical Smoothing of Spatial and Temporal Voltage Noise Interference in GPU Architectures 161
Jingwen Leng, Yazhou Zu, Vijay Janapa Reddi (The University of Texas at Austin)

Mascar: Speeding up GPU Warps by Reducing Memory Pitstops 174
Ankit Sethia, Davoud Anoushe Jamshidi, Scott Mahlke (University of Michigan)

Hierarchical Private/Shared Classification: the Key to Simple and Efficient Coherence for Clustered Cache Hierarchies 186
Alberto Ros (Universidad de Murcia), Mahdad Davari, Stefanos Kaxiras (Uppsala Universitet)

Flask Coherence: A Morphable Hybrid Coherence Protocol to Balance Energy, Performance and Scalability 198
Lucia G. Menezo, Valentin Puente, Jose Angel Gregorio (Univ. de Cantabria)

Prediction-Based Superpage-Friendly TLB Designs 210
Myrto Papadopoulou, Xin Tong (University of Toronto), Andre Seznec (INRIA), Andreas Moshovos (University of Toronto)

Supporting Superpages in Non-Contiguous Physical Memory 223
Yu Du, Miao Zhou, Bruce R. Childers, Daniel Mossé, Rami Melhem (University of Pittsburgh)

Paying to Save: Reducing Cost of Colocation Data Center via Rewards 235
Mohammad A. Islam, A.S.M. Hasan Mahmud, Shaolei Ren (Florida International University), Xiaorui Wang (The Ohio State University)

Octopus-Man: QoS-Driven Task Management for Heterogeneous Multicore in Warehouse-Scale Computers 246
Vinicius Petrucci (Federal University of Bahia, Salvador), Michael A. Laurenzano (University of Michigan, Ann Arbor), John Doherty (University of Michigan, Ann Arbor), Yunqi Zhang (University of Michigan, Ann Arbor), Daniel Mossé (University of Pittsburgh, Pittsburgh), Jason Mars (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor)

Understanding the Virtualization 'Tax' of Scale-out Pass-Through GPUs in GaaS Clouds: An Empirical Study 259
Ming Liu (University of Florida), Tao Li (University of Florida), Neo Jia, Andy Currid, Vladimir Troy (NVIDIA)

Adrenaline: Pinpointing and Reining in Tail Queries with Quick Voltage Boosting 271
Chang-Hong Hsu (University of Michigan, Ann Arbor), Yunqi Zhang (University of Michigan, Ann Arbor), Michael A. Laurenzano (University of Michigan, Ann Arbor), David Meisner (Facebook, Inc.), Thomas Wenisch (University of Michigan, Ann Arbor), Jason Mars (University of Michigan, Ann Arbor), Lingjia Tang (University of Michigan, Ann Arbor), Ronald G. Dreslinski (University of Michigan, Ann Arbor)

NDA: Near-DRAM Acceleration Architecture Leveraging Commodity DRAM Devices and Standard Memory Modules 283
Amin Farmahini-Farahani (University of Wisconsin-Madison), Jung Ho Ahn (Seoul National University), Katherine Morrow (University of Wisconsin-Madison), Nam Sung Kim (University of Wisconsin-Madison)

Alloy: Parallel-Serial Memory Channel Architecture for Single-Chip Heterogeneous Processor Systems 296
Hao Wang (UW-Madison), Chang-Jae Park (Samsung Electronics), Gyung-su Byun (Southern Methodist University), Jung Ho Ahn (Seoul National University), Nam Sung Kim (UW-Madison)

Reducing Read Latency of Phase Change Memory via Early Read and Turbo Read 309
Prashant J. Nair (Georgia Institute of Technology), Chiachen Chou (Georgia Institute of Technology), Bipin Rajendran (Indian Institute of Technology – Bombay), Moinuddin K. Qureshi (Georgia Institute of Technology)

CAFO: Cost Aware Flip Optimization for Asymmetric Memories 320
Rakan Maddah, Seyed Mohammad Seyedzadeh, Rami Melhem (University of Pittsburgh)

Understanding GPU Errors on Large-scale HPC Systems and the Implications for System Design and Operation 331
Devesh Tiwari, Saurabh Gupta, James Rogers, Don Maxwell (ORNL), Paolo Rech (Federal University of Rio Grande do Sul), Sudharshan Vazhkudai (ORNL), Daniel Oliveira (Federal University of Rio Grande do Sul), Dave Londo (Cray Inc.), Nathan Debardeleben (LANL), Philippe Navaux, Luigi Carro (Federal University of Rio Grande do Sul), Arthur Bland (ORNL)

High Performing Cache Hierarchies for Server Workloads 343
Aamer Jaleel, Joseph Nuzman, Adrian Moga, Simon Steely, Joel Emer (Intel)

Unlocking Bandwidth for GPUs in CC-NUMA Systems 354
Neha Agarwal (NVIDIA and University of Michigan), David Nellans, Mike O'Connor, Stephen W. Keckler (NVIDIA), Thomas F. Wenisch (University of Michigan)

Understanding Idle Behavior and Power Gating Mechanisms in the Context of Modern Benchmarks on CPU-GPU Integrated Systems 366
Manish Arora (AMD Research and UC San Diego), Srilatha Manne (Cavium Networks), Indrani Paul (AMD Research and Georgia Tech), Nuwan Jayasena (AMD Research), Dean M. Tullsen (UC San Diego)

Power Punch: Towards Non-blocking Power-gating of NoC Routers 378
Lizhong Chen (Oregon State University), Di Zhu, Massoud Pedram, Timothy M. Pinkston (University of Southern California)

Augmenting Low-latency HPC Network with Free-space Optical Links 390
Ikki Fujiwara (National Institute of Informatics), Michihiro Koibuchi (National Institute of Informatics), Tomoya Ozaki (Keio University), Hiroki Matsutani (Keio University), Henri Casanova (University of Hawai at Manoa)

SCOC: High-Radix Switches Made of Bufferless Clos Networks 402
Nikolaos Chrysos (IBM Research — Zurich), Cyriel Minkenberg (IBM Research — Zurich), Mark Rudquist, Claude Basso, Brian Vanderpool (IBM Systems & Technology Group)

Overcoming Far-end Congestion in Large-Scale Networks 415
Jongmin Won, Gwangsun Kim, John Kim (KAIST), Ted Jiang (NVIDIA), Mike Parker (Intel), Steve Scott (Cray)

iPatch: Intelligent Fault Patching to Improve Energy Efficiency 428
David J. Palframan, Nam Sung Kim, Mikko H. Lipasti (University of Wisconsin-Madison)

Balancing Reliability, Cost, and Performance Tradeoffs with FreeFault 439
Dong Wan Kim, Mattan Erez (University of Texas at Austin)

FTXen: Making Hypervisor Resilient to Hardware Faults on Relaxed Cores 451
Xinxin Jin (University of California, San Diego), Soyeon Park (Whova), Tianwei Sheng(Whova), Rishan Chen (University of California, San Diego), Zhiyong Shan(University of California, San Diego), Yuanyuan Zhou(University of California, San Diego)

Correction Prediction: Reducing Error Correction Latency for On-Chip Memories 463
Henry Duwe, Xun Jian, Rakesh Kumar (UIUC)

Overcoming the Challenges of Cross-Point Resistive Memory Architectures 476
Cong Xu (Pennsylvania State University), Dimin Niu (Pennsylvania State University) Naveen Muralimanohar (HP Labs), Rajeev Balasubramonian (University of Utah and HP Labs) Tao Zhang (Pennsylvania State University), Shimeng Yu (Arizona State University) Yuan Xie (University of California, Santa Barbara)

Adaptive-Latency DRAM: Optimizing DRAM Timing for the Common-Case 489
Donghyuk Lee, Yoongu Kim, Gennady Pekhimenko, Samira Khan, Vivek Seshadri, Kevin Chang, Onur Mutlu (Carnegie Mellon University)

CiDRA: A Cache-inspired DRAM Resilience Architecture 502
Young Hoon Son (Seoul National University), Sukhan Lee (Seoul National University), Seongil O (Seoul National University), Sanghyuk Kwon(Seoul National University), Nam Sung Kim (University of Wisconsin-Madison), Jung Ho Ahn (Seoul National University)

Tag Tables 514
Sean Franey, Mikko Lipasti (University of Wisconsin - Madison)

Architecture Exploration for Ambient Energy Harvesting Nonvolatile Processors 526
Kaisheng Ma (PSU), Yang Zheng (PSU), Shuangchen Li (PSU), Karthik Swaminathan (PSU), Xueqing Li (PSU), Yongpan Liu (Tsinghua University), Jack Sampson (PSU), Yuan Xie (UCSB), Vijaykrishnan Narayanan (PSU)

Scaling Distributed Cache Hierarchies through Computation and Data Co-Scheduling 538
Nathan Beckmann, Po-An Tsai, Daniel Sanchez (MIT)

Data Retention in MLC NAND Flash Memory: Characterization, Optimization and Recovery 551
Yu Cai, Yixin Luo (Carnegie Mellon University), Erich Haratsch (LSI Corporation), Ken Mai, Onur Mutlu (Carnegie Mellon University)

GPGPU Performance and Power Estimation Using Machine Learning 564
Gene Wu (The University of Texas at Austin), Joseph Greathouse (AMD), Alexander Lyashevsky (AMD), Nuwan Jayasena (AMD), Derek Chiou (The University of Texas at Austin)

Quantifying Sources of Error in McPAT and their Potential Impacts on Architectural Studies 577
Sam (Likun) Xi (Harvard University), Hans Jacobson (IBM), Pradip Bose (IBM), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

Studying the Impact of Multicore Processor Scaling on Directory Techniques via Reuse Distance Analysis 590
Minshu Zhao, Donald Yeung (University of Maryland, College Park)

SNNAP: Approximate Computing on Programmable SoCs via Neural Acceleration 603
Thierry Moreau, Mark Wyse , Jacob Nelson, Adrian Sampson (University of Washington), Hadi Esmaeilzadeh (Georgia Tech), Luis Ceze, Mark Oskin (University of Washington)

BRAINIAC: Bringing Reliable Accuracy Into Neurally-Implemented Approximate Computing 615
Beayna Grigorian, Nazanin Farahpour, Glenn Reinman (University of California, Los Angeles)

Scalable Communication Architecture for Network-Attached Accelerators 627
Sarah Neuwirth (University of Heidelberg), Dirk Frey (University of Heidelberg), Mondrian Nuessle (EXTOLL GmbH), Ulrich Bruening (University of Heidelberg)

Understanding Contention-based Covert Channels and Using Them for Defense..... 639
Casen Hunger, Mikhail Kazdagli, Ankit Rawat, Alex Dimakis, Sriram Vishwanath, Mohit Tiwari (UT Austin)

Malware-Aware Processors: A Framework for Efficient Online Malware Detection..... 651
Meltem Ozsoy, Caleb Donovick, Iakov Gorelik (Binghamton University), Nael Abu-Ghazaleh (University of California, Riverside), Dmitry Ponomarev (Binghamton University)

Run-Time Monitoring with Adjustable Overheads Using Dataflow-Guided Filtering..... 662
Daniel Lo, Tao Chen, Mohamed Ismail, G. Edward Suh (Cornell University)