

2014 IEEE 20th International Symposium on High Performance Computer Architecture

(HPCA 2014)

**Orlando, Florida, USA
15-19 February 2014**



IEEE Catalog Number: CFP14013-POD
ISBN: 978-1-4799-3098-2

Table of Contents

Session 1A: Caches

1. Locality-Aware Data Replication in the Last-Level Cache.....	1
<i>George Kurian (Massachusetts Institute of Technology), Srinivas Devadas (Massachusetts Institute of Technology), Omer Khan (University of Connecticut)</i>	
2. Adaptive Placement and Migration Policy for an STT-RAM-Based Hybrid Cache	13
<i>Zhe Wang (Texas A&M University), Daniel A. Jimenez (Texas A&M University), Cong Xu (Pennsylvania State University), Guangyu Sun (Peking University), Yuan Xie (Pennsylvania State University/AMD Research)</i>	
3. DASCA: Dead Write Prediction Assisted STT-RAM Cache Architecture	25
<i>Junwhan Ahn (Seoul National University), Sungjoo Yoo (Pohang University of Science and Technology), Kiyoung Choi (Seoul National University)</i>	
4. A Detailed GPU Cache Model Based on Reuse Distance Theory	37
<i>Cedric Nugteren (Eindhoven University of Technology), Gert-Jan van den Braak (Eindhoven University of Technology), Henk Corporaal (Eindhoven University of Technology), Henri Bal (Vrije Universiteit Amsterdam)</i>	

Session 1B: Reliability and Process Variation

1. Precision-Aware Soft Error Protection for GPUs.....	49
<i>David J. Palframan (University of Wisconsin-Madison), Nam Sung Kim (University of Wisconsin-Madison), Mikko H. Lipasti (University of Wisconsin-Madison)</i>	
2. Understanding the Impact of Gate-Level Physical Reliability Effects on Whole Program Execution	60
<i>Raghuraman Balasubramanian (University of Wisconsin-Madison), Karthikeyan Sankaralingam (University of Wisconsin-Madison)</i>	
3. Accordion: Toward Soft Near-Threshold Voltage Computing	72
<i>Ulya R. Karpuzcu (University of Minnesota), Ismail Akturk (University of Minnesota), Nam Sung Kim (University of Wisconsin-Madison)</i>	

4. Mosaic: Exploiting the Spatial Locality of Process Variation to Reduce Refresh Energy in On-Chip eDRAM Modules	84
<i>Aditya Agrawal (University of Illinois at Urbana-Champaign), Amin Ansari (University of Illinois at Urbana- Champaign), Josep Torrellas (University of Illinois at Urbana-Champaign)</i>	

Session 2A: Race Detection and Instruction Monitoring

1. Low-Overhead and High Coverage Run-Time Race Detection Through Selective Meta-data Management	96
<i>Ruirui Huang (Intel), Erik Halberg (Cornell University), Andrew Ferraiuolo (Cornell University), G. Edward Suh (Cornell)</i>	
2. FADE: A Programmable Filtering Accelerator for Instruction-Grain Monitoring	108
<i>Sotiria Fytraki (EPFL), Evangelos Vlachos (Oracle Labs), Onur Kocberber (EPFL), Babak Falsafi (EPFL), Boris Grot (University of Edinburgh)</i>	
3. Dynamically Detecting and Tolerating IF-Condition Data Races.....	120
<i>Shanxiang Qi (Google), Abdullah A. Muzahid (University of Texas at San Antonio), Wonsun Ahn (University of Illinois at Urbana-Champaign), Josep Torrellas (University of Illinois at Urbana-Champaign)</i>	

Session 2B: Datacenters

1. Exploiting Thermal Energy Storage to Reduce Data Center Capital and Operating Expenses	132
<i>Wenli Zheng (The Ohio State University), Kai Ma (The Ohio State University), Xiaorui Wang (The Ohio State University)</i>	
2. Implications of High Energy Proportional Servers on Cluster-wide Energy Proportionality	142
<i>Daniel Wong (University of Southern California), Murali Annavaram (University of Southern California)</i>	
3. Strategies for Anticipating Risk in Heterogeneous System Design	154
<i>Marisabel Guervara (Duke University), Benjamin Lubin (Boston University) and Benjamin C. Lee (Duke University)</i>	

Session 3A: Coherence and Consistency

1. TSO-CC: Consistency directed cache coherence for TSO	165
<i>Marco Elver (University of Edinburgh), Vijay Nagarajan (University of Edinburgh)</i>	

2. Stash Directory: A Scalable Directory for Many-Core Coherence.....	177
<i>Socrates Demetriadis (University of Pittsburgh), Sangyeun Cho (University of Pittsburgh)</i>	
3. QuickRelease: A Throughput Oriented Approach to Release Consistency on GPUs	189
<i>Blake Hechtman (Duke University/AMD Research), Brad Beckmann (AMD Research), Derek Hower (Qualcomm Research), Mark Hill (University of Wisconsin-Madison), David Wood (University of Wisconsin-Madison), Steve Reinhardt (AMD Research), Shuai Che (AMD Research), Yingying Tian (Texas A&M, AMD Research)</i>	

Session 3B: Best of CAL BB#5

1. The Netflix Challenge: Datacenter Edition,

Christina Delimitrou (Stanford), Christos Kozyrakis (Stanford)

2. High Performance, Energy Efficient Chipkill Correct Memory with Multidimensional Parity,

Xun Jian (University of Illinois at Urbana-Champaign) John Sartori (University of Illinois at Urbana-Champaign), Henry Duwe (University of Illinois at Urbana-Champaign), Rakesh Kumar (University of Illinois at Urbana-Champaign)

3. Shrink-Fit: A Framework for Flexible Accelerator Sizing,

Michael Lyons (Harvard University), Gu-Yeon Wei (Harvard University), David Brooks (Harvard University)

4. Clumsy Flow Control for High-Throughput Bufferless On-Chip Networks,

Hanjoon Kim (KAIST), Yonggon Kim (KAIST), John Kim (KAIST)

Session 4A: Security and Cloning

1. A Non-Inclusive Memory Permissions Architecture for Protection Against Cross-Layer Attacks.....	201
<i>Jesse Elwell (SUNY Binghamton), Ryan Riley (Qatar University), Nael Abu-Ghazaleh (SUNY Binghamton), Dmitry Ponomarev (SUNY Binghamton)</i>	
2. Suppressing the Oblivious RAM Timing Channel While Making Information Leakage and Program Efficiency Trade-offs	213
<i>Christopher W. Fletcher (Massachusetts Institute of Technology), Ling Ren (Massachusetts Institute of Technology), Xiangyao Yu (Massachusetts Institute of Technology), Marten Van Dijk (University of Connecticut), Omer Khan (University of Connecticut), Srinivas Devadas (Massachusetts Institute of Technology)</i>	

3. Timing Channel Protection for Memory Controllers	225
<i>Yao Wang (Cornell University), Andrew Ferraiuolo (Cornell University), G. Edward Suh (Cornell University)</i>	
4. STM : Cloning the Spatial and Temporal Memory Access Behavior.....	237
<i>Amro Awad (North Carolina State University), Yan Solihin (North Carolina State University)</i>	

Session 4B: GPUs

1. A Scalable Multi-Path Microarchitecture for Efficient GPU Control Flow.....	248
<i>Ahmed ElTantawy (University of British Columbia), Jessica Wenjie Ma (University of British Columbia), Mike O'Connor (NVIDIA/University of Texas at Austin), Tor Aamodt (University of British Columbia)</i>	
2. Improving GPGPU Resource Utilization and Performance Through Alternative Thread Block Scheduling	260
<i>Minseok Lee (KAIST), Sukwoo Song (KAIST), Joosik Moon (KAIST), John Kim (KAIST), Woong Seo (Samsung), Yeongon Cho (Samsung), Soojung Ryu (Samsung)</i>	
3. MRPB: Memory Request Prioritization for Massively Parallel Processors	272
<i>Wenhai Jia (Princeton University), Kelly A. Shaw (University of Richmond), Margaret Martonosi (Princeton University)</i>	
4. Warp-Level Divergence in GPUs: Characterization, Impact and Mitigation ..	284
<i>Ping Xiang (North Carolina State University), Yi Yang (NEC Labs), Huiyang Zhou (North Carolina State University)</i>	

Session 5A: Interconnection Networks

1. MP3: Minimizing Performance Penalty for Power-gating of Clos Network-on-Chip	296
<i>Lizhong Chen (University of Southern California), Lihang Zhao (University of Southern California), Ruisheng Wang (University of Southern California), Timothy Mark Pinkston (University of Southern California)</i>	
2. Up By Their Bootstraps: Online Learning in Artificial Neural Networks for CMP Uncore Power Management	308
<i>Jae-Yeon Won (Texas A&M University), Xi Chen (Texas A&M University), Paul V. Gratz (Texas A&M University), Jiang Hu (Texas A&M University), Vassos Soteriou (Cyprus University of Technology)</i>	
3. QORE: A Fault Tolerant Network-on-Chip Architecture with Power-Efficient Quad-Function Channel (QFC) Buffers.....	320
<i>Dominic DiTomaso (Ohio University), Avinash Kodi (Ohio University), Ahmed Louri (The University of Arizona)</i>	

4. Transportation-Network Inspired Network-on-Chip	332
<i>Hanjoon Kim (KAIST), Gwangsun Kim (KAIST), Hwasoo Yeo (KAIST), Seungryoul Maeng (KAIST), John Kim (KAIST)</i>	

Session 5B: DRAM

1. Improving System Throughput and Fairness Simultaneously in CMP Systems via Dynamic Bank Partitioning	344
<i>Mingli Xie (Peking University), Dong Tong (Peking University), Kan Huang (Peking University), Xu Cheng (Peking University)</i>	
2. Improving DRAM Performance by Parallelizing Refreshes with Accesses ...	356
<i>Kevin Kai-Wei Chang (Carnegie Mellon University), Donghyuk Lee (Carnegie Mellon University), Zeshan Chishti (Intel), Chris Wilkerson (Intel), Alaa Alameldeen (Intel), Yoongu Kim (Carnegie Mellon University), Onur Mutlu (Carnegie Mellon University)</i>	
3. CREAM: A Concurrent-Refresh-Aware DRAM Memory System.....	368
<i>Tao Zhang (Pennsylvania State University), Matt Poremba (Pennsylvania State University), Cong Xu (Pennsylvania State University), Guangyu Sun (Peking University), Yuan Xie (Pennsylvania State University/AMD Research)</i>	
4. DraMon: Predicting Memory Bandwidth Usage of Multi-threaded Programs with High Accuracy and Low Overhead	380
<i>Wei Wang (University of Virginia), Tania Dey (University of Virginia), Jack Davidson (University of Virginia), Mary Lou Soffa (University of Virginia)</i>	

Session 6: Best Paper I

1. PVCoherence: Designing Flat Coherence Protocols for Scalable Verification	392
<i>Meng Zhang (Duke University), Jesse D. Bingham (Intel), John Erickson (Intel), Daniel J. Sorin (Duke University)</i>	
2. Atomic SC for Simple In-order Processors	404
<i>Dibakar Gope (University of Wisconsin-Madison), Mikko H. Lipasti (University of Wisconsin-Madison)</i>	
3. Concurrent and Consistent Virtual Machine Introspection with Hardware Transactional Memory	416
<i>Yutao Liu (Shanghai Jiao Tong University), Yubin Xia (Shanghai Jiao Tong University), Haibing Guan (Shanghai Jiao Tong University), Binyu Zang (Shanghai Jiao Tong University), Haibo Chen (Shanghai Jiao Tong University)</i>	

4. Practical Data Value Speculation for Future High-end Processors	428
<i>Arthur Perais (INRIA), André Seznec (INRIA)</i>	

Session 7: Best Paper II

1. Tangle: Route-Oriented Dynamic Voltage Minimization for Variation-Afflicted, Energy-Efficient On-Chip Networks 440
Amin Ansari (University of Illinois at Urbana-Champaign), Asit Mishra (Intel), Jianping Xu (Intel), Josep Torrellas (University of Illinois at Urbana-Champaign)
2. Improving Cache Performance by Exploiting Read-Write Disparity 452
Samira Khan (Intel/Carnegie Mellon University), Alaa R. Alameldeen (Intel), Chris Wilkerson (Intel), Onur Mutlu (Carnegie Mellon University), Daniel A. Jimenez (Texas A&M University)
3. NUAT: A Non-Uniform Access Time Memory Controller..... 464
Wongyu Shin (KAIST), Jeongmin Yang (KAIST), Jungwhan Choi (KAIST), Lee-Sup Kim (KAIST)

Session 8A: Industrial Track

1. Improving In-Memory Database Index Performance with Intel® Transactional Synchronization Extensions..... 476
Tomas Karnagel (TU Dresden), Roman Dementiev (Intel), Ravi Rajwar (Intel), Konrad Lai (Intel), Thomas Legler (SAP AG), Benjamin Schlegel (TU Dresden), Wolfgang Lehner (TU Dresden)
2. BigDataBench: a Big Data Benchmark Suite from Internet Services 488
Lei Wang (ICT, Chinese Academy of Sciences), Chunjie Luo (ICT, Chinese Academy of Sciences), Yongqiang He (Facebook), Jianfeng Zhan (ICT, Chinese Academy of Sciences), Kent Zhan (Tencent), Xiaona Li (Baldu), Yuqing Zhu (ICT, Chinese Academy of Sciences), Shujie Zhang (Huawei), Qiang Yang (ICT, Chinese Academy of Sciences), Bizhu Qiu (Yahoo!), Zhen Jia (ICT, Chinese Academy of Sciences)
3. 3D Stacking of High-Performance Processors..... 500
Philip Emma (IBM), Alper Buyuktosunoglu (IBM), Michael Healy (IBM), Krishnan Kailas (IBM), Valentin Puente (IBM), Roy Yu (IBM), Allan Hartstein (IBM), Pradip Bose (IBM), Jaime Moreno (IBM)

Session 8B: Non-volatile memory

1. Reducing the Cost of Persistence for Nonvolatile Heaps in End User Devices 512
Sudarsun Kannan (Georgia Institute of Technology), Ada Gavrilovska (Georgia Institute of Technology), Karsten Schwan (Georgia Institute of Technology)

2. Sprinkler: Maximizing Resource Utilization in Many-Chip Solid State Disks	524
<i>Myoungsoo Jung (The University of Texas at Dallas), Mahmut Kandemir (Pennsylvania State University)</i>	
3. Over-Clocked SSD: Safely Running Beyond Flash Memory Chip I/O Clock Specs.....	536
<i>Kai Zhao (Rensselaer Polytechnic Institute), Kalyana Venkataraman (Cavium), Xuebin Zhang (Rensselaer Polytechnic Institute), Jiangpeng Li (Shanghai Jiao Tong University), Ning Zheng (Rensselaer Polytechnic Institute), Tong Zhang (Rensselaer Polytechnic Institute)</i>	

Session 9A: Memory Management

1. GPUdmm: A High-Performance and Memory-Oblivious GPU Architecture Using Dynamic Memory Management.....	546
<i>Youngsok Kim (POSTECH), Jaewon Lee (POSTECH), Jae-Eon Jo (POSTECH), Jangwoo Kim (POSTECH)</i>	
2. Increasing TLB Reach by Exploiting Clustering in Page Translations.....	558
<i>Binh Pham (Rutgers University), Abhishek Bhattacharjee (Rutgers University), Yasuko Eckert (AMD Research), Gabriel H. Loh (AMD Research)</i>	
3. Supporting x86-64 Address Translation for 100s of GPU Lanes.....	568
<i>Jason Power (University of Wisconsin-Madison), Mark D. Hill (University of Wisconsin-Madison), David A. Wood (University of Wisconsin-Madison)</i>	

Session 9B: Power

1. Scalably Verifiable Dynamic Power Management.....	579
<i>Opeoluwa Matthews (Duke University), Meng Zhang (Duke University), Daniel J. Sorin (Duke University)</i>	
2. Revolver: Processor Architecture for Power Efficient Loop Execution	591
<i>Mitchell Hayenga (ARM), Mikko H. Lipasti (University of Wisconsin-Madison), Vignyan Reddy (University of Wisconsin-Madison)</i>	
3. Dynamic Management of TurboMode in Modern Multi-core Chips	603
603 <i>David Lo (Stanford), Christos Kozyrakis (Stanford)</i>	

Session 10A: Prefetching and Compression

1. Spare Register Aware Prefetching for Graph Algorithms on GPUs	614
<i>Nagesh B. Lakshminarayana (Georgia Institute of Technology), Hyesoon Kim (Georgia Institute of Technology)</i>	

2. Sandbox Prefetching: Safe, Run-Time Evaluation of Aggressive Prefetchers 626
Seth Pugsley (University of Utah), Zeshan Chishti (Intel), Chris Wilkerson (Intel), Troy Chuang (Intel), Robert Scott (Intel), Aamer Jaleel (Intel), Shih-Lien Lu (Intel), Kingsum Chow (Intel), Rajeev Balasubramonian (University of Utah)

3. Memzip: Exploiting Unconventional Benefits from Memory Compression 638
Ali Shafiee (University of Utah), Meysam Taassori (University of Utah), Rajeev Balasubramonian (University of Utah), Al Davis (University of Utah)

Session 10B: Threading

1. CDTT: Compiler-generated data-triggered threads..... 650
Hung-Wei Tseng (University of California, San Diego), Dean Tullsen (University of California, San Diego)

2. Accelerating Decoupled Look-ahead via Weak Dependence Removal: A Metaheuristic Approach 662
Raj Parihar (University of Rochester), Michael C. Huang (University of Rochester)

3. Undersubscribed Threading for High-Performance and Energy-Efficient Many-Core Execution 671
Wim Heirman (Ghent University), Trevor Carlson (Ghent University), Kenzo Van Craeynest (Ghent University), Ibrahim Hur (Intel), Aamer Jaleel (Intel), Lieven Eeckhout (Ghent University)

Author Index..... 687