

2013 IEEE 19th International Symposium on High Performance Computer Architecture

(HPCA 2013)

**Shenzhen, China
23-27 February 2013**



**IEEE Catalog Number: CFP13013-PRT
ISBN: 978-1-4673-5585-8**

Table of Contents

Session 1: Power and Energy I

Power Struggles: Revisiting the RISC v CISC Debate on Contemporary ARM and x86 Architectures	1
<i>Emily Blem, Jaikrishnan Menon, and Karthikeyan Sankaralingam (University of Wisconsin)</i>	
High-performance and Energy-efficient Mobile Web Browsing	13
<i>Yuhao Zhu and Vijay Janapa Reddi (University of Texas)</i>	
Skinflint DRAM System: Minimizing DRAM Chip Writes for Low Power	25
<i>Yebin Lee, Soontae Kim, Seokin Hong, and Jongmin Lee (KAIST)</i>	
Enabling Distributed Generation Powered Sustainable High-Performance Data Center	35
<i>Chao Li, Ruijin Zhou, and Tao Li (University of Florida)</i>	

Session 2A: Microarchitecture

A Group-Commit Mechanism for ROB-Based Processors Implementing the x86 ISA	47
<i>Furat Afram, Hui Zeng, and Kanad Ghose (SUNY at Binghamton)</i>	
Store-Load-Branch (SLB) Predictor: A Compiler Assisted Branch Prediction for Data Dependent Branches	59
<i>Muhammad Umar Farooq and Lizy Kurian John (University of Texas)</i>	
Two Level Bulk Preload Branch Prediction	71
<i>James Bonanno, Adam Collura, Daniel Lipetz, Ulrich Mayer, Brian Prasky, and Anthony Saporito (IBM)</i>	

Session 2B: Scheduling and Virtual Machines

ReCaP: A Region-Based Cure for the Common Cold (Cache)	83
<i>Jason Zebchuk, Harold W. Cain, Xin Tong, Vijayalakshmi Srinivasan, and Andreas Moshovos (University of Toronto, IBM)</i>	
Navigating Heterogeneous Processors with Market Mechanisms	95
<i>Marisabel Guevara, Benjamin Lubin, and Benjamin C. Lee (Duke University, Boston University)</i>	

Application-to-Core Mapping Policies to Reduce Memory System Interference in Multi-Core Systems	107
<i>Reetuparna Das, Rachata Ausavarungnirun, Onur Mutlu, Akhilesh Kumar, and Mani Azimi (University of Michigan, Carnegie Mellon University, Intel)</i>	

Session 3A: Caches

Improving Multi-Core Performance Using Mixed-Cell Cache Architecture	119
<i>Samira Khan, Alaa R. Alameldeen, Chris Wilkerson, Jaydeep Kulkarni, and Daniel A. Jimenez (University of Texas at San Antonio, Intel)</i>	

ECM : Effective Capacity Maximizer for High-Performance Compressed Caching.....	131
<i>Seungcheol Baek, Hyung Gyu Lee, Chrysostomos Nicopoulos, Junghee Lee, and Jongman Kim (Georgia Institute of Technology, University of Daegu, University of Cyprus)</i>	

Technology Comparison for Large Last-Level Caches (L3Cs):Low-Leakage SRAM, Low Write-Energy STT-RAM, and Refresh-Optimized eDRAM.....	143
<i>Mu-Tien Chang, Paul Rosenfeld, Shih-Lien Lu, and Bruce Jacob (University of Maryland, Intel)</i>	

Modeling Performance Variation Due to Cache Sharing	155
<i>Andreas Sandberg, Andreas Sembrant, Erik Hagersten, and David Black-Schaffer (Uppsala University)</i>	

Session 3B: Industrial Track

A Novel System Architecture for Web Scale Applications Using Lightweight CPUs and Virtualized I/O	167
<i>Kshitij Sudan, Saisanthosh Balakrishnan, Sean Lie, Min Xu, Dhiraj Mallick, Gary Lauterbach, and Rajeev Balasubramonian (University of Utah, SeaMicro, AMD))</i>	

Cost Effective Data Center Servers	179
<i>Rui Hou, Tao Jiang, Liuhan Zhang, Pengfei Qi, Jianbo Dong, Haibin Wang, Xiongli Gu, and Shujie Zhang (Chinese Academy of Sciences, Huawei)</i>	

Optimizing Google’s Warehouse Scale Computers: The NUMA Experience.....	188
<i>Lingjia Tang, Jason Mars, Xiao Zhang, Robert Hagmann, Robert Hundt, and Eric Tune (UCSD, Google)</i>	

Runnemed: An Architecture for Ubiquitous High-Performance Computing	198
<i>Nicholas Carter, Aditya Agrawal, Shekhar Borkar, Romain Cledat, Howard David, Dave Dunning, Joshua Fryman, Ivan Ganey, Roger Golliver, Rob Knauerhase, Richard Lethin, Benoit Meister, Asit Mishra, Wilfred Pinfold, Justin Teller, Josep Torrellas, Nicolas Vasilache, Ganesh Venkatesh, and Jianping Xu (Intel, UIUC, Reservoir Labs)</i>	

Session 4A: Non-volatile Memory

- Exploring High-Performance and Energy Proportional Interface for Phase Change Memory Systems..... 210
Zhongqi Li, Ruijin Zhou and Tao Li (University of Florida)
- Coset Coding to Extend the Lifetime of Memory 222
Adam N. Jacobvitz, Robert Calderbank, and Daniel J. Sorin (Duke University)
- i²WAP: Improving Non-Volatile Cache Lifetime by Reducing Inter- and Intra-Set Write Variations..... 234
Jue Wang, Xiangyu Dong, Yuan Xie , and Norman P. Jouppi (Pennsylvania State University, Hewlett-Packard Labs, AMD)

Session 4B: Secure and Reliable Architectures

- Architecture Support for Guest-Transparent VM Protection from Untrusted Hypervisor and Physical Attacks..... 246
Yubin Xia, Yutao Liu, and Haibo Chen (Shanghai Jiao Tong University)
- SCRAP: Architecture for Signature-based Detection of Code Reuse Attacks 258
Mehmet Kayaalp, Timothy Schmitt, Junaid Nomani, Dmitry Ponomarev, and Nael Abu-Ghazaleh (Binghamton University)
- Adaptive Reliability Chipkill Correct (ARCC) 270
Xun Jian and Rakesh Kumar (UIUC)

Session 5: Best Paper Nominees

- Accelerating Write by Exploiting PCM Asymmetries 282
Jianhui Yue and Yifeng Zhu (University of Maine)
- Hybrid Latency Tolerance for Robust Energy-Efficiency on 1000-Core Processors..... 294
Neal C. Crago, Omid Azizi, Steven S. Lumetta, and Sanjay J. Patel (Intel, UIUC, HiCAMP Systems)
- Optimizing Virtual Machine Scheduling in NUMA Multicore Systems..... 306
Jia Rao, Kun Wang, Xiaobo Zhou, and Cheng-Zhong Xu (University of Colorado at Colorado Springs, Wayne State University)
- Sonic Millip3De: A Massively Parallel 3D Stacked Accelerator for 3D Ultrasound..... 318
Richard Sampson, Ming Yang, Siyuan Wei, and Chaitali Chakrabarti, and Thomas F. Wenisch (University of Michigan, Arizona State University)

Session 6A: GPUs I

Power-efficient Computing for Compute-intensive GPGPU Applications	330
<i>Syed Zohaib Gilani, Nam Sung Kim, and Michael J. Schulte (University of Wisconsin, AMD Research)</i>	
Power-performance Co-optimization of GPU Core Architecture using Resistive Memory	342
<i>Nilanjan Goswami, Bingyi Cao, and Tao Li (University of Florida)</i>	
Reducing GPU Offload Latency Via Fine-Grained CPU-GPU Synchronization	354
<i>Daniel Lustig and Margaret Martonosi (Princeton University)</i>	

Session 6B: Interconnection Networks I

Worm-Bubble Flow Control	366
<i>Lizhong Chen and Timothy Mark Pinkston (USC)</i>	
Breaking the on-chip latency barrier using SMART	378
<i>Tushar Krishna, Chia-Hsin Owen Chen, Woo Cheol Kwon, and Li-Shiuan Peh (MIT)</i>	
TS-Router: On Maximizing the Quality-of-Allocation in the On-Chip Network	390
<i>Yuan-Ying Chang, Yoshi Shih-Chieh Huang, Matthew Poremba, Vijaykrishnan Narayanan, Yuan Xie, and Chung-Ta King (National Tsing Hua University, Pennsylvania State University)</i>	

Session 7B: Power and Energy II

Refrint: Intelligent Refresh to Minimize Power in On-Chip Multiprocessor Cache Hierarchies.....	400
<i>Aditya Agrawal, Prabhat Jain, Amin Ansari, and Josep Torrellas (UIUC)</i>	
Warped Register File: A Power Efficient Register File for GPGPUs	412
<i>Mohammad Abdel-Majeed and Murali Annavaram (USC)</i>	
Disintgerated Control for Energy-Efficient and Heterogeneous Memory Systems	424
<i>Tae Jun Ham, Bharath K. Chelepalli, Neng Xue, and Benjamin C. Lee (Duke University)</i>	

Session 8A: Heterogeneous and Adaptive Architectures

Illusionist: Transforming Lightweight Cores into Aggressive Cores on Demand	436
<i>Amin Ansari, Shuguang Feng, Shantanu Gupta, Josep Torrellas, and Scott Mahlke (UIUC, University of Michigan)</i>	
ESESC: A Fast Multicore Simulator Using Time-Based Sampling.....	448
<i>Ehsan K. Ardestani and Jose Renau (UC Santa Cruz)</i>	
How to Implement Effective Prediction and Forwarding for Fusable Dynamic Multicore Architectures	460

Behnam Robotmili, Dong Li, Hadi Esmaeilzadeh, Sibi Govindan, Aaron Smith, Andrew Putnam, Doug Burger, and Stephen W. Keckler (Qualcomm, University of Texas, University of Washington, AMD, Microsoft Research, NVIDIA)

Bridging the Semantic Gap: Emulating Biological Neuronal Behaviors with Simple Digital Neurons 472
Andrew Nere, Atif Hashmi, Mikko Lipasti, and Giulio Tononi (University of Wisconsin)

Session 8B: Interconnection Networks II

Layout-conscious Random Topologies for HPC Off-chip Interconnects..... 484
*Michihiro Koibuchi, Ikki Fujiwara, Hiroki Matsutani, and Henri Casanova
(National Institute of Informatics, Keio University, University of Hawaii at Manoa)*

Scaling Towards Kilo-Core Processors with Asymmetric High Radix Topologies..... 496
*Nilmini Abeyratne, Reetuparna Das, Qingkun Li, Korey Sewell, Bharan Giridhar,
Ronald Dreslinski, David Blaauw, and Trevor Mudge (University of Michigan)*

Energy-Efficient Interconnect via Router Parking 508
Ahmad Samih, Ren Wang, Anil Krishna, Christian Maciocco, and Solihin Yan (NCSU, Intel Labs)

In-Network Traffic Regulation for Transactional Memory 520
Lihang Zhao, Woojin Choi, Lizhong Chen, and Jeffrey Draper (USC)

Session 9A: Near-threshold Computing

Macho: A Failure Model-oriented Adaptive Cache Architecture to enable Near-Threshold
Voltage Scaling..... 532
Tayyeb Mahmood, Soontae kim, and Seokin Hong (KAIST)

EnergySmart:Toward Energy-Efficient Manycores for Near-Threshold Computing 542
*Ulya R. Karpuzcu, Abhishek Sinkar, Nam Sung Kim, and Josep Torrellas (University of Minnesota,
University of Wisconsin, UIUC)*

Session 9B: Coherence and Consistency

Rainbow: Efficient Memory Dependence Recording with High Replay Parallelism for Relaxed Memory
Model 554
*Xuehai Qian, He Huang, Benjamin Sahelices, and Depei Qian
(UIUC, AMD, University of Valladolid, Beihang University)*

High-Speed Formal Verification of Heterogeneous Coherence Hierarchies..... 566
Jesse G Beu, Jason A. Poovey, Eric R. Hein, and Thomas M. Conte (Georgia Institute of Technology)

Session 10A: GPUS II

Cache Coherence for GPU Architectures	578
<i>Inderpreet Singh, Arrvinth Shriraman, Wilson W. L. Fung, Mike O'Connor, and Tor M. Aamodt (University of British Columbia, Simon Fraser University, AMD, Stanford)</i>	
The Dual-Path Execution Model for Efficient GPU Control Flow	591
<i>Minsoo Rhu and Mattan Erez (The University of Texas)</i>	
A Multiple SIMD, Multiple Data (MSMD) Architecture: Parallel Execution of Dynamic and Static SIMD Fragments.....	603
<i>Yaohua Wang, Shuming Chen, Jianghua Wan, Jiayuan Meng, Kai Zhang, Wei Liu, and Xi Ning (National University of Defense Technology, Argonne National Laboratory)</i>	

Session 10B: DRAM/Memory

Tiered-Latency DRAM: A Low Latency and Low Cost DRAM Architecture	615
<i>Donghyuk Lee, Yoongu Kim, Vivek Seshadri, Jamie Liu, Lavanya Subramanian, and Onur Mutlu (CMU)</i>	
A Case for Refresh Pausing in DRAM Memory Systems.....	627
<i>Prashant Nair, Chia Chen Chou, and Moinuddin K. Qureshi (Georgia Tech)</i>	
MISE: Providing Performance Predictability and Improving Fairness in Shared Main Memory Systems	639
<i>Lavanya Subramanian, Vivek Seshadri, Yoongu Kim, Ben Jaiyen, and Onur Mutlu (CMU)</i>	
Author Index	651