

2017 50th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 2017)

**Boston, Massachusetts, USA
14-17 October 2017**



**IEEE Catalog Number: CFP17071-POD
ISBN: 978-1-4503-4952-9**

**Copyright © 2017, Association for Computing Machinery (ACM)
All Rights Reserved**

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP17071-POD
ISBN (Print-On-Demand):	978-1-4503-4952-9
ISSN:	1072-4451

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Contents

MICRO-50 Organizing Committee	xiii
MICRO-50 Program Committee	xiv
MICRO-50 External Review Committee	xv
MICRO-50 Sponsors	xvii
Keynote I “To a Trillion and Beyond: the Future of the Internet of Things” Krisztián Flautner, VP Technology, ARM	xviii
Keynote II “Specialization and Accelerated AI at Hyperscale” Doug Burger, Distinguished Engineer (MSR NExT), Microsoft	xix
Session 1-A: DRAM	
Banshee: Bandwidth-Efficient DRAM Caching Via Software/Hardware Cooperation	1
Xiangyao Yu:MIT; Christopher J. Hughes:Intel; Nadathur Satish:Intel; Onur Mutlu:ETH; Srinivas Devadas:MIT	
ConTutto - A Novel FPGA-based Prototyping Platform Enabling Innovation in the Memory Subsystem of a Server Class Processor	15
Bharat Sukhwani:IBM Research; Thomas Roewer:IBM Research; Charles L. Haymes:IBM Research; Kyu-Hyoun Kim:IBM Research; Adam J. McPadden:IBM; Daniel M. Dreps: IBM; Dean Sanner:IBM; Jan Van Lunteren:IBM Research; Sameh Asaad:IBM Research	
Detecting and Mitigating Data-Dependent DRAM Failures by Exploiting Current Memory Content	27
Samira Khan:University of Virginia; Chris Wilkerson:Intel; Zhe Wang:Intel; Alaa Alameldeen:Intel; Donghyuk Lee:NVIDIA; Onur Mutlu:ETH Zurich	
Fine-Grained DRAM: Energy Efficient DRAM for Extreme Bandwidth Systems	41
Mike O’Connor:NVIDIA / UT-Austin; Niladrish Chatterjee:NVIDIA; Donghyuk Lee: NVIDIA; John Wilson:NVIDIA; Aditya Agrawal:NVIDIA; Stephen W. Keckler:NVIDIA / UT-Austin; William J. Dally:NVIDIA / Stanford	
Session 1-B: Accelerators	
UDP: A Programmable Accelerator for Extract-Transform-Load Workloads and More	55
Yuanwei Fang:University of Chicago; Chen Zou:University of Chicago; Aaron J. Elmore: University of Chicago; Andrew A. Chien:University of Chicago and Argonne National Laboratory	
UNFOLD: A Memory-Efficient Speech Recognizer Using On-The-Fly WFST Composition	69
Reza Yazdani:Universitat Politecnica de Catalunya; Jose-Maria Arnau: Universitat Politecnica de Catalunya; Antonio Gonzalez:Universitat Politecnica de Catalunya	
IDEAL: Image DEnoising AcceLerator	82
Mostafa Mahmoud:University of Toronto; Bojian Zheng:University of Toronto; Alberto Delmas Lascorz:University of Toronto; Felix Heide: Stanford University/Algolux; Jonathan Assouline: Algolux; Paul Boucher: Algolux; Emmanuel Onzon: Algolux; Andreas Moshovos: University of Toronto	
Pipelining a Triggered Processing Element	96
Thomas J. Repetti:Columbia University; Joao Pedro Cerqueira:Columbia University; Martha A. Kim:Columbia University; Mingoo Seok:Columbia University	

Session 2-A: GPU-I

Efficient Exception Handling Support for GPUs	109
Ivan Tanasic:Universitat Politecnica de Catalunya / Barcelona Supercomputing Center; Isaac Gelado:NVIDIA; Marc Jorda:Barcelona Supercomputing Center; Eduard Ayguade: Universitat Politecnica de Catalunya / Barcelona Supercomputing Center; Nacho Navarro:Universitat Politecnica de Catalunya / Barcelona Supercomputing Center	
Beyond the Socket: NUMA-Aware GPUs	123
Ugljesa Milic:Barcelona Supercomputing Center / Universitat Politecnica de Catalunya; Oreste Villa:Nvidia; Evgeny Bolotin:Nvidia; Akhil Arunkumar:Arizona State University; Eiman Ebrahimi:Nvidia; Aamer Jaleel:Nvidia; Alex Ramirez:Google; David Nellans: Nvidia	
Mosaic: A GPU Memory Manager with Application-Transparent Support for Multiple Page Sizes	136
Rachata Ausavarungnirun:Carnegie Mellon University; Joshua Landgraf:UT Austin; Vance Miller:UT Austin; Saugata Ghose:Carnegie Mellon University; Jayneel Gandhi: VMware Research Group; Christopher J. Rossbach:UT Austin and VMware Research Group; Onur Mutlu:ETH Zurich and Carnegie Mellon University	
RegLess: Just-in-Time Operand Staging for GPUs	151
John Kloosterman:University of Michigan; Jon Beaumont:University of Michigan; Davoud Anoushe Jamshidi:University of Michigan; Jonathan Bailey:University of Michigan; Trevor Mudge:University of Michigan; Scott Mahlke:University of Michigan	
SCRATCH: An End-to-end Application-aware Soft-GPGPU Architecture and Trimming Tool	165
Pedro Duarte:Instituto de Telecomunicacoes, University of Coimbra, Portugal; Pedro Tomas:INESC-ID, Instituto Superior Tecnico, Universidade de Lisboa, Portugal; Gabriel Falcao:Instituto de Telecomunicacoes, University of Coimbra, Portugal	

Session 2-B: Non-Volatile Memory/Storage

Proteus: A Flexible and Fast Software Supported Hardware Logging approach for NVM	178
Seunghee Shin:North Carolina State University; Satish Kumar Tirukkovalluri:North Carolina State University; James Tuck:North Carolina State University; Yan Solihin: North Carolina State University	
Efficient Support of Position Independence on Non-Volatile Memory	191
Guoyang Chen:Qualcomm; Lei Zhang:NCSU; Richa Budhiraja:Qualcomm; Xipeng Shen:NCSU; Youfeng Wu:Intel	
Incidental Computing on IoT Nonvolatile Processors	204
Kaisheng Ma: Penn State; Xuqing Li: Penn State; Jinyang Li:Tsinghua Univ.; Yongpan Liu: Tsinghua Univ.; Yuan Xie:UCSB; Jack Sampson:Penn State; Mahmut Taylan Kandemir:Penn State; Vijaykrishnan Narayanan:Penn State;	
Summarizer: Trading Communication with Computing Near Storage	219
Gunjae Koo:University of Southern California; Kiran Kumar Matam:University of Southern California; Te I:North Carolina State University; H. V. Krishna Giri Narra: University of Southern California; Jing Li:University of California, San Diego; Hung-Wei Tseng:North Carolina State University; Steven Swanson:University of California, San Diego; Murali Annavaram:University of Southern California	
Memory Cocktail Therapy: A General Learning-Based Framework to Optimize Dynamic Tradeoffs in NVMs	232
Zhaoxia Deng:University of California, Santa Barbara; Lunkai Zhang:University of Chicago; Nikita Mishra:University of Chicago; Henry Hoffmann:University of Chicago; Fred Chong:University of Chicago	

Session 3-A: In/Near Memory Computing

A Many-core Architecture for In-Memory Data Processing	245
---	-----

Sandeep R Agrawal:Oracle Labs; Sam Idicula:Oracle Labs; Arun Raghavan:Oracle Labs; Evangelos Vlachos:Oracle Labs; Venkatraman Govindaraju:Oracle Labs; Venkatanathan Varadarajan:Oracle Labs; Cagri Balkesen:Oracle Labs; Georgios Giannikis:Oracle Labs; Charlie Roth:NXP; Nipun Agarwal:Oracle Labs; Eric Sedlar:Oracle Labs	
Cache Automaton	259
Arun Subramaniyan:University of Michigan; Jingcheng Wang:University of Michigan; Ezhil R. M. Balasubramanian:University of Michigan; David Blaauw:University of Michigan; Dennis Sylvester:University of Michigan; Reetuparna Das:University of Michigan	
Ambit: In-Memory Accelerator for Bulk Bitwise Operations Using Commodity DRAM Technology	273
Vivek Seshadri:Microsoft Research India; Donghyuk Lee:NVIDIA Research; Thomas Mullins:Intel; Hasan Hassan:ETH Zurich; Amirali Boroumand:CMU; Jeremie Kim:ETH Zurich; Michael A. Kozuch:Intel; Onur Mutlu:ETH Zurich; Phillip B. Gibbons:CMU; Todd C. Mowry:CMU	
DRISA: A DRAM-based Reconfigurable In-Situ Accelerator	288
Shuangchen Li:University of California, Santa Barbara; Dimin Niu:Samsung; Krishna T. Malladi:Samsung; Hongzhong Zheng:Samsung; Bob Brennan:Samsung; Yuan Xie:University of California, Santa Barbara	
PageForge: A Near-Memory Content-Aware Page-Merging Architecture	302
Dimitrios Skarlatos:University of Illinois at Urbana-Champaign; Nam Sung Kim:University of Illinois at Urbana-Champaign; Josep Torrellas:University of Illinois at Urbana-Champaign	
Session 3-B: Security	
RHMD: Evasion-Resilient Hardware Malware Detectors	315
Khaled N. Khasawneh:University of California, Riverside; Nael Abu-Ghazaleh:University of California, Riverside; Dmitry Ponomarev:Binghamton University; Lei Yu:Binghamton University	
Software-based Gate-level Information Flow Security for IoT Systems	328
Hari Cherupalli:University of Minnesota; Henry Duwe:University of Illinois; Weidong Ye:University of Illinois; Rakesh Kumar:University of Illinois; John Sartori:University of Minnesota	
How secure is your cache against side-channel attacks?	341
Zecheng He:Princeton University; Ruby B. Lee:Princeton University	
Constructing and Characterizing Covert Channels on GPGPUs	354
Hoda Naghibijouybari:University of California, Riverside; Khaled Khasawneh:University of California, Riverside; Nael Abu-Ghazaleh:University of California, Riverside	
Session 4-A: Deep Learning	
Scale-Out Acceleration for Machine Learning?	367
Jongse Park:Georgia Institute of Technology; Hardik Sharma:Georgia Institute of Technology; Divya Mahajan:Georgia Institute of Technology; Joon Kyung Kim:Georgia Institute of Technology; Preston Olds:Georgia Institute of Technology; Hadi Esmaeilzadeh:University of California, San Diego	
Bit-Pragmatic Deep Neural Network Computing	382
Jorge Albericio:NVIDIA; Patrick Judd:University of Toronto; Alberto Delmas:University of Toronto; Sayeh Sharify:University of Toronto; Gerard O'Leary:University of Toronto; Roman Genov:University of Toronto; Andreas Moshovos:University of Toronto	
CirCNN: Accelerating and Compressing Deep Neural Networks Using Block-Circulant Weight Matrices	395

Caiwen Ding:Syracuse University; Siyu Liao:City University of New York, City College;
 Yanzhi Wang:Syracuse University; Zhe Li:Syracuse University; Ning Liu:Syracuse University;
 Youwei Zhuo:University of Southern California; Chao Wang:University of Southern
 California;Xuehai Qian:University of Southern California; Yu Bai:California State University
 Fullerton; Geng Yuan:Syracuse University; Xiaolong Ma:Syracuse University; Yipeng
 Zhang:Syracuse University; Jian Tang:Syracuse University; Qinru Qiu:Syracuse University;
 Xue Lin:Northeastern University; Bo Yuan:City University of New York, City College

Session 4-B: Prediction

- Using Branch Predictors to Predict Brain Activity in Brain-Machine Implants** . 409
 Abhishek Bhattacharjee:Rutgers University/Princeton University
- Load Value Prediction via Path-based Address Prediction: Avoiding Mispredictions
 due to Conflicting Stores** 423
 Rami Sheikh:Qualcomm Technologies, Inc.; Harold W. Cain:Qualcomm Datacenter
 Technologies, Inc.; Raguram Damodaran:Qualcomm Technologies, Inc.
- Multiperspective Reuse Prediction** 436
 Daniel A. Jiménez:Texas A&M; Elvira Teran:Texas A&M

Session 5-A: Coherence/Consistency, Translation and Virtual Memory

- CSALT: Context Switch Aware Large TLB** 449
 Yashwant Marathe:The University of Texas at Austin; Nagendra Guler:Texas Instruments;
 Jee Ho Ryoo:The University of Texas at Austin; Shuang Song:The University of Texas
 at Austin; Lizy K. John:The University of Texas at Austin
- RTLCheck: Verifying the Memory Consistency of RTL Designs** 463
 Yatin A. Manerkar:Princeton University; Daniel Lustig:NVIDIA; Margaret Martonosi:
 Princeton University; Michael Pellauer:NVIDIA
- Architecting Hierarchical Coherence Protocols for Push-button Parametric Verification**
 477
 Opeoluwa Matthews:Duke University; Daniel J. Sorin:Duke University
- PARSNIP: Performant Architecture for Race Safety with No Impact on Precision** 490
 Yuanfeng Peng:University of Pennsylvania; Benjamin P. Wood:Wellesley College; Joseph
 Devietti:University of Pennsylvania

Session 5-B: Energy

- Harnessing Voltage Margins for Energy Efficiency in Multicore CPUs** 503
 George Papadimitriou:University of Athens; Manolis Kaliorakis:University of Athens;
 Athanasios Chatzidimitriou:University of Athens; Dimitris Gizopoulos:University of
 Athens; Peter Lawthers:AppliedMicro; Shidhartha Das:ARM
- Race-To-Sleep + Content Caching + Display Caching: A Recipe for Energy-efficient
 Video Streaming on Handhelds** 517
 Haibo Zhang:Penn State; Prasanna Venkatesh Rengasamy:Penn State; Shulin Zhao:Penn
 State; Nachiappan Chidambaram Nachiappan:Penn State; Anand Sivasubramaniam:Penn
 State; Mahmut Kandemir:Penn State; Ravi Iyer:Intel; Chita R. Das:Penn State
- BVF: Enabling Significant On-Chip Power Savings via Bit-Value-Favor for Throughput
 Processors** 532
 Ang Li:Pacific Northwest National Laboratory; Wenfeng Zhao:University of Minnesota;
 Shuaiwen Leon Song:Pacific Northwest National Laboratory
- Xylem: Enhancing Vertical Thermal Conduction in 3D Processor-Memory Stacks** 546
 Aditya Agrawal:UIUC; Josep Torrellas:UIUC; Sachin Idgunji:Nvidia

Session 6-A: GPU-II

- Unleashing the Power of GPU for Physically Based Rendering via Dynamic Ray
 Shuffling** 560

Yashuai Lv:Astronautical Engineering University; Libo Huang:National University of Defense Technology; Li Shen:National University of Defense Technology; Zhiying Wang:National University of Defense Technology	
GPUd: A Fast and Scalable Multi-GPU Architecture Using Cooperative Projection and Distribution	574
Youngsok Kim:Seoul National University; Jae-Eon Jo:POSTECH; Hanhwi Jang:POSTECH; Minsoo Rhu:POSTECH; Hanjun Kim:POSTECH; Jangwoo Kim:Seoul National University	
VersaPipe: A Versatile Programming Framework for Pipelined Computing on GPU	587
Zhen Zheng:Tsinghua University; Chanyoung Oh:University of Seoul; Jidong Zhai:Tsinghua University; Xipeng Shen:North Carolina State University; Youngmin Yi:University of Seoul; Wenguang Chen:Tsinghua University	
WIREFRAME: Supporting Data-dependent Parallelism through Dependency Graph Execution in GPUs	600
AmirAli Abdolrashidi:University of California Riverside; Devashree Tripathy:University of California Riverside; Mehmet Esat Belviranli:Oak Ridge National Laboratories; Daniel Wong:University of California Riverside; Laxmi Narayan Bhuyan:University of California Riverside	
Session 6-B: OS and System Design	
SchedTask: A Hardware-Assisted Task Scheduler	612
Prathmesh Kallurkar:IIT Delhi; Smruti R. Sarangi:IIT Delhi	
Exploiting Heterogeneity for Tail Latency and Energy Efficiency	625
Md E. Haque:Rutgers University; Yuxiong He:Microsoft Research; Sameh Elnikety:Microsoft Research; Thu D. Nguyen:Rutgers University; Ricardo Bianchini:Microsoft Research; Kathryn S. McKinley:Google	
TMI: Thread Memory Isolation for False Sharing Repair	639
Christian DeLozier:University of Pennsylvania; Ariel Eizenberg:University of Pennsylvania; Shiliang Hu:Intel; Gilles Pokam:Intel; Joseph Devietti:University of Pennsylvania	
Estimating and Understanding Architectural Risk	651
Weilong Cui:University of California, Santa Barbara; Timothy Sherwood:University of California, Santa Barbara	
Session 7-A: Unconventional Architectures	
Hybrid Analog-Digital Solution of Nonlinear Partial Differential Equations	665
Yipeng Huang:Columbia University; Ning Guo:Columbia University; Mingoo Seok:Columbia University; Yannis Tsividis:Columbia University; Kyle Mandli:Columbia University; Simha Sethumadhavan:Columbia University	
Taming the Instruction Bandwidth of Quantum Computers via Hardware-Managed Error Correction	679
Swamit S. Tannu:Georgia Tech; Zachary A. Myers:Stanford; Prashant J. Nair:Georgia Tech; Douglas M. Carmean:Microsoft; Moinuddin K. Qureshi:Georgia Tech	
Optimized Surface Code Communication in Superconducting Quantum Computers	692
Ali Javadi-Abhari:Princeton University; Pranav Gokhale:University of Chicago; Adam Holmes:University of Chicago; Diana Franklin:University of Chicago; Kenneth R. Brown:Georgia Institute of Technology; Margaret Martonosi:Princeton University; Frederic T. Chong:University of Chicago	
Architectural Tradeoffs for Biodegradable Computing	706
Ting-Jung Chang:Princeton University; Zhuozhi Yao:Princeton University; Paul J. Jackson:Princeton University; Barry P. Rand:Princeton University; David Wentzlaff:Princeton University	

Session 7-B: Compilers and Microarchitecture

Improving the Effectiveness of Searching for Isomorphic Chains in Superword Level Parallelism	718
Joonmoo Huh:North Carolina State University; James Tuck:North Carolina State University	
Data Movement Aware Computation Partitioning	730
Xulong Tang:Penn State; Orhan Kislal:Penn State; Mahmut Kandemir:Penn State; Mustafa Karakoy:TOBB University	
Mirage Cores: The Illusion of Many Out-of-order Cores Using In-order Hardware	745
Shruti Padmanabha:University of Michigan; Andrew Lukefahr:University of Michigan; Reetuparna Das:University of Michigan; Scott Mahlke:University of Michigan	
Using Intra-Core Loop-Task Accelerators to Improve the Productivity and Performance of Task-Based Parallel Programs	759
Ji Kim: Cornell University; Shunning Jiang: Cornell University; Christopher Torng: Cornell University; Moyang Wang: Cornell University; Shreesha Srinath: Cornell University; Berkin Ilbeyi: Cornell University; Khalid Al-Hawaj: Cornell University; Christopher Batten: Cornell University	

Session 8: Best Paper Nominees

Architectural Opportunities for Novel Dynamic EMI Shifting (DEMIS)	774
Daphne I. Gorman:UC Santa Cruz; Jose Renau:UC Santa Cruz; Matthew Guthaus:UC Santa Cruz	
DeftNN: Addressing Bottlenecks for DNN Execution on GPUs via Synapse Vector Elimination and Near-compute Data Fission?	786
Parker Hill:University of Michigan; Animesh Jain:University of Michigan; Mason Hill: University of Nevada, Las Vegas; Babak Zamirai:University of Michigan; Chang-Hong Hsu:University of Michigan; Michael A. Laurenzano:University of Michigan; Scott Mahlke:University of Michigan; Lingjia Tang:University of Michigan; Jason Mars: University of Michigan	
Hardware Supported Persistent Object Address Translation	800
Tiancong Wang:NC State University; Sakthikumaran Sambasivam:NC State University; Yan Solihin:NC State University; James Tuck:NC State University	
An Experimental Microarchitecture for a Superconducting Quantum Processor .	813
X. Fu: QuTech, Delft University of Technology, Computer Engineering Lab, Delft University of Technology; M. A. Rol: QuTech, Delft University of Technology \Kavli Institute of Nanoscience, Delft University of Technology; C. C. Bultink: QuTech, Delft University of Technology \Kavli Institute of Nanoscience, Delft University of Technology; H. van Someren: QuTech, Delft University of Technology, Computer Engineering Lab, Delft University of Technology; N. Khammassi: QuTech, Delft University of Technology, Computer Engineering Lab, Delft University of Technology; I. Ashraf: QuTech, Delft University of Technology, Computer Engineering Lab, Delft University of Technology; R. F. L. Vermeulen: QuTech, Delft University of Technology \Kavli Institute of Nanoscience, Delft University of Technology; J. C. de Sterke: Topic Embedded Systems \QuTech, Delft University of Technology; W. J. Vlothuizen: Netherlands Organisation for Applied Scientific Research (TNO) \QuTech, Delft University of Technology; R. N. Schouten: QuTech, Delft University of Technology \Kavli Institute of Nanoscience, Delft University of Technology; C. G. Almudever: QuTech, Delft University of Technology, Computer Engineering Lab, Delft University of Technology; L. DiCarlo: QuTech, Delft University of Technology; Kavli Institute of Nanoscience, Delft University of Technology; K. Bertels: QuTech, Delft University of Technology, Computer Engineering Lab, Delft University of Technology	