

2017 26th International Conference on Parallel Architectures and Compilation Techniques (PACT 2017)

**Portland, Oregon, USA
9-13 September 2017**



**IEEE Catalog Number: CFP17073-POD
ISBN: 978-1-5090-6765-7**

**Copyright © 2017 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. 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:	CFP17073-POD
ISBN (Print-On-Demand):	978-1-5090-6765-7
ISBN (Online):	978-1-5090-6764-0
ISSN:	1089-795X

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

2017 26th International Conference on Parallel Architectures and Compilation Techniques

PACT 2017

Table of Contents

Welcome Message from the General Chair and Program

Chair.....	x
Organizing Committee.....	xii
Program Committee.....	xiii
Steering Committee.....	xiv
Artifact Evaluation Committee.....	xv
Reviewers.....	xvi
Keynotes.....	xix
Sponsors.....	xxii

Session 1: Algorithms and Data Structures

RCU-HTM: Combining RCU with HTM to Implement Highly Efficient Concurrent Binary Search Trees	1
<i>Dimitrios Siakavaras, Konstantinos Nikas, Georgios Goumas, and Nectarios Koziris</i>	
Redesigning Go's Built-In Map to Support Concurrent Operations	14
<i>Louis Jenkins, Tingzhe Zhou, and Michael Spear</i>	
MultiGraph: Efficient Graph Processing on GPUs	27
<i>Changwan Hong, Aravind Sukumaran-Rajam, Jinsung Kim, and P. Sadayappan</i>	
An Ultra Low-Power Hardware Accelerator for Acoustic Scoring in Speech Recognition	41
<i>Hamid Tabani, Jose-Maria Arnau, Jordi Tubella, and Antonio González</i>	

Session 2: Approximate and Speculative Computations

DrMP: Mixed Precision-Aware DRAM for High Performance Approximate and Precise Computing	53
<i>Xianwei Zhang, Youtao Zhang, Bruce R. Childers, and Jun Yang</i>	
SAM: Optimizing Multithreaded Cores for Speculative Parallelism	64
<i>Maleen Abeydeera, Suvinay Subramanian, Mark C. Jeffrey, Joel Emer, and Daniel Sanchez</i>	

Performance Improvement via Always-Abort HTM	79
<i>Joseph Izraelevitz, Lingxiang Xiang, and Michael L. Scott</i>	
DRUT: An Efficient Turbo Boost Solution via Load Balancing in Decoupled Look-Ahead Architecture	91
<i>Raj Parihar and Michael C. Huang</i>	

Session 3: Data & Emerging Use Cases

Proxy Benchmarks for Emerging Big-Data Workloads	105
<i>Reena Panda and Lizy Kurian John</i>	
Lightweight Provenance Service for High-Performance Computing	117
<i>Dong Dai, Yong Chen, Philip Carns, John Jenkins, and Robert Ross</i>	

Poster Papers

POSTER: Putting the G back into GPU/CPU Systems Research	130
<i>Andreas Sembrant, Trevor E. Carlson, Erik Hagersten, and David Black-Schaffer</i>	
POSTER: Application-Driven Near-Data Processing for Similarity Search	132
<i>Vincent T. Lee, Amrita Mazumdar, Carlo C. del Mundo, Armin Alaghi, Luis Ceze, and Mark Oskin</i>	
POSTER: Improving Datacenter Efficiency Through Partitioning-Aware Scheduling	134
<i>Harshad Kasture, Xu Ji, Nosayba El-Sayed, Nathan Beckmann, Xiaosong Ma, and Daniel Sanchez</i>	
POSTER: The Liberation Day of Nondeterministic Programs	136
<i>Enrico Armenio Deiana, Vincent St-Amour, Peter Dinda, Nikos Hardavellas, and Simone Campanoni</i>	
POSTER: Location-Aware Computation Mapping for Manycore Processors	138
<i>Orhan Kislal, Jagadish Kotra, Xulong Tang, Mahmut Taylan Kandemir, and Myoungsoo Jung</i>	
POSTER: BACM: Barrier-Aware Cache Management for Irregular Memory-Intensive GPGPU Workloads	140
<i>Yuxi Liu, Xia Zhao, Zhibin Yu, Zhenlin Wang, Xiaolin Wang, Yingwei Luo, and Lieven Eeckhout</i>	
POSTER: DaQueue: A Data-Aware Work-Queue Design for GPGPUs	142
<i>Yashuai Lü, Libo Huang, and Li Shen</i>	
POSTER: Accelerate GPU Concurrent Kernel Execution by Mitigating Memory Pipeline Stalls	144
<i>Hongwen Dai, Zhen Lin, Chao Li, Chen Zhao, Fei Wang, Nanning Zheng, and Huiyang Zhou</i>	
POSTER: Design Space Exploration for Performance Optimization of Deep Neural Networks on Shared Memory Accelerators	146
<i>Swagath Venkataramani, Jungwook Choi, Vijayalakshmi Srinivasan, Kailash Gopalakrishnan, and Leland Chang</i>	
POSTER: Bridge the Gap Between Neural Networks and Neuromorphic Hardware	148
<i>Yu Ji, YouHui Zhang, WenGuang Chen, and Yuan Xie</i>	

POSTER: Improving NUMA System Efficiency with a Utilization-Based Co-scheduling	150
<i>Younghyun Cho, Camilo A. Celis Guzman, and Bernhard Egger</i>	
POSTER: Bridging the Gap Between Deep Learning and Sparse Matrix Format Selection	152
<i>Yue Zhao, Jiajia Li, Chunhua Liao, and Xipeng Shen</i>	
POSTER: Cutting the Fat: Speeding Up RBM for Fast Deep Learning Through Generalized Redundancy Elimination	154
<i>Lin Ning, Randall Pittman, and Xipeng Shen</i>	
POSTER: Exploiting Approximations for Energy/Quality Tradeoffs in Service-Based Applications	156
<i>Liu Liu, Sibren Isaacman, Abhishek Bhattacharjee, and Ulrich Kremer</i>	
POSTER: Statement Reordering to Alleviate Register Pressure for Stencils on GPUs	158
<i>Prashant Singh Rawat, Aravind Sukumaran-Rajam, Atanas Rountev, Fabrice Rastello, Louis-Noël Pouchet, and P. Sadayappan</i>	
POSTER: NUMA-Aware Power Management for Chip Multiprocessors	160
<i>Changmin Ahn, Camilo A. Celis Guzman, and Bernhard Egger</i>	
POSTER: BigBus: A Scalable Optical Interconnect	162
<i>Eldhose Peter, Janibul Bashir, and Smruti R. Sarangi</i>	
POSTER: Elastic Reconfiguration for Heterogeneous NoCs with BiNoCHS	164
<i>Amirhossein Mirhosseini, Mohammad Sadrosadati, Behnaz Soltani, Hamid Sarbazi-Azad, and Thomas F. Wenisch</i>	

Session 4: Memory 1

Nexus: A New Approach to Replication in Distributed Shared Caches	166
<i>Po-An Tsai, Nathan Beckmann, and Daniel Sanchez</i>	
Leeway: Addressing Variability in Dead-Block Prediction for Last-Level Caches	180
<i>Priyank Faldu and Boris Grot</i>	
Application Clustering Policies to Address System Fairness with Intel's Cache Allocation Technology	194
<i>Vicent Selfa, Julio Sahuquillo, Lieven Eeckhout, Salvador Petit, and María E. Gómez</i>	
Transparent Dual Memory Compression Architecture	206
<i>Seikwon Kim, Seonyoung Lee, Taehoon Kim, and Jaehyuk Huh</i>	

Session 5: Best Papers (GPU Computing and Energy Efficiency)

End-to-End Deep Learning of Optimization Heuristics	219
<i>Chris Cummins, Pavlos Petoumenos, Zheng Wang, and Hugh Leather</i>	
Graphie: Large-Scale Asynchronous Graph Traversals on Just a GPU	233
<i>Wei Han, Daniel Mawhirter, Bo Wu, and Matthew Buland</i>	
A GPU-Friendly Skiplist Algorithm	246
<i>Nurit Moscovici, Nachshon Cohen, and Erez Petrank</i>	

Sthira: A Formal Approach to Minimize Voltage Guardbands under Variation in Networks-on-Chip for Energy Efficiency	260
<i>Raghavendra Pradyumna Pothukuchi, Amin Ansari, Bhargava Gopireddy, and Josep Torrellas</i>	

Session 6: Memory 2

Avoiding TLB Shootdowns Through Self-Invalidating TLB Entries	273
<i>Amro Awad, Arkaprava Basu, Sergey Blagodurov, Yan Solihin, and Gabriel H. Loh</i>	
Weak Memory Models: Balancing Definitional Simplicity and Implementation Flexibility	288
<i>Sizhuo Zhang, Muralidaran Vijayaraghavan, and Arvind</i>	
Near-Memory Address Translation	303
<i>Javier Picorel, Djordje Jevdjic, and Babak Falsafi</i>	
Efficient Checkpointing of Loop-Based Codes for Non-volatile Main Memory	318
<i>Hussein Elnawawy, Mohammad Alshboul, James Tuck, and Yan Solihin</i>	

Session 7: Translation

SuperGraph-SLP Auto-Vectorization	330
<i>Vasileios Porpodas</i>	
Exploiting Asymmetric SIMD Register Configurations in ARM-to-x86 Dynamic Binary Translation	343
<i>Yu-Ping Liu, Ding-Yong Hong, Jan-Jan Wu, Sheng-Yu Fu, and Wei-Chung Hsu</i>	
A Generalized Framework for Automatic Scripting Language Parallelization	356
<i>Taewook Oh, Stephen R. Beard, Nick P. Johnson, Sergiy Popovych, and David I. August</i>	

ACM SRC Presentations

Architecting a Novel Hybrid Cache with Low Energy	370
<i>Jiacong He and Joseph Callenes-Sloan</i>	
Introspective Computing	371
<i>Karl Taht and Rajeev Balasubramonian</i>	
A DSL for Performance Orchestration	372
<i>Thiago Santos Faria Xavier Teixeira, David Padua, and William Gropp</i>	
Cache Automaton: Repurposing Caches for Automata Processing	373
<i>Arun Subramaniyan, Jingcheng Wang, Ezhil R. M. Balasubramanian, David Blaauw, Dennis Sylvester, and Reetuparna Das</i>	
Large Scale Data Clustering Using Memristive k-Median Computation	374
<i>Yomi Karthik Rupesh and Mahdi Nazm Bojnordi</i>	
In-memory Data Flow Processor	375
<i>Daichi Fujiki, Scott Mahlke, and Reetuparna Das</i>	
Multilayer Compute Resource Management with Robust Control Theory	376
<i>Raghavendra Pradyumna Pothukuchi, Sweta Yamini Pothukuchi, Petros Voulgaris, and Josep Torrellas</i>	

Author Index	377
---------------------------	------------