2019 28th International Conference on Parallel Architectures and Compilation Techniques (PACT 2019)

Seattle, Washington, USA
23 – 26 September 2019

IEEE Catalog Number: CFP19073-POD
ISBN: 978-1-7281-3614-1

Copyright © 2019 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:	CFP19073-POD
ISBN (Print-On-Demand):	978-1-7281-3614-1
ISBN (Online):	978-1-7281-3613-4
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

2019 28th International Conference on Parallel Architectures and Compilation Techniques (PACT)

PACT 2019

Table of Contents

Message from the Chairs xii
Program Committee xiii
External Review Committee xiv
External Reviewers xv

Session 1: Best-Papers

MASR: A Modular Accelerator for Sparse RNNs 1
Udit Gupta (Harvard University), Brandon Reagen (Harvard University), Lillian Pentecost (Harvard University), Marco Donato (Harvard University), Thierry Tambe (Harvard University), Alexander M. Rush (Harvard University), Gu-Yeon Wei (Harvard University), and David Brooks (Harvard University)

Gluon-Async: A Bulk-Asynchronous System for Distributed and Heterogeneous Graph Analytics 15
Roshan Dathathri (University of Texas at Austin), Gurbinder Gill (University of Texas at Austin), Loc Hoang (University of Texas at Austin), Vishwesh Jatala (University of Texas at Austin), Keshav Pingali (University of Texas at Austin), V. Krishna Nandivada (Indian Institute of Technology Madras), Hoang-Vu Dang (University of Illinois at Urbana-Champaign), and Marc Snir (University of Illinois at Urbana-Champaign)

BOLT: Optimizing OpenMP Parallel Regions with User-Level Threads 29
Shintaro Iwasaki (The University of Tokyo), Abdelhalim Amer (Argonne National Laboratory), Kenjiro Taura (The University of Tokyo), Sangmin Seo (Argonne National Laboratory), and Pavan Balaji (Argonne National Laboratory)

SMT-COP: Defeating Side-Channel Attacks on Execution Units in SMT Processors 43
Daniel Townley (Binghamton University) and Dmitry Ponomarev (Binghamton University)

Session 2A: Compiler Optimization and Code Generation 1

Type-Directed Program Synthesis and Constraint Generation for Library Portability .55
Bruce Collie (University of Edinburgh), Philip Ginsbach (University of Edinburgh), and Michael F.P. O'Boyle (University of Edinburgh)

Deepframe: A Profile-Driven Compiler for Spatial Hardware Accelerators .68
Apala Guha (Simon Fraser University), Naveen Vedula (Simon Fraser University), and Arrvindh Shriraman (Simon Fraser University)

Fast Parallel Equivalence Relations in a Datalog Compiler .82
Patrick Nappa (The University of Sydney), David Zhao (The University of Sydney), Pavle Suboti (Amazon), and Bernhard Scholz (The University of Sydney)

Session 2B: Memory/Storage Systems 1

Enforcing Last-Level Cache Partitioning through Memory Virtual Channels .97
Jongwook Chung (Seoul National University), Yuhwan Ro (Samsung Electronics), Joonsung Kim (Seoul National University), Jaehyung Ahn (Samsung Electronics), Jangwoo Kim (Seoul National University), John Kim (KAIST), Jae W. Lee (Seoul National University), and Jung Ho Ahn (Seoul National University)

To Stack or Not To Stack .110
Richard Afoakwa (University of Rochester), Lejie Lu (University of Rochester), Hui Wu (University of Rochester), and Michael Huang (University of Rochester)

Enforcing Crash Consistency of Evolving Network Analytics in Non-Volatile Main Memory Systems .124
Soklong Lim (Washington State University Vancouver), Zaixin Lu (Washington State University Vancouver), Bin Ren (College of William and Mary), and Xuechen Zhang (Washington State University Vancouver)

Session 3A: Hardware/Software for Security

Fooling the Sense of Cross-Core Last-Level Cache Eviction Based Attacker by Prefetching Common Sense.138
Biswabandan Panda (Indian Institute of Technology, Kanpur)

SpecShield: Shielding Speculative Data from Microarchitectural Covert Channels .151
Kristin Barber (The Ohio State University), Anys Bacha (University of Michigan), Li Zhou (The Ohio State University), Yinqian Zhang (The Ohio State University), and Radu Teodorescu (The Ohio State University)

Session 3B: Hardware/Software for Machine Learning

MOSAIC: Heterogeneity-, Communication-, and Constraint-Aware Model Slicing and Execution for Accurate and Efficient Inference .165
Myeonggyun Han (UNIST), Jihoon Hyun (UNIST), Seongbeom Park (UNIST), Jinsu Park (UNIST), and Woongki Baek (UNIST)

Acorns: A Framework for Accelerating Deep Neural Networks with Input Sparsity .178
Xiao Dong (Chinese Academy of Sciences; University of Chinese Academy of Sciences), Lei Liu (Chinese Academy of Sciences), Peng Zhao (Chinese Academy of Sciences; University of Chinese Academy of Sciences), Guangli Li (Chinese Academy of Sciences; University of Chinese Academy of Sciences), Jiansong Li (Chinese Academy of Sciences; University of Chinese Academy of Sciences), Xueying Wang (Chinese Academy of Sciences; University of Chinese Academy of Sciences), and Xiaobing Feng (Chinese Academy of Sciences; University of Chinese Academy of Sciences)

Session 4A: Concurrency Management

Forgive-TM: Supporting Lazy Conflict Detection In Eager Hardware Transactional Memory .192
Sunjae Park (Georgia Institute of Technology), Christopher J. Hughes (Intel), and Milos Prvulovic (Georgia Institute of Technology)

Unfair Scheduling Patterns in NUMA Architectures .205
Naama Ben-David (Carnegie Mellon University), Ziv Scully (Carnegie Mellon University), and Guy E. Blelloch (Carnegie Mellon University)

Optimizing Persistent Memory Transactions .219
Pantea Zardoshti (Lehigh University), Tingzhe Zhou (Facebook), Yujie Liu (Google), and Michael Spear (Lehigh University)

Session 4B: Heterogeneous Systems and Accelerators 1

HeTM: Transactional Memory for Heterogeneous Systems .231
Daniel Castro (Universidade de Lisboa, Portugal), Paolo Romano (Universidade de Lisboa, Portugal), Aleksandar Ilic (Universidade de Lisboa, Portugal), and Amin M. Khan (UiT The Arctic University of Norway, Norway)

Achieving Scalability in a k-NN Multi-GPU Network Service with Centaur .244
Amir Watad (Technion), Alexander Libov (Amazon), Ohad Shacham (Yahoo! Labs), Edward Bortnikov (Yahoo! Labs), and Mark Silberstein (Technion)

Analyzing and Leveraging Remote-Core Bandwidth for Enhanced Performance in GPUs .257
Mohamed Assem Ibrahim (William & Mary), Hongyuan Liu (William & Mary), Onur Kayiran (Advanced Micro Devices, Inc.), and Adwait Jog (William & Mary)

Session 5A: Domain/Application-Specific Hardware/Software

Specialization Opportunities in Graphical Workloads .271
Lewis Crawford (The University of Edinburgh) and Michael O'Boyle (The University of Edinburgh)

FindeR: Accelerating FM-Index-Based Exact Pattern Matching in Genomic Sequences through ReRAM Technology .283
Farzaneh Zokaee (Indiana University), Mingzhe Zhang (ICT, CAS, China), and Lei Jiang (Indiana University)

SLAMBooster: An Application-Aware Online Controller for Approximation in Dense SLAM 295
Yan Pei (The University of Texas at Austin), Swarnendu Biswas (Indian Institute of Technology Kanpur), Donald S. Fussell (The University of Texas at Austin), and Keshav Pingali (The University of Texas at Austin)

Session 5B: Heterogeneous Systems and Accelerators 2

Exploring Memory Persistency Models for GPUs 310
Zhen Lin (North Carolina State University), Mohammad Alshboul (North Carolina State University), Yan Solihin (University of Central Florida), and Huiyang Zhou (North Carolina State University)

Adaptive Task Aggregation for High-Performance Sparse Solvers on GPUs 323
Ahmed E. Helal (Virginia Tech), Ashwin M. Aji (AMD), Michael L. Chu (AMD), Bradford M. Beckmann (AMD), and Wu-chun Feng (Virginia Tech)

EDGE: Event-Driven GPU Execution 336
Tayler Hicklin Hetherington (The University of British Columbia), Maria Lubeznov (The University of British Columbia), Deval Shah (The University of British Columbia), and Tor M. Aamodt (The University of British Columbia)

Session 6A: Compiler Optimization and Code Generation 2

Generating Portable High-Performance Code via Multi-Dimensional Homomorphisms 353
Ari Rasch (University of Münster), Richard Schulze (University of Münster), and Sergei Gorlatch (University of Münster)

Absinthe: Learning an Analytical Performance Model to Fuse and Tile Stencil Codes in One Shot 369
Tobias Gysi (ETH Zurich), Tobias Grosser (ETH Zurich), and Torsten Hoefler (ETH Zurich)

Session 6B: Memory/Storage Systems 2

Reducing Data Movement and Energy in Multilevel Cache Hierarchies without Losing Performance: Can you have it all? 382
Jiajun Wang (University of Texas at Austin), Prakash Ramrakhyani (Arm), Wendy Elsasser (Arm), and Lizy Kurian John (University of Texas at Austin)

Multiversioned Page Overlays: Enabling Faster Serializable Hardware Transactional Memory 394
Ziqi Wang (Carnegie Mellon University), Michael A. Kozuch (Intel Labs), Todd C. Mowry (Carnegie Mellon University), and Vivek Seshadri (Microsoft Research India)

Session 7: Parallel Algorithms and Applications

Computing Three-Dimensional Constrained Delaunay Refinement Using the GPU 408
Zhenghai Chen (National University of Singapore) and Tiow-Seng Tan (National University of Singapore)

A Synchronization-Avoiding Distance-1 Grundy Coloring Algorithm for Power-Law Graphs 420
Jesun Sahariar Firoz (Pacific Northwest National Laboratory), Marcin Zalewski (Pacific Northwest National Laboratory), and Andrew Lumsdaine (Pacific Northwest National Laboratory)

Accelerating DCA++ (Dynamical Cluster Approximation) Scientific Application on the Summit Supercomputer 432
Giovanni Balduzzi (ETH Zurich), Arghya Chatterjee (Oak Ridge National Laboratory), Ying Wai Li (Los Alamos National Laboratory), Peter W. Doak (Oak Ridge National Laboratory), Urs Haehner (ETH Zurich), Ed F. D'Azevedo (Oak Ridge National Laboratory), Thomas A. Maier (Oak Ridge National Laboratory), and Thomas Schulthess (ETH Zurich)

A Methodology for Characterizing Sparse Datasets and Its Application to SIMD Performance Prediction 444
Gangyi Zhu (Ohio State University), Peng Jiang (Ohio State University), and Gagan Agrawal (Ohio State University)

Posters

POSTER: Precise Capacity Planning for Database Public Clouds 456
Ningxin Zheng (Shanghai Jiao Tong University), Quan Chen (Shanghai Jiao Tong University), Yong Yang (Alibaba Cloud), Jin Li (Shanghai Jiao Tong University), Wenli Zheng (Shanghai Jiao Tong University), and Minyi Guo (Shanghai Jiao Tong University)

POSTER: BioSEAL: In-Memory Biological Sequence Alignment Accelerator for Large-Scale Genomic Data 458
Roman Kaplan (Technion, Israel Institute of Technology), Leonid Yavits (Technion, Israel Institute of Technology), and Ran Ginosar (Technion, Israel Institute of Technology)

POSTER: The Performance Impact of Thread Packing on Synchronization-Intensive Applications 460
Jinsu Park (UNIST), Seongbeom Park (UNIST), Myeonggyun Han (UNIST), and Woongki Baek (UNIST)

POSTER: Leveraging Run-Time Feedback for Efficient ASR Acceleration 462
Reza Yazdani (Universitat Politecnica de Catalunya), Jose-Maria Arnau (Universitat Politecnica de Catalunya), and Antonio González (Universitat Politecnica de Catalunya)

POSTER: Automatic Parallelization Targeting Asynchronous Task-Based Runtimes 464
Charles Jin (Reservoir Labs), Muthu Baskaran (Reservoir Labs), and Benoit Meister (Reservoir Labs)

POSTER: Memory Hotspot Optimization for Data-Intensive Applications 466
Xi Wang (Texas Tech University), Jie Li (Texas Tech University), Antonino Tumeo (Pacific Northwest National Laboratory), John D. Leidel (Tactical Computing Laboratories), and Yong Chen (Texas Tech University)

POSTER: GPU Based Near Data Processing for Image Processing with Pattern Aware Data Allocation and Prefetching 468
Jungwoo Choi (Seoul National University), Boyeal Kim (Seoul National University), Ji-Ye Jeon (Seoul National University), Hyuk-Jae Lee (Seoul National University), Euicheol Lim (SKHynix), and Chae Eun Rhee (Inha University)

POSTER: Variable Sized Cache-Block Compaction .470
Sayantan Ray (IIT Madras) and Madhu Mutyam (IIT Madras)

POSTER: Domain-Specialized Cache Management for Graph Analytics .472
Priyank Faldu (The University of Edinburgh), Jeff Diamond (Oracle Labs), and Boris Grot (The University of Edinburgh)

POSTER: Runtime Adaptations for Energy-Efficient VSLAM .474
Abdullah Khalufa (The University of Manchester), Graham Riley (The University of Manchester), and Mikel Lujan (The University of Manchester)

POSTER: GIRAF: General Purpose In-Storage Resistive Associative Framework .476
Leonid Yavits (Technion Israel Institute of Technology), Roman Kaplan (Technion Israel Institute of Technology), and Ran Ginosar (Technion Israel Institute of Technology)

POSTER: An Optimized Predication Execution for SIMD Extensions .478
Adrián Barredo (Barcelona Supercomputing Center), Juan M. Cebrián (Barcelona Supercomputing Center), Miquel Moretó (Barcelona Supercomputing Center), Marc Casas (Barcelona Supercomputing Center), and Mateo Valero (Barcelona Supercomputing Center)

POSTER: Tango: An Optimizing Compiler for Just-In-Time RTL Simulation .480
Blaise-Pascal Tine (Georgia Institute of technology), Sudhakar Yalamanchili (Georgia Institute of Technology), Hyesoon Kim (Georgia Institute of Technology), and Jeff Vetter (Oak Ridge National Laboratory)

POSTER: SPiDRE: Accelerating Sparse Memory Access Patterns .482
Adrián Barredo (Barcelona Supercomputing Center), Jonathan C. Beard (Arm Research), and Miquel Moretó (Barcelona Supercomputing Center)

POSTER: CogR: Exploiting Program Structures for Machine-Learning Based Runtime Solutions .484
Hyojin Sung (Pohang University of Science and Technology), Tong Chen (IBM Research), Alenxandre Eichenberger (IBM Research), and Kevin K. O'Brien (IBM Research)

POSTER: A Collaborative Multi-Factor Scheduler for Asymmetric Multicore Processors .486
Teng Yu (University of St Andrews), Pavlos Petoumenos (University of Edinburgh), Vladimir Janjic (University of St Andrews), Mingcan Zhu (University of St Andrews), Hugh Leather (University of Edinburgh), and John Thomson (University of St Andrews)

POSTER: Space and Time Optimal DNN Primitive Selection with Integer Linear Programming .488
Yuan Wen (Trinity College Dublin), Andrew Anderson (Trinity College Dublin), Valentin Radu (The University of Edinburgh), Michael F.P. O'Boyle (The University of Edinburgh), and David Gregg (Trinity College Dublin)

POSTER: Quiescent and Versioned Shadow Copies for NVM .490
Zhenwei Wu (National University of Defense Technology; University of Manchester), Kai Lu (National University of Defense Technology), Wenzhe Zhang (National University of Defense Technology), Andrew Nisbet (University of Manchester), and Mikel Luján (University of Manchester)

POSTER: AR-MMAP: Write Performance Improvement of Memory-Mapped File 492
Satoshi Imamura (Fujitsu Laboratories Ltd.) and Eiji Yoshida (Fujitsu Laboratories Ltd.)

POSTER: Exploiting Multi-Level Task Dependencies to Prune Redundant Work in Relax-Ordered Task-Parallel Algorithms 494
Masab Ahmad (University of Connecticut), Mohsin Shan (University of Connecticut), Akif Rehman (University of Connecticut), and Omer Khan (University of Connecticut)

POSTER: Quantifying the Direct Overhead of Virtual Function Calls on Massively Parallel Architectures 496
Mengchi Zhang (Purdue University), Roland N. Green (Purdue University), and Timothy G. Rogers (Purdue University)

POSTER: A Polyhedral+Dataflow Intermediate Language for Performance Exploration 498
Eddie C. Davis (Boise State University) and Catherine RM. Olschanowsky (Boise State University)

POSTER: Pairing Up CNNs for High Throughput Deep Learning 500
Babak Zamirai (University of Michigan), Salar Latifi (University of Michigan), and Scott Mahlke (University of Michigan)

POSTER: A Memory-Access-Efficient Adaptive Implementation of kNN on FPGA through HLS 502
Xiaojia Song (San Diego State university), Tao Xie (San Diego State University), and Stephen Fischer (Samsung Semiconductor, Inc.)

Author Index 505