

2023 32nd International Conference on Parallel Architectures and Compilation Techniques (PACT 2023)

**Vienna, Austria
21-25 October 2023**



**IEEE Catalog Number: CFP23073-POD
ISBN: 979-8-3503-4255-0**

**Copyright © 2023 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:	CFP23073-POD
ISBN (Print-On-Demand):	979-8-3503-4255-0
ISBN (Online):	979-8-3503-4254-3

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

2023 32nd International Conference on Parallel Architectures and Compilation Techniques (PACT) **PACT 2023**

Table of Contents

Welcome Message	x
Conference Organization	xi
Keynotes	xiii
Sponsors and Supporters	xvi

Compilers

CELLO: Compiler-Assisted Efficient Load-Load Ordering in Data-Race-Free Regions	1
<i>Sawan Singh (University of Murcia, Spain), Josue Feliu (Universitat Politècnica de València), Manuel E. Acacio (University of Murcia, Spain), Alexandra Jimborean (University of Murcia, Spain), and Alberto Ros (University of Murcia, Spain)</i>	
Automatic Code Generation for High-Performance Graph Algorithms	14
<i>Zhen Peng (Pacific Northwest National Laboratory, USA), Rizwan A. Ashraf (Pacific Northwest National Laboratory, USA), Luanzheng Guo (Pacific Northwest National Laboratory, USA), Ruiqin Tian (Horizon Robotics, China; University of California, USA), and Gokcen Kestor (Pacific Northwest National Laboratory, USA; University of California, USA)</i>	
UWomp_pro : UWOMP++ with Point-to-Point Synchronization, Reduction and Schedules	27
<i>Aditya Agrawal (IIT Madras, India) and V. Krishna Nandivada (IIT Madras, India)</i>	
mlirSynth: Automatic, Retargetable Program Raising in Multi-Level IR using Program Synthesis	39
<i>Alexander Brauckmann (University of Edinburgh, United Kingdom), Elizabeth Polgreen (University of Edinburgh, United Kingdom), Tobias Grosser (University of Edinburgh, United Kingdom), and Michael F. P. O'Boyle (University of Edinburgh, United Kingdom)</i>	

Memory System

Drishyam: An Image is Worth a Data Prefetcher	51
<i>Shubdeep Mohapatra (BITS Pilani, India) and Biswabandan Panda (Indian Institute of Technology Bombay, India)</i>	

HugeGPT: Storing Guest Page Tables on Host Huge Pages to Accelerate Address Translation	62
<i>Weiwei Jia (The University of Rhode Island), Jiyuan Zhang (University of Illinois at Urbana-Champaign), Jianchen Shan (Hofstra University), Yiming Du (The University of Rhode Island), Xiaoning Ding (New Jersey Institute of Technology), and Tianyin Xu (University of Illinois at Urbana-Champaign)</i>	
PreFlush: Lightweight Hardware Prediction Mechanism for Cache Line Flush and Writeback	74
<i>Hussein Elnawawy (North Carolina State University, USA), James Tuck (North Carolina State University, USA), and Gregory T. Byrd (North Carolina State University, USA)</i>	
SDM: Sharing-Enabled Disaggregated Memory System with Cache Coherent Compute Express Link ..	86
<i>Hyokeun Lee (North Carolina State University), Kwansoek Choi (Seoul National University), Hyuk-Jae Lee (Seoul National University), and Jaewoong Sim (Seoul National University)</i>	
SimplePIM: A Software Framework For Productive And Efficient In-Memory Processing	99
<i>Jinfan Chen (ETH Zürich), Juan Gómez-Luna (ETH Zürich), Izzat El Hajj (American University of Beirut), Yuxin Guo (ETH Zürich), and Onur Mutlu (ETH Zürich)</i>	
Virtual PIM: Resource-Aware Dynamic DPU Allocation and Workload Scheduling Framework for Multi-DPU PIM Architecture	112
<i>Donghyeon Kim (Hanyang University), Taehoon Kim (Hanyang University), Inyong Hwang (Yonsei University), Taehyeong Park (Yonsei University), Hanjun Kim (Yonsei University), Youngsok Kim (Yonsei University), and Yongjun Park (Yonsei University)</i>	

GPUs

Boustrophedonic Frames: Quasi-Optimal L2 Caching for Textures in GPUs	124
<i>Diya Joseph (Universitat Politècnica de Catalunya, Spain), Juan L. Aragón (Universidad de Murcia, Spain), Joan-Manuel Parcerisa (Universitat Politècnica de Catalunya, Spain), and Antonio González (Universitat Politècnica de Catalunya, Spain)</i>	
G-Sparse: Compiler-Driven Acceleration for Generalized Sparse Computation for Graph Neural Networks on Modern GPUs	137
<i>Yue Jin (Ant Group), Chengying Huan (Chinese Academy of Sciences), Heng Zhang (Chinese Academy of Sciences), Yongchao Liu (Ant Group), Shuaiwen Leon Song (University of Sydney), Rui Zhao (Ant Group), Yao Zhang (Microsoft), Changhua He (Dipeak Ltd), and Wenguang Chen (Ant Group)</i>	
TSUNAMI: A GPU Implementation of the WFA Algorithm	150
<i>Giulia Gerometta (Informatica e Bioingegneria, Politecnico di Milano, Italy), Alberto Zeni (Informatica e Bioingegneria, Politecnico di Milano, Italy), and Marco D. Santambrogio (Informatica e Bioingegneria, Politecnico di Milano, Italy)</i>	

Parallelizing Maximal Clique Enumeration on GPUs	162
<i>Mohammad Almasri (ECE), Yen-Hsiang Chang (ECE), Izzat El Hajj (American University of Beirut, Lebanon), Rakesh Nagi (ISE, University of Illinois at Urbana-Champaign, USA), Jinjun Xiong (University at Buffalo, USA), and Wen-Mei Hwu (ECE; Nvidia Corporation, USA)</i>	

Algorithms

Accelerating Decision-Tree-based Inference through Adaptive Parallelization	176
<i>Jan van Lunteren (IBM Research Europe, Switzerland)</i>	
Automatic Algorithm-Based Fault Tolerance (AABFT) of Stencil Computations	187
<i>Louis Narmour (Univ Rennes, Inria, CNRS, IRISA, Colorado State University, USA), Steven Derrien (Univ Rennes, Inria, CNRS, IRISA, France), and Sanjay Rajopadhye (Colorado State University, USA)</i>	
Performance Characterization of Popular DNN Models on Out-of-Order CPUs	199
<i>Pablo Prieto (Universidad de Cantabria, Spain), Pablo Abad (Universidad de Cantabria, Spain), Jose Angel Gregorio (Universidad de Cantabria, Spain), and Valentin Puente (Universidad de Cantabria, Spain)</i>	
GraphMini: Accelerating Graph Pattern Matching Using Auxiliary Graphs	211
<i>Juelin Liu (University of Massachusetts Amherst, USA), Sandeep Polisetty (University of Massachusetts Amherst, USA), Hui Guan (University of Massachusetts Amherst, USA), and Marco Serafini (University of Massachusetts Amherst, USA)</i>	

Architecture

Barad-dur: Near-Storage Accelerator for Training Large Graph Neural Networks	225
<i>Jiyoung An (University of California, Irvine, USA), Esmerald Aliaj (University of California, Irvine, USA), and Sang-Woo Jun (University of California, Irvine, USA)</i>	
A Silicon Photonic Multi-DNN Accelerator	238
<i>Yuan Li (George Washington University, USA), Ahmed Louri (George Washington University, USA), and Avinash Karanth (Ohio University, USA)</i>	
Architecture-Aware Currying	250
<i>Mahmut Taylan Kandemir (Pennsylvania State University, USA), Gulsum Gudukbay Akbulut (Pennsylvania State University, USA), Wonil Choi (Hanyang University, Republic of Korea), and Mustafa Karakoy (TUBITAK-BILGEM, Turkey)</i>	
SpecCheck: A Tool for Systematic Identification of Vulnerable Transient Execution in gem5	265
<i>Zack McKeivitt (University of Colorado Boulder, USA), Ashutosh Trivedi (University of Colorado Boulder, USA), and Tamara Silbergleit Lehman (University of Colorado Boulder, USA)</i>	

Optimization

Separating Mechanism from Policy in STM	279
<i>Yaodong Sheng (Lehigh University, USA), Ahmed Hassan (Lehigh University, USA), and Michael Spear (Lehigh University, USA)</i>	
MBAPIS: Multi-Level Behavior Analysis Guided Program Interval Selection for Microarchitecture Studies	297
<i>Hongwei Cui (Peking University, China), Yujie Cui (Peking University, China), Honglan Zhan (Peking University, China), Shuhao Liang (Peking University, China), Xianhua Liu (Peking University, China), Chun Yang (Peking University, China), and Xu Cheng (Peking University, China)</i>	
INTERPRET: Inter-Warp Register Reuse for GPU Tensor Core	309
<i>Jae Seok Kwak (Yonsei University, Korea), Myung Kuk Yoon (Ewha Womans University, Korea), Ipoom Jeong (University of Illinois Urbana-Champaign, USA), Seunghyun Jin (Yonsei University, Korea), and Won Woo Ro (Yonsei University, Korea)</i>	

Posters

A CPU-FPGA Holistic Source-To-Source Compilation Approach for Partitioning and Optimizing C/C++ Applications	320
<i>Tiago Santos (Univ. of Porto, FEUP and INESC-TEC, Portugal), João Bispo (Univ. of Porto, FEUP and INESC-TEC, Portugal), and João M. P. Cardoso (Univ. of Porto, FEUP and INESC-TEC, Portugal)</i>	
Dynamic Allocation of Processor Cores to Graph Applications on Commodity Servers	323
<i>Lucia Pons (Universitat Politècnica de València, Spain), Julio Sahuquillo (Universitat Politècnica de València, Spain), and Timothy M. Jones (University of Cambridge, United Kingdom)</i>	
QeiHaN: An Energy-Efficient DNN Accelerator that Leverages Log Quantization in NDP Architectures	325
<i>Bahareh Khabbazan (Universitat Politècnica de Catalunya (UPC), Spain), Marc Riera (Universitat Politècnica de Catalunya (UPC), Spain), and Antonio González (Universitat Politècnica de Catalunya (UPC), Spain)</i>	
Quickloop: An Efficient, FPGA-Accelerated Exploration of Parameterized DNN Accelerators	327
<i>Tayyeb Mahmood (Incheon National University, South Korea), Kashif Inayat (Barcelona Supercomputing Center, Spain), and Jaeyong Chung (Incheon National University, South Korea)</i>	
Retargeting Applications for Heterogeneous Systems with the Tribble Source-to-Source Framework	329
<i>Luís Miguel Sousa (University of Porto & INESC TEC, Portugal), João Bispo (University of Porto & INESC TEC, Portugal), and Nuno Paulino (University of Porto & INESC TEC, Portugal)</i>	
SLIDEX: Sliding Window Extension for Image Processing	332
<i>Raúl Taranco (Universitat Politècnica de Catalunya, Spain), José-María Arnau (Universitat Politècnica de Catalunya, Spain), and Antonio González (Universitat Politècnica de Catalunya, Spain)</i>	

Thread-to-Core Allocation in ARM Processors Building Synergistic Pairs	335
<i>Marta Navarro (Universitat Politècnica de València, Spain), Josué Feliu (Universitat Politècnica de València, Spain), Salvador Petit (Universitat Politècnica de València, Spain), María E. Gómez (Universitat Politècnica de València, Spain), and Julio Sahuquillo (Universitat Politècnica de València, Spain)</i>	
SparseFT: Sparsity-Aware Fault Tolerance for Reliable CNN Inference on GPUs	337
<i>Gwangeun Byeon (Sungkyunkwan University), Seungtae Lee (Sungkyunkwan University), Seongwook Kim (Sungkyunkwan University), Yongjun Kim (Samsung Electronics), Prashant J. Nair (The University of British Columbia), and Seokin Hong (Sungkyunkwan University)</i>	
Author Index	339