

# **2023 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2023)**

**Raleigh, North Carolina, USA  
23 – 25 April 2023**



**IEEE Catalog Number: CFP23PER-POD  
ISBN: 979-8-3503-9740-6**

**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:	CFP23PER-POD
ISBN (Print-On-Demand):	979-8-3503-9740-6
ISBN (Online):	979-8-3503-9739-0

**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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2023 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) **ISPASS 2023**

## Table of Contents

Message from the General Chair .....	xi
Organizing Committee .....	xii
Program Committee .....	xiii
Steering Committee .....	xiv
Sponsors .....	xv

### Best Paper Session

Characterization of Data Compression in Datacenters .....	1
<i>Geonhwa Jeong (Georgia Institute of Technology, USA), Bikash Sharma (Meta, USA), Nick Terrell (Meta, USA), Abhishek Dhanotia (Meta, USA), Zhiwei Zhao (Meta, USA), Niket Agarwal (NVIDIA, USA), Arun Kejariwal (Meta, USA), and Tushar Krishna (Georgia Institute of Technology, USA)</i>	
PES: An Energy and Throughput Model for Energy Harvesting IoT Systems .....	13
<i>Fatemeh Ghasemi (Norwegian University of Science and Technology (NTNU), Norway), Lukas Liedtke (Norwegian University of Science and Technology (NTNU), Norway), and Magnus Jahre (Norwegian University of Science and Technology (NTNU), Norway)</i>	
PyTFHE: An End-to-End Compilation and Execution Framework for Fully Homomorphic Encryption Applications .....	24
<i>Jiaao Ma (Duke University, USA), Ceyu Xu (Duke University, USA), and Lisa Wu Wills (Duke University, USA)</i>	
Evaluating Machine Learning Workloads on Memory-Centric Computing Systems .....	35
<i>Juan Gomez Luna (ETH Zürich), Yuxin Guo (ETH Zürich), Sylvan Brocard (UPMEM), Julien Legriel (UPMEM), Remy Cimadomo (UPMEM), Geraldo F. Oliveira (ETH Zürich), Gagandeep Singh (ETH Zürich), and Onur Mutlu (ETH Zürich)</i>	
ML: ML-Assisted Queuing Latency Analysis for Data Center Networks .....	50
<i>Shruti Yadav Narayana (University of Wisconsin-Madison, USA), Jie Tong (University of Wisconsin-Madison, USA), Anish Krishnakumar (University of Wisconsin-Madison, USA), Nuriye Yildirim (University of Wisconsin-Madison, USA), Emily Shriver (Intel Labs, USA), Mahesh Ketkar (Intel Labs, USA), and Umit Y. Ogras (University of Wisconsin-Madison, USA)</i>	

## CPU Microarchitecture

A Characterization of the Effects of Software Instruction Prefetching on an Aggressive Front-end .....	61
<i>Gino Chacon (Texas A&amp;M University, USA), Nathan Gober (Texas A&amp;M University, USA), Krishnendra Nathella (Arm, USA), Paul V. Gratz (Texas A&amp;M University, USA), and Daniel A. Jiménez (Texas A&amp;M University, USA)</i>	
MBPlib: Modular Branch Prediction Library .....	71
<i>Emilio Domínguez-Sánchez (University of Murcia, Spain) and Alberto Ros (University of Murcia, Spain)</i>	
Evaluating the Impact of Optimizations for Dynamic Binary Modification on 64-bit RISC-V .....	81
<i>John Alistair Kressel (University of Manchester, United Kingdom), Guillermo Callaghan (University of Manchester, United Kingdom), Cosmin Gorgovan (University of Manchester, United Kingdom), and Mikel Luján (University of Manchester, United Kingdom)</i>	
An Application-Oriented Approach to Designing Hybrid CPU Architectures .....	92
<i>Anna Yue (Hewlett-Packard Enterprise, USA) and Sanyam Mehta (Hewlett-Packard Enterprise, USA)</i>	

## Exploring Simulations

Profiling gem5 Simulator .....	103
<i>Johnson Umeike (University of Kansas), Neel Patel (University of Kansas), Alex Manley (University of Kansas), Amin Mamandipoor (University of Kansas), Heechul Yun (University of Kansas), and Mohammad Alian (University of Kansas)</i>	
A Novel Simulation Methodology for Silicon Photonic Switching Fabrics .....	114
<i>Markos Kynigos (University of Manchester, United Kingdom), Javier Navaridas (University of the Basque Country, Spain), Jose Pascual (University of the Basque Country, Spain), and Mikel Lujan (University of Manchester, United Kingdom)</i>	
Simulating Wrong-Path Instructions in Decoupled Functional-First Simulation .....	124
<i>Stijn Eyerman (Intel, Belgium), Sam Van den Steen (Intel, Belgium), Wim Heirman (Intel, Belgium), and Ibrahim Hur (Intel, USA)</i>	
Is the Future Cold or Tall? Design Space Exploration of Cryogenic and 3D Embedded Cache Memory .....	134
<i>Alexander Hankin (Harvard University, USA; Intel Labs, USA), Lillian Pentecost (Amherst College, USA), Dongmoon Min (Seoul National University, Republic of Korea), David Brooks (Harvard University, USA), and Gu-Yeon Wei (Harvard University, USA)</i>	

# Machine Learning 1

MergePath-SpMM: Parallel Sparse Matrix-Matrix Algorithm for Graph Neural Network Acceleration .....	145
<i>Mohsin Shan (University of Connecticut, USA), Deniz Gurevin (University of Connecticut, USA), Jared Nye (University of Connecticut, USA), Caiwen Ding (University of Connecticut, USA), and Omer Khan (University of Connecticut, USA)</i>	
CFU Playground: Full-Stack Open-Source Framework for Tiny Machine Learning (TinyML) Acceleration on FPGAs .....	157
<i>Shvetank Prakash (Harvard University, USA), Tim Callahan (Google, USA), Joseph Bushagour (Purdue University, USA), Colby Banbury (Harvard University, USA), Alan V. Green (Google, Australia), Pete Warden (Stanford University, USA), Tim Ansell (Google, USA), and Vijay Janapa Reddi (Harvard University, USA)</i>	
Characterizing the Scalability of Graph Convolutional Networks on Intel® PIUMA .....	168
<i>Matthew Adiletta (Harvard University; Intel Corporation), Jesmin Jahan Tithi (Intel Corporation), Emmanouil-Ioannis Farsarakis (Intel Corporation), Gerasimos Gerogiannis (University of Illinois at Urbana-Champaign; Intel Corporation), Robert Adolf (Harvard University; Intel Corporation), Robert Benke (Intel Corporation), Sidharth Kashyap (Intel Corporation), Samuel Hsia (Harvard University), Kartik Lakhota (Intel Corporation), Fabrizio Petrini (Intel Corporation), Gu-Yeon Wei (Harvard University), and David Brooks (Harvard University)</i>	

# Exploring Workloads

Genomics-GPU: A Benchmark Suite for GPU-Accelerated Genome Analysis .....	178
<i>Zhuren Liu (University of North Texas, USA), Shouzhe Zhang (University of North Texas, USA), Justin Garrigus (University of North Texas, USA), and Hui Zhao (University of North Texas, USA)</i>	
Exploring the Efficiency of Data-Oblivious Programs .....	189
<i>Lauren Biernacki (University of Michigan), Biniyam Mengist Tiruye (University of Michigan), Meron Zerihun Demissie (University of Michigan), Fitsum Assamnew Andargie (Addis Ababa University), Brandon Reagen (New York University), and Todd Austin (University of Michigan)</i>	
Redwood: Flexible and Portable Heterogeneous Tree Traversal Workloads .....	201
<i>Yanwen Xu (University of California, Santa Cruz), Ang Li (Princeton University), and Tyler Sorensen (University of California, Santa Cruz)</i>	
Community-Based Matrix Reordering for Sparse Linear Algebra Optimization .....	214
<i>Vignesh Balaji (NVIDIA), Neal C. Crago (NVIDIA), Aamer Jaleel (NVIDIA), and Stephen W. Keckler (NVIDIA)</i>	

## Accelerators

Sieve: Stratified GPU-Compute Workload Sampling .....	224
<i>Mahmood Naderan-Tahan (Ghent University, Belgium), Hossein SeyyedAghaei (Ghent University, Belgium), and Lieven Eeckhout (Ghent University, Belgium)</i>	
TransPimLib: Efficient Transcendental Functions for Processing-in-Memory Systems .....	235
<i>Maurus Item (ETH Zürich), Juan Gomez Luna (ETH Zürich), Yuxin Guo (ETH Zürich), Geraldo F. Oliveira (ETH Zürich), Mohammad Sadrosadati (ETH Zürich), and Onur Mutlu (ETH Zürich)</i>	
Early-Adaptor: An Adaptive Framework for Proactive UVM Memory Management .....	248
<i>Seokjin Go (Yonsei University, South Korea), Hyunwuk Lee (Yonsei University, South Korea), Junsung Kim (Yonsei University, South Korea), Jiwon Lee (Yonsei University, South Korea), Myung Kuk Yoon (Ewha Womans University, South Korea), and Won Woo Ro (Yonsei University, South Korea)</i>	
Sunstone: A Scalable and Versatile Scheduler for Mapping Tensor Algebra on Spatial Accelerators .....	259
<i>MohammadHossein Olyaiy (The University of British Columbia, Canada), Christopher Ng (The University of British Columbia, Canada), Alexandra Fedorova (The University of British Columbia, Canada), and Mieszko Lis (The University of British Columbia, Canada)</i>	
RPU: The Ring Processing Unit .....	272
<i>Deepraj Soni (New York University), Negar Neda (New York University), Naifeng Zhang (Carnegie Mellon University), Benedict Reynwar (USC Information Sciences Institute), Homer Gamil (New York University Abu Dhabi), Benjamin Heyman (New York University), Mohammed Nabeel (New York University Abu Dhabi), Ahmad Al Badawi (Duality Technologies), Yuriy Polyakov (Duality Technologies), Kellie Canida (USC Information Sciences Institute), Massoud Pedram (USC Viterbi School of Engineering), Michail Maniatakos (New York University Abu Dhabi), David Bruce Cousins (Duality Technologies), Franz Franchetti (Carnegie Mellon University), Matthew French (USC Information Sciences Institute), Andrew Schmidt (USC Information Sciences Institute), and Brandon Reagen (New York University)</i>	

## Machine Learning 2

ASTRA-sim2.0: Modeling Hierarchical Networks and Disaggregated Systems for Large-Model Training at Scale .....	283
<i>William Won (Georgia Institute of Technology, USA), Taekyung Heo (Georgia Institute of Technology, USA), Saeed Rashidi (Georgia Institute of Technology, USA), Srinivas Sridharan (Meta, USA), Sudarshan Srinivasan (Intel, India), and Tushar Krishna (Georgia Institute of Technology, USA)</i>	

Boreas: A Cost-Effective Mitigation Method for Advanced Hotspots Using Machine Learning and Hardware Telemetry .....	295
<i>Maziar Amiraski (Tufts University, USA), David Werner (Tufts University, USA), Alexander Hankin (Harvard University, USA; Intel Labs, USA), Julien Sebot (Intel Corp., USA), Kaushik Vaidyanathan (Google Inc., USA), and Mark Hempstead (Tufts University, USA)</i>	

AMPeD: An Analytical Model for Performance in Distributed Training of Transformers .....	306
<i>Diksha Moolchandani (Interuniversity Microelectronics Centre, Belgium), Joyjit Kundu (Interuniversity Microelectronics Centre, Belgium), Frederik Ruelens (Interuniversity Microelectronics Centre, Belgium), Peter Vrancx (Interuniversity Microelectronics Centre, Belgium), Timon Evenblij (Interuniversity Microelectronics Centre, Belgium), and Manu Perumkunnil (Interuniversity Microelectronics Centre, Belgium)</i>	

## Poster Session

LoopTree: Enabling Exploration of Fused-Layer Dataflow Accelerators .....	316
<i>Michael Gilbert (MIT, USA), Yannan Nellie Wu (MIT, USA), Angshuman Parashar (NVIDIA, USA), Vivienne Sze (MIT, USA), and Joel S. Emer (MIT, USA; NVIDIA, USA)</i>	

Degree-Aware Kernel Mapping for Graph Processing on GPUs .....	319
<i>Sanya Srivastava (University of California, Santa Cruz) and Tyler Sorensen (University of California, Santa Cruz)</i>	

lfbench: A Lock-Free Microbenchmark Suite .....	322
<i>Mahita Nagabhiru (North Carolina State University, USA) and Greg Byrd (North Carolina State University, USA)</i>	

A Benchmark Suite for Improving Performance Portability of the SYCL Programming Model .....	325
<i>Zheming Jin (Oak Ridge National Laboratory) and Jeffrey Vetter (Oak Ridge National Laboratory)</i>	

Impact of Optimal Design Point on Performance Metrics of DNN Accelerators in FPGA .....	328
<i>Tom Glint (IIT Gandhinagar, India), Aryan Gupta (IIT Gandhinagar, India), Daniel Giftson (IIT Gandhinagar, India), Gaurav Shah (IIT Gandhinagar, India), Vrajesh Patel (IIT Gandhinagar, India), Ruchit Chudasama (IIT Gandhinagar, India), Sukanya More (IIT Gandhinagar, India), and Joyce Meki (IIT Gandhinagar, India)</i>	

Workload Characterization Using Hierarchical PC .....	331
<i>Lina Sawalha (Western Michigan University) and Grant Deljevic (Western Michigan University)</i>	

Analyzing Energy Efficiency of a Server with a SmartNIC Under SLO Constraints .....	334
<i>Jinghan Huang (University of Illinois at Urbana-Champaign), Jiaqi Lou (University of Illinois at Urbana-Champaign), Yan Sun (University of Illinois at Urbana-Champaign), Tianchen Wang (University of Illinois at Urbana-Champaign), Eun Kyung Lee (IBM Research), and Nam Sung Kim (University of Illinois at Urbana-Champaign)</i>	

KORDI: A Framework for Real-Time Performance and Cost Optimization of Apache Spark Streaming .....	337
<i>Athanasios Kordelas (Vrije Universiteit Brussel, Belgium; University of Patras, Greece), Thanasis Spyrou (n/a), Spyros Voulgaris (Athens University of Economics and Business, Greece), Vasileios Megalooikonomou (University of Patras, Greece), and Nikos Deligiannis (Vrije Universiteit Brussel, Belgium)</i>	
Enabling Design Space Exploration of DRAM Caches for Emerging Memory Systems .....	340
<i>Maryam Babaie (University of California, USA), Ayaz Akram (University of California, USA), and Jason Lowe-Power (University of California, USA)</i>	
A Regression-Based Model for End-to-End Latency Prediction for DNN Execution on GPUs .....	343
<i>Ying Li (William &amp; Mary, USA), Yifan Sun (William &amp; Mary, USA), and Adwait Jog (William &amp; Mary, USA; University of Virginia, USA)</i>	
A Survey and Comparison of Consistent Hashing Algorithms .....	346
<i>Massimo Coluzzi (University of Applied Sciences and Arts of Southern Switzerland, Switzerland), Amos Brocco (University of Applied Sciences and Arts of Southern Switzerland, Switzerland), Patrizio Contu (Natzka SA, Switzerland), and Tiziano Leidi (University of Applied Sciences and Arts of Southern Switzerland, Switzerland)</i>	
Analysis of Conventional, Near-Memory, and In-Memory DNN Accelerators .....	349
<i>Tom Glint (IIT Gandhinagar, India), Chandan Kumar Jha (DFKI, Germany), Manu Awasthi (Ashoka University, India), and Joycee Mekie (IIT Gandhinagar, India)</i>	
RAINBOW: Multi-Dimensional Hardware-Software Co-Design for DL Accelerator On-Chip Memory .	352
<i>Stavroula Zouzoula (Chalmers University of Technology), Muhammad Waqar Azhar (Chalmers University of Technology), and Pedro Trancoso (Chalmers University of Technology)</i>	
Stream: A Modeling Framework for Fine-Grained Layer Fusion on Multi-Core DNN Accelerators .	355
<i>Arne Symons (KU Leuven, Belgium), Linyan Mei (KU Leuven, Belgium), Steven Coleman (KU Leuven, Belgium), Pouya Houshmand (KU Leuven, Belgium), Sebastian Karl (TU Munich, Germany; KU Leuven, Belgium), and Marian Verhelst (KU Leuven, Belgium)</i>	
<b>Author Index .....</b>	<b>359</b>