

2024 IEEE 36th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2024)

**Hilo, Hawaii, USA
13-15 November 2024**



**IEEE Catalog Number: CFP24307-POD
ISBN: 979-8-3503-5617-5**

**Copyright © 2024 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:	CFP24307-POD
ISBN (Print-On-Demand):	979-8-3503-5617-5
ISBN (Online):	979-8-3503-5616-8
ISSN:	1550-6533

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

2024 IEEE 36th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD) **SBAC-PAD 2024**

Table of Contents

Message from General Chairs	ix
Message from Program Chairs	x
Organizing Committee	xi
Program Committee	xii

Session #1

TangramFP: Energy-Efficient, Bit-Parallel, Multiply-Accumulate for Deep Neural Networks 1 <i>Yuan Yao (Uppsala University, Sweden), Xiaoyue Chen (Uppsala University, Sweden), Hannah Atmer (Uppsala University, Sweden), and Stefanos Kaxiras (Uppsala University, Sweden)</i>	1
DYAD: Locality-Aware Data Management for Accelerating Deep Learning Training 13 <i>Hariharan Devarajan (Lawrence Livermore National Laboratory), Ian Lumsden (University of Tennessee, Knoxville), Chen Wang (Lawrence Livermore National Laboratory), Konstantia Georgouli (Lawrence Livermore National Laboratory), Tom Scogland (Lawrence Livermore National Laboratory), Jae-Seung Yeom (Lawrence Livermore National Laboratory), and Michela Taufer (University of Tennessee, Knoxville)</i>	13
Janus: A Simple and Efficient Speculative Defense using Reinforcement Learning 25 <i>Pavlos Aimoniotis (Uppsala University, Sweden) and Stefanos Kaxiras (Uppsala University, Sweden)</i>	25
Exploiting Processor Heterogeneity to Improve Throughput and Reduce Latency for Deep Neural Network Inference 37 <i>Olivier Beaumont (Inria Center of the University of Bordeaux, France), Jean-François David (Inria Center of the University of Bordeaux, France), Lionel Eyraud-Dubois (Inria Center of the University of Bordeaux, France), and Samuel Thibault (University of Bordeaux, France)</i>	37

DeepWave: A Software Stack for Parallelizing Deep Learning Models Used in Geophysics	49
<i>Allan Pinto (Brazilian Center for Research in Energy and Materials (CNPEM), Brazil; University of Campinas (UNICAMP), Brazil), Gustavo Leite (University of Campinas (UNICAMP), Brazil), Marcio Pereira (University of Campinas (UNICAMP), Brazil), Hervé Yvoiquel (University of Campinas (UNICAMP), Brazil), Sandro Rigo (University of Campinas (UNICAMP), Brazil), and Guido Araujo (University of Campinas (UNICAMP), Brazil)</i>	

Session #2

A New Level-Set Analysis and Sparse Storage Format for the SPTRSV in GPUs	59
<i>Manuel Freire (n/a), Ernesto Dufrechou (n/a), and Pablo Ezzatti (n/a)</i>	
A Performance and Energy Study of GPU-Resident Preconditioners for Conjugate Gradient Solvers: In the Context of Existing and Novel Approaches	70
<i>Kasia Świrzydowicz (Advanced Micro Devices, Inc., USA), Jesun Firoz (Pacific Northwest National Laboratory, USA), Joseph Manzano (Pacific Northwest National Laboratory, USA), Mahantesh Halappanavar (Pacific Northwest National Laboratory, USA), Stephen Thomas (Advanced Micro Devices, Inc., USA), and Kevin Barker (Pacific Northwest National Laboratory, USA)</i>	

Session #3

TailorFS: An Adaptive File System to Support Dynamic I/O Requirements of HPC Workloads	81
<i>Hariharan Devarajan (Lawrence Livermore National Laboratory) and Kathryn Mohror (Lawrence Livermore National Laboratory)</i>	
High-Quality I/O Bandwidth Prediction with Minimal Data via Transfer Learning Workflow	93
<i>Dmytro Povaliaiev (RWTH Aachen University, Germany), Radita Liem (RWTH Aachen University, Germany), Julian Kunkel (Georg-August Universität Göttingen/GWDG, Germany), Jay Lofstead (Sandia National Laboratory, USA), and Philip Carns (Argonne National Laboratory, USA)</i>	
To Derive or Not to Derive: I/O Libraries Take Charge of Derived Quantities Computation	105
<i>Ana Gainaru (Oak Ridge National Laboratory), Greg Eisenhauer (Georgia Institute of Technology), Norbert Podhorszki (Oak Ridge National Laboratory), Liz Dulac (Oak Ridge National Laboratory), Qian Gong (Oak Ridge National Laboratory), Anthony Kougkas (Illinois Institute of Technology), Xian-He Sun (Illinois Institute of Technology), Gerarld Lofstead (Sandia National Laboratory), and Scott Klasky (Oak Ridge National Laboratory)</i>	
Characterizing the Impact of Job Execution on the Occurrence of Memory Failures on a Petascale HPC System	116
<i>Scott Levy (Sandia National Laboratories, USA), Josh Hemmert (Sandia National Laboratories, USA), Kurt B. Ferreira (Sandia National Laboratories, USA), and Kevin Pedretti (Sandia National Laboratories, USA)</i>	

Session #4

Towards Performance Portability of an Oil and Gas Application on Heterogeneous Architectures	127
<i>Arthur F. Lorenzon (Institute of Informatics - UFRGS, Brazil), Philippe O. A. Navaux (Institute of Informatics - UFRGS, Brazil), Alexandre Sardinha (Petróleo Brasileiro S.A., Brazil), and Bronson Messer (Oak Ridge National Laboratory, USA)</i>	
IDS-DEEP: A Strategy for Selecting the Best IDS for Drones with Heterogeneous EmbEdded Platforms	138
<i>Louis Morge-Rollet (ENSTA Bretagne, France; CNRS, France), Camélia Slimani (ENSTA Bretagne, France; CNRS, France), Laurent Lemarchand (Université de Bretagne Occidentale, France; CNRS, France), Frédéric Le Roy (ENSTA Bretagne, France; CNRS, France), David Espes (Université de Bretagne Occidentale, France; CNRS, France), and Jalil Boukhobza (ENSTA Bretagne, France; CNRS, France)</i>	
Optimal Time and Energy-Aware Client Selection Algorithms for Federated Learning on Heterogeneous Resources	148
<i>Alan L. Nunes (Fluminense Federal University, Brazil; University of Bordeaux, France), Cristina Boeres (Fluminense Federal University, Brazil), Lúcia M. A. Drummond (Fluminense Federal University, Brazil), and Laércio L. Pilla (University of Bordeaux, France)</i>	
Disaggregated Memory with SmartNIC Offloading: A Case Study on Graph Processing	159
<i>Jacob Wahlgren (KTH Royal Institute of Technology, Sweden), Gabin Schieffer (KTH Royal Institute of Technology, Sweden), Maya Gokhale (Lawrence Livermore National Laboratory, USA), Roger Pearce (Lawrence Livermore National Laboratory, USA), and Ivy Peng (KTH Royal Institute of Technology, Sweden)</i>	
Analyzing HPC Monitoring Data With a View Towards Efficient Resource Utilization	170
<i>Samuel Maloney (Forschungszentrum Jülich, Germany), Estela Suarez (Forschungszentrum Jülich, Germany; University of Bonn, Germany), Norbert Eicker (Forschungszentrum Jülich, Germany; University of Wuppertal, Germany), Filipe Guimarães (Forschungszentrum Jülich, Germany), and Wolfgang Frings (Forschungszentrum Jülich, Germany)</i>	

Session #5

DeVAS: Decoupled Virtual Address Spaces	182
<i>Mirco Mannino (University of Siena, Italy), Biagio Peccerillo (University of Siena, Italy), Andrea Mondelli (Huawei), and Sandro Bartolini (University of Siena, Italy)</i>	
DRAM Errors and Cosmic Rays: Space Invaders or Science Fiction?	194
<i>Isaac Boixaderas (Barcelona Supercomputing Center), Jorge Amaya (European Space Agency), Sergi Moré (Barcelona Supercomputing Center), Javier Bartolome (Barcelona Supercomputing Center), David Vicente (Barcelona Supercomputing Center), Osman Unsal (Barcelona Supercomputing Center), Dimitris Gizopoulos (University of Athens), Paul M. Carpenter (Barcelona Supercomputing Center), Petar Radojković (Barcelona Supercomputing Center), and Eduard Ayguadé (Universitat Politècnica de Catalunya)</i>	

Session #6

Memory Sandbox: A Versatile Tool for Analyzing and Optimizing HBM Performance in FPGA	206
<i>Elias Perdomo (Universitat Politècnica de Catalunya (UPC), Spain), Xavier Martorell (UPC, Spain), Teresa Cervero (BSC, Spain), and Behzad Salami (BSC, Spain)</i>	
S-Clflush: Securing Against Flush-Based Cache Timing Side-Channel Attacks	218
<i>Tejeshwar Bhagatsing Thorawade (Indian Institute of Technology Bombay, India), Prajakta Yeola (Indian Institute of Technology Bombay, India), Varun Venkitaraman (Indian Institute of Technology Bombay, India), and Virendra Singh (Indian Institute of Technology Bombay, India)</i>	
To Protect or Not To Protect: Probability-Aware Selective Protection for Sparse Iterative Solvers	229
<i>Daniel Ryley Johnson (University of Kansas, USA), Hongyang Sun (University of Kansas, USA), Joshua Dennis Booth (University of Alabama in Huntsville, USA), and Padma Raghavan (Vanderbilt University, USA)</i>	
Author Index	239