

2024 IEEE 31st International Conference on High Performance Computing, Data, and Analytics (HiPC 2024)

**Bangalore, India
18-21 December 2024**



**IEEE Catalog Number: CFP24176-POD
ISBN: 979-8-3315-0910-1**

**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:	CFP24176-POD
ISBN (Print-On-Demand):	979-8-3315-0910-1
ISBN (Online):	979-8-3315-0909-5
ISSN:	1094-7256

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 31th International Conference on High Performance Computing, Data, and Analytics (HiPC) **HiPC 2024**

Table of Contents

2024 Message from the General Co-chairs	ix
2024 Message from the Program Chairs	xii
HiPC 2024 Organization	xv
HiPC 2024 Steering Committee	xviii
HiPC 2024 Technical Program Committee	xix

Technical Session 1

Allocation Strategies for Disaggregated Memory in HPC Systems	1
<i>Robin Božennec (University of Rennes, Inria, CNRS, IRISA, France), Danilo Carastan-Santos (Univ. Grenoble Alpes, Grenoble INP, Inria, CNRS, LIG, France), Fanny Dufossé (Univ. Grenoble Alpes, Grenoble INP, Inria, CNRS, LIG, France), and Guillaume Pallez (Inria, France)</i>	
HPC Application Parameter Autotuning on Edge Devices: A Bandit Learning Approach	12
<i>Abrar Hossain (The University of Toledo), Abdel-Hameed A. Badawy (New Mexico State University), Mohammad Atiqul Islam (The University of Texas at Arlington), Tapasya Patki (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory), and Kishwar Ahmed (The University of Toledo)</i>	
Effective and Efficient Offloading Designs for One-Sided Communication to SmartNICs	23
<i>Benjamin Michalowicz (The Ohio State University, USA), Kaushik Kandadi Suresh (The Ohio State University, USA), Hari Subramoni (The Ohio State University, USA), Mustafa Abduljabbar (The Ohio State University, USA), Dhabaleswar K. Panda (The Ohio State University, USA), and Stephen Poole (Los Alamos National Laboratory)</i>	
Retrospection on the Performance Analysis Tools for Large-Scale HPC Programs	34
<i>Zhibo Xuan (Beihang University, China), Xin You (Beihang University, China), Hailong Yang (Beihang University, China), Mingzhen Li (State Key Lab of Processors, Institute of Computing Technology, Chinese Academy, China), Zhongzhi Luan (Beihang University, China), Yi Liu (Beihang University, China), and Depei Qian (Beihang University, China)</i>	

BigThrill: MPI-Based Data Processing Engine	45
<i>Anastasia Khartikova (Huawei Russian Research Institute, Russia), Denis Shaikhislamov (Huawei Russian Research Institute, Russia; Lomonosov Moscow State University, Russia), Ilya Timokhin (HSE University, Russia), Roman Kostromin (ISDCT SB RAS, Russia), Vladislav Muratov (Huawei Russian Research Institute, Russia), Aleksey Demakov (Huawei Russian Research Institute, Russia), Maxim Belov (MIPT, Russia), and Aleksey Teplov (T-Bank RnD, Russia)</i>	

Technical Session 2

Scaling Large Language Model Training on Frontier with Low-Bandwidth Partitioning	57
<i>Lang Xu (The Ohio State University, Ohio), Quentin Anthony (The Ohio State University, Ohio), Jacob Hatef (The Ohio State University, Ohio), Aamir Shafi (The Ohio State University, Ohio), Hari Subramoni (The Ohio State University, Ohio), and Dhbaleswar K. Panda (The Ohio State University, Ohio)</i>	
Transformer-Based Self-Supervised Imputation and Attention GANs Oversampling for Medical Data Processing	68
<i>Aryan Kumar Singh (National Institute of Technology Silchar, India), Arpit Saikia (National Institute of Technology Silchar, India), Pranita Baro (National Institute of Technology Silchar, India), and Malaya Dutta Borah (National Institute of Technology Silchar, India)</i>	
Exploring Algorithmic Design Choices for Low Latency CNN Deployment	78
<i>Changxin Li (Case Western Reserve University, USA) and Sanmukh Kuppannagari (Case Western Reserve University, USA)</i>	
CAR-LLM: Cloud Accelerator Recommender for Large Language Models	89
<i>Ashwin Krishnan (TCS Research, India), Venkatesh Pasumarti (TCS Research, India), Samarth Inamdar (TCS Research, India), Arghyajoy Mondal (TCS Research, India), Manoj Nambiar (TCS Research, India), and Rekha Singhal (TCS Research, India)</i>	
HyperSack: Distributed Hyperparameter Optimization for Deep Learning Using Resource-Aware Scheduling on Heterogeneous GPU Systems	100
<i>Nawras Alnaasan (The Ohio State University, USA), Bharath Ramesh (The Ohio State University, USA), Jinghan Yao (The Ohio State University, USA), Aamir Shafi (The Ohio State University, USA), Hari Subramoni (The Ohio State University, USA), and Dhbaleswar K. Panda (The Ohio State University, USA)</i>	

Technical Session 3

GDBOD: Density-Based Outlier Detection Exploiting Efficient Tree Traversals on the GPU	111
<i>Revanth Reddy Munugala (Northern Arizona University, USA) and Michael Gowanlock (Northern Arizona University, USA)</i>	

Design and Implementation of Kernel-Based MPI Reduction Operations for Intel GPUs	122
<i>Chen-Chun Chen (The Ohio State University, USA), Goutham Kalikrishna Reddy Kuncham (The Ohio State University, USA), Hari Subramoni (The Ohio State University, USA), and Dhabaleswar K. Panda (The Ohio State University, USA)</i>	
Multi-Space Tree with Incremental Construction for GPU-Accelerated Range Queries	132
<i>Brian Donnelly (Northern Arizona University, USA) and Michael Gowanlock (Northern Arizona University, USA)</i>	
A More Scalable Sparse Dynamic Data Exchange	143
<i>Andrew Geyko (Universitat des Saarlandes, Germany), Gerald Collom (University of New Mexico, USA), Derek Schafer (University of New Mexico, USA), Patrick Bridges (University of New Mexico, USA), and Amanda Bienz (University of New Mexico, USA)</i>	
Using BlueField-3 SmartNICs to Offload Vector Operations in Krylov Subspace Methods	155
<i>Kaushik Kandadi Suresh (The Ohio State University, USA), Benjamin Michalowicz (The Ohio State University, USA), Nick Contini (The Ohio State University, USA), Bharath Ramesh (The Ohio State University, USA), Mustafa Abduljabbar (The Ohio State University, USA), Aamir Shafi (The Ohio State University, USA), Hari Subramoni (The Ohio State University, USA), and Dhabaleswar Panda (The Ohio State University, USA)</i>	

Technical Session 4

From Bits to Qubits: Challenges in Classical-Quantum Integration	166
<i>Sudhanshu Kulkarni (San Francisco State University, USA) and E. Wes Bethel (San Francisco State University, USA; Lawrence Berkeley National Laboratory, USA)</i>	
Circuit Partitioning and Full Circuit Execution: A Comparative Study of GPU-Based Quantum Circuit Simulation	177
<i>Kartikey Sarode (San Francisco State University, USA)</i>	
Dual Channel Dual Staging: Hierarchical and Portable Staging for GPU-Based In-Situ Workflow	188
<i>Bo Zhang (University of Utah, USA), Philip E. Davis (University of Utah, USA), Zhao Zhang (Rutgers University, USA), Keita Teranishi (Oak Ridge National Laboratory, USA), and Manish Parashar (University of Utah, USA)</i>	
Mini-Combust—An Open-Source Unstructured FGM Combustion Mini-App for Co-Designing Aero-Engines at Extreme Scale.	199
<i>Samuel Curtis (University of Warwick), Harry Waugh (University of Bristol), Tom Deakin (University of Bristol), and Gihan R. Mudalige (University of Warwick)</i>	
Training Photonic Mach Zehnder Meshes for Neural Network Acceleration	210
<i>Andy Wolff (Ohio University, OH) and Avinash Karanth (Ohio University, OH)</i>	

Technical Session 5

ML-Based Modeling to Predict I/O Performance on Different Storage Sub-Systems	221
<i>Yiheng Xu (Palantir Technologies, Inc., USA), Pranav Sivaraman (University of Maryland, USA), Hariharan Devarajan (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, USA), Kathryn Mohror (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, USA), and Abhinav Bhatele (University of Maryland, USA)</i>	
Simulation of Large-Scale HPC Storage Systems: Challenges and Methodologies	232
<i>Julien Monniot (University of Rennes, France), François Tessier (University of Rennes, France), Henri Casanova (University of Hawai'i, USA), and Gabriel Antoniu (University of Rennes, France)</i>	
Graph Sampling Quality Prediction for Algorithm Recommendation	243
<i>Seyedehhaleh Seyeddizaji (University of Klagenfurt, Austria), Reza Farahani (University of Klagenfurt, Austria), Joze Martin Rozanec (Jozef Stefan Institute, Slovenia), Dragi Kimovski (University of Klagenfurt, Austria), Ahmet Soylu (Kristiania University College, Norway), and Radu Aurel Prodan (University of Klagenfurt, Austria)</i>	
Efficient Resource-Constrained Federated Learning Clustering with Local Data Compression on the Edge-to-Cloud Continuum	255
<i>Cédric Prigent (University of Rennes, Inria, CNRS, IRISA - Rennes, France), Melvin Chelli (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarland Informatics Campus, Germany), Alexandru Costan (University of Rennes, France), Loïc Cudennec (DGA Maîtrise de l'Information, France), René Schubotz (Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarland Informatics Campus, Germany), and Gabriel Antoniu (University of Rennes, France)</i>	
When Less is More: Achieving Faster Convergence in Distributed Edge Machine Learning	266
<i>Advik Raj Basani (BITS Pilani, India), Siddharth Chaitra Vivek (BITS Pilani, India), Advait Krishna (BITS Pilani, India), and Arnab K. Paul (BITS Pilani, India)</i>	
Leveraging LLVM OpenMP GPU Offload Optimizations for Kokkos Applications	277
<i>Rahul Kumar Gayatri (NERSC, Lawrence Berkeley National Laboratory, USA), Shilei Tian (Stony Brook University, USA), Stephen L. Olivier (Center for Computing Research, Sandia National Laboratories, USA), Eric Wright (Livermore Computing, Lawrence Livermore National Laboratory, USA), and Johannes Doerfert (Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, USA)</i>	
Author Index	289