

2025 IEEE 32nd International Conference on High Performance Computing, Data, and Analytics (HiPC 2025)

**Hyderabad, India
17-20 December 2025**

IEEE Catalog Number: CFP25176-POD
ISBN: 979-8-3315-6665-4



Copyright © 2025, IEEE

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:	CFP25176-POD
ISBN (Print-On-Demand):	979-8-3315-6665-4
ISBN (Online):	979-8-3315-6664-7
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

TABLE OF CONTENTS

HOPPS: Hardware-Aware Optimal Phase Polynomial Synthesis with Blockwise Optimization for Quantum Circuits	1
<i>Xinpeng Li, Ji Liu, Shuai Xu, Paul Hovland, Vipin Chaudhary</i>	
An Accelerator for Low-Computational Overhead Privacy-Preserving GNN Inference	12
<i>Heonhui Jung, Whoi Ree Ha, Kevin Nam, Youyeon Joo, Lucas Oros, Yunheung Paek</i>	
Efficient Fine-Grained Gpu Performance Modeling for Distributed Deep Learning of Llm	22
<i>Biyao Zhang, Mingkai Zheng, Debargha Ganguly, Xuecen Zhang, Vikash Singh, Vipin Chaudhary Zhao Zhang</i>	
Grey-Box Machine Learning Prediction of Parallel Application Scaling	33
<i>Akhil Alasandagutti, Patrick G. Bridges, Trilce Estrada</i>	
A Parallel Alternative for Energy-Efficient Neural Network Training and Inferencing	44
<i>Sudip K. Seal, Maksudul Alam, Jorge Ramirez, Sajal Dash, Hao Lu</i>	
Latency-Aware Deduplication for Efficient Object-Based Big Data Transfers in Heterogeneous Networks	55
<i>Preethika Kasu, Prince Hamandawana, Tae-Sun Chung</i>	
DistFNE: A Distributed-Memory Algorithm for Force-Directed Node Embedding	66
<i>Isuru Ranawaka, Ariful Azad</i>	
Predictive Execution of Workflows in a HPC+Cloud Environment	76
<i>Subhendu Behera, Jae-Seung Yeom, Daniel Milroy, Marc Niethammer, Frank Mueller</i>	
Layer-Wise Relevance Propagation (LRP)-Guided Clustering-Based Compression Framework for Deploying Deep Neural Network Models on Memory-Constrained Devices	87
<i>Debajyoti Sahoo, Nithyashree Ravi, Divij Ghose, Sashikumaar Ganesan</i>	
K ⁴ : Online Log Anomaly Detection Via Unsupervised Typicality Learning	96
<i>Weicong Chen, Vikash Singh, Zahra Rahmani, Debargha Ganguly, Mohsen Hariri, Vipin Chaudhary</i>	
Multi-Objective Loss Balancing in Physics-Informed Neural Networks for Fluid Flow Applications	108
<i>Afra Farea, Saiful Khan, Mustafa Serdar Celebi</i>	
Assessing Processor Allocation Strategies for Online Scheduling of Moldable Task Graphs	119
<i>Mary Jeevana Pudota, Krishna Chaitanya Reddy Chitta, Sai Rithvik Gundla, Hongyang Sun</i>	
Scalable XML Parsing and XPath Querying with Regular Expression Support	130
<i>Robert Samuel, Rupesh Nasre</i>	
Energy-Aware Runtime Resource Harmonizer for Co-Running Applications	140
<i>Vanshika Jain, Varun Parashar, Vivek Kumar, Chiranjib Sur</i>	
Machine Learning Driven Auto-Tuning for Non-Uniform All-to-All Collectives	151
<i>Kunting Qi, Ke Fan, Jens Domke, Seydou Ba, Venkatram Vishwanath, Michael E. Papka, Sidharth Kumar</i>	
Enhanced MPI Intra-Node Communication Framework: A Hybrid Approach with Cooperative DMA Channel-Based Data Transfer	161
<i>Shulei Xu, Tu Tran, Dhableswar K. DK Panda</i>	

CALL: Context-Aware Low-Latency Retrieval in Disk-Based Vector Databases	172
<i>Yeonwoo Jeong, Hyunji Cho, Kyuli Park, Youngjae Kim, Sungyong Park</i>	
IMBPS - Iterative MLP Blocks with Parameter Splits for Improving LLM Inference	183
<i>Ganesh Prasad Nagaraja, Arun Coimbatore Ramachandran, Shubhendu Sharma, R. Govindarajan Prakash Sathyanath Raghavendra</i>	
SFCC: A Scalable and Flexible RDMA Congestion Control Algorithm	194
<i>Yi Liao, Anran Xu, Wenming Zheng, Biyao Che, Jian Tang, Yonghang Zhang, Xiaoping Fan, Luren Liu Ying Chen</i>	
Tuple Spaces for Workflow Scheduling and Core-Level Malleability in Hpc	205
<i>Juan Asensio Ayesa, Darío Suárez Gracia, Lluís Castrillo-Acuña</i>	
Selection of Supervised Learning-Based Sparse Matrix Reordering Algorithms	216
<i>Tao Tang, Youfu Jiang, Yingbo Cui, Jianbin Fang, Peng Zhang, Lin Peng, Chun Huang</i>	
ZEUS: An Efficient GPU Optimization Method Integrating PSO, BFGS, and Automatic Differentiation	225
<i>Dominik Soós, Marc Paterno, Desh Ranjan, Mohammad Zubair</i>	
QIEDP: A Quantum-Inspired Two-Bit Error Correction Protocol for Low-Power Serial Communication in IoT Systems	236
<i>Om Maheshwari, Bikram Paul</i>	
Embracing Dynamism: Control State Serialization for High-Performance Python	246
<i>Zane Fink, Laxmikant Kale</i>	
Optimizing Deployment of Unstructured Group Convolutions for Low Latency Inference	257
<i>Changxin Li, Sanmukh Kuppannagari</i>	
Distributed Metadata Querying on HPC Systems	268
<i>Suben Kumer Saha, Houjan Tang, Wei Zhang, Suren Byna</i>	
Dynamic Resource Management in HPC Systems Using Dynamic Processes with PSets	279
<i>Dominik Huber, Keerthi Gaddameedi, Tobias Neckel, Hans-Joachim Bungartz, Martin Schulz Pierre-François Dutot, Olivier Richard, Martin Schreiber, Sergio Iserte, Antonio Pena</i>	
NiceSched : Memory Locality Aware Dynamic Priority Scheduling in Tiered Memory	290
<i>Binwon Song, Minwoo Jo, Hayong Jeong, Heeseung Jo</i>	
Performance-Portable Optimization and Analysis of Multiple Right-Hand Sides in a Lattice QCD Solver	301
<i>Shiting Long, Gustavo Ramirez-Hidalgo, Stepan Nassyr, Jose Jimenez-Merchan, Andreas Frommer Dirk Pleiter</i>	
UniOMP: Unified Optimization Framework for OpenMP Offload Under Machine Learning Guidance	312
<i>Jianan Li, Lin Han, Shaoliang Peng, Yingying Li, Wei Gao</i>	
Maximizing Insights, Minimizing Data: I/O Time Prediction Using Transfer Learning	322
<i>Adrian Voß, Radita Liem, Julian Kunkel, Jay Lofstead, Philip Carns, Matthias Müller</i>	
LLM4VV:: Evaluating Cutting-Edge LLMs for Generation and Evaluation of Directive-Based Parallel Programming Model Compiler Tests	333
<i>Zachariah Sollenberger, Rahul Patel, Saieda Ali Zada, Sunita Chandrasekaran</i>	

LatentTune: Efficient Tuning of High Dimensional Database Parameters via Latent Representation Learning	343
<i>Sein Kwon, Youngwan Jo, Seungyeon Choi, Jieun Lee, Huijun Jin, Sanghyun Park</i>	
Predicting Executability and Performance of CNN Kernels on Tenstorrent Hardware Using Machine Learning	354
<i>Param Gandhi, Sharvil Potdar, Nayan Gogari, Gargi Alavani Prabhu, Sankar Manoj, Santonu Sarkar</i>	
Context-Driven Performance Modeling for Causal Inference Operators on Neural Processing Units	365
<i>Neelesh Gupta, Rakshith Jayanth, Dhruv Parikh, Viktor Prasanna</i>	