June 8 – 11, 2025
Salt Lake City, USA

Association for Computing Machinery

Advancing Computing as a Science & Profession

ACM ICS '25

Proceedings of the 39th ACM International Conference on Supercomputing

Sponsored by:

ACM SIGARCH & SIGHPC

Association for Computing Machinery

Advancing Computing as a Science & Profession

The Association for Computing Machinery
1601 Broadway, 10th Floor
New York, New York 10019, USA

ISBN: 979-8-4007-1537-2

Sponsors

Association for Computing Machinery

ACM SIGARCH

sighpc

Message from the General Chair

Welcome to the 39th ACM International Conference on Supercomputing (ICS 2025), held June 8-11 in Salt Lake City, USA. ICS has been a premiere forum for presentation of research in all aspects of high-performance computing, including architecture, algorithms/applications, compilers and runtime systems, performance modeling/measurement, programming languages, and systems software. The key to the success of the conference is the creation of a high-quality program covering a wide range of topics with results that advance the state-of-the-art. I thank the Program Chairs Michael Ferdman and Xipeng Shen for their tireless efforts in managing the dual submissions (for the first time at ICS 2025) and the creation of a very strong technical program with 83 papers.

Grateful thanks to ACM and the co-sponsorship of ICS by SIGARCH and SIGHPC, without whose support the conference would not be possible. Thanks to SIGHPC for providing funds for travel grants to student and early career attendees.

I am very thankful for all the heavy lifting by the rest of the team that made ICS 2025 possible: Hari Sundar, the Finance and Registration Chair; Varun Shankar, the Local Arrangements Chair; Yuke Wang, the Publications Chair; Wenqian Dong, the Publicity Chair; Rohan Basu Roy, the Travel Grants Chair; Yufan Xu, the Web Liaison; and Mehmet Belviranli, the Workshops Chair.

I wish to express my gratitude to the current and former steering committee chairs, Dimitrios Nikolopoulos and Alex Veidenbaum, for their valuable advice and help on so many organizational matters.

I hope that you will enjoy the program and interactions at ICS 2025!

P. (Saday) Sadayappan
University of Utah
USA

Message from the Program Chairs

Welcome to the 39th ACM International Conference on Supercomputing (ICS 2025), held June 8–11, 2025, in Salt Lake City, USA. ICS continues its tradition as a premier forum for high-performance computing research. We are pleased to again convene *in person*, fostering vibrant discussion and collaboration across the community.

This year we used a two-cycle submission model, with back-to-back deadlines in January and February 2025. We received 320 submissions—178 in the first cycle and 142 in the second. The deadline for the second cycle was intentionally set before the decisions for the first cycle were released to prevent re-submission of the same papers to both cycles. In each cycle, the authors had a chance to submit a written rebuttal to the reviews prior to the commencement of the discussion period among the reviewers. We followed a strict double-blind review process, where the reviewers were unaware of the author identities and vice versa. Moreover, we retained anonymity throughout the online discussion period, where the reviewer names were not disclosed to the other reviewers.

Decisions were made through a combination of online discussion and virtual PC meetings via Zoom. During online discussion, unanimity among all of a paper's reviewers was needed to decide if a paper was rejected or accepted. Moreover, for a paper to be accepted during the online discussion phase, the paper had to have at least one reviewer act as champion. Papers that had a champion, but for which a unanimous decision was not reached during online discussion, were discussed at the virtual PC meeting. At the PC meeting, each of the undecided papers was discussed by the present reviewers, followed by a vote where the vote of any missing reviewers was taken from their latest review score. If a unanimous decision was not reached at that time, each paper was brought to a vote of all Program Committee members that were attending the PC meeting, with a simple majority of the attendees determining the final outcome. We accepted 83 papers in total (50 from the first cycle and 33 from the second), yielding an overall acceptance rate of 26%.

The accepted papers span a broad array of topics—from architectures and systems to programming models and applications—reflecting the dynamic nature of supercomputing research. This largest-ever ICS technical program comprises 19 sessions, including a *Best Papers session* that highlights work selected based on the highest review scores and reviewer endorsements for best paper consideration.

We express our gratitude to the 122 members of the Program Committee and the nine external reviewers for their expert evaluations. We are especially thankful to the 12 Program Committee members who served as shepherds for 14 conditionally accepted papers and the seven Program Committee members who agreed to participate on the best-paper selection committee. We are grateful to Saday Sadayappan (University of Utah), our General Chair, for taking care of all the conference organization details beyond paper selection. We thank Yuke Wang (Rice University) for liaising between the authors of the large number of accepted papers and our publisher. Finally, we would like to thank Xinning Hui and Ruifeng Zhang, graduate students from NCSU, who assisted us in running the Zoom virtual PC meetings.

We hope you find ICS 2025 intellectually rewarding and that it inspires your future work. Thank you for joining us in Salt Lake City.

Michael Ferdman (Stony Brook University)
Xipeng Shen (North Carolina State University)
Program Co-Chairs, ICS 2025

ICS 2025 Organizing Committee

General Chair

P. (Saday) Sadayappan University of Utah

Program co-chairs

Michael Ferdman Stony Brook University

Xipeng Shen North Carolina State University

Finance chair

Hari Sundar Tufts University

Publication chair

Yuke Wang Rice University

Publicity chair

Wenqian Dong Oregon State University

Web liaison

Yufan Xu Uber Technologies Inc.

Workshops and tutorials co-chairs

Mehmet Belviranli Colorado School of Mines

Local Arrangement

Varun Shankar University of Utah

ICS 2025 Program Committee

Almutaz Adileh	Huawei
Varun Agrawal	Advanced Micro Devices Inc
Alaa Alameldeen	Simon Fraser University
Rachata Ausavarungnirun	Mangoboost Inc.
Amro Awad	University of Oxford
Saurabh Bagchi	Purdue University
Rajeev Balasubramonian	University of Utah
Kevin Barker	Pacific Northwest National Laboratory
Tekin Bicer	Argonne National Laboratory
Mehmet Belviranli	Colorado School of Mines
Martin Burtscher	Texas State University
Suren Byna	The Ohio State University
Franck Cappello	Argonne National Laboratory
Trevor E. Carlson	National University of Singapore
Vito Giovanni Castellana	Pacific Northwest National Laboratory
Sunita Chandrasekaran	University of Delaware
Mainak Chaudhuri	Indian Institute of Technology Kanpur
Guoyang Chen	Apple Inc.
Jou-Ann Chen	Qualcomm
Wenguang Chen	Tsinghua University
Quan Chen	Shanghai Jiao Tong University
Zizhong Chen	UC Riverside
Preyesh Dalmia	NVIDIA
Sina Darabi	Università della Svizzera italiana
Sheng Di	Argonne National Laboratory
Yufei Ding	UCSD
Chen Ding	University of Rochester
Christian Engelmann	Oak Ridge National Laboratory
Simon Garcia de Gonzalo	Sandia National Laboratories
Jayesh Gaur	Intel Labs
Rong Ge	Clemson University
R Govindarajan	Indian Institute of Science
Hui Guan	UMass Amherst
Kyle Hale	Oregon State University
Martin Herbordt	Boston University
Jin Huang	Brookhaven National Lab
Nikhil Jain	Nvidia
Ali Jannesari	Iowa State University
Peng Jiang	University of Iowa
Daniel Jiménez	Texas A&M University

Mahmut Kandemir	Pennsylvania State University
David Kaeli	Northeastern University
Sudarsun Kannan	Rutgers University
Ahmad Maroof Karimi	Oak Ridge National Laboratory
Omer Khan	University of Connecticut
John Kim	KAIST
Youngsok Kim	Yonsei University
Milind Kulkarni	Purdue University
Hyoukjun Kwon	University of California, Irvine
Ignacio Laguna	Lawrence Livermore National Laboratory
Zhiling Lan	University of Illinois Chicago
Dongyoon Lee	Stony Brook University
Gushu Li	University of Pennsylvania
Jiajia Li	NCSU
Dong Li	UC Merced
Lingda Li	Brookhaven National Laboratory
Ang Li	PNNL and UW
Chao Li	Shanghai Jiao Tong University
Guanpeng Li	University of Iowa
Chunhua Liao	Lawrence Livermore National Laboratory
Mieszko Lis	University of British Columbia
Xu Liu	Google
Jinyang Liu	University of Houston
Liu Liu	Rensselaer Polytechnic Institute
Pejman Lotfi-Kamran	Institute for Research in Fundamental Sciences
Jason Lowe-Power	University of California, Davis
Jason Mars	University of Michigan
Marco Minutoli	Pacific Northwest National Laboratory
William Moses	UIUC/Google
Abdullah Muzahid	Texas A&M University
Bogdan Nicolae	Argonne National Laboratory
Dimitrios S. Nikolopoulos	Virginia Tech
Soner Onder	Michigan Tech
Dhabaleswar K Panda	The Ohio State University
Prashant Pandey	Northeastern University
Tirthak Patel	Rice University
Suchita Pati	AMD Research & Advanced Development
Arnab K. Paul	BITS Pilani, Goa Campus
Jacques Pienaar	Google
Aleksandar Prokopec	Oracle Labs
Moinuddin Qureshi	Georgia Tech
Yihui Ren	Brookhaven National Laboratory

Tahsin Reza	University of Waterloo
Bin Ren	William & Mary
John (Jack) Sampson	Penn State
Bertil Schmidt	JGU Mainz
Joseph Schuchart	Stony Brook University
Martin Schulz	Technical University of Munich
Prateek Sharma	Indiana University
Seunghee Shin	State University of New York at Binghamton
Arrvindh Shriraman	Simon Fraser University
Yan Solihin	UCF
Ravi Soundararajan	VMware by Broadcom
Aravind Sukumaran Rajam	Meta
Hsin-Hsuan Sung	Qualcomm
Guangming Tan	Chinese Academy of Sciences
Kenjiro Taura	The University of Tokyo
Antonino Tumeo	Pacific Northwest National Laboratory
Nilay Vaish	Google
Hans Vandierendonck	Queens University Belfast
Ashish Venkat	University of Virginia
Guru Venkataramani	George Washington University
Chen Wang	Lawrence Livermore National Laboratory
Bo Wu	Colorado School of Mines
Yuanchao Xu	University of California, Santa Cruz
Helen Xu	Georgia Tech
Wei Xue	Tsinghua Uviersity
Feng Yan	University of Houston
Jun Yang	University of Pittsburgh
Yifan Yang	Nvidia
Chencheng Ye	Huazhong University of Science and Technology
Pen-Chung Yew	University of Minnesota at Twin Cities
Youngmin Yi	Sogang University
Cliff Young	Google DeepMind
Xiaodong Yu	Stevens Institute of Technology
Jianping Zeng	Samsung MSL
Jidong Zhai	Tsinghua University
Youtao Zhang	University of Pittsburgh
Zheng Zhang	Rutgers University
Feng Zhang	Renmin University of China
Zhijia Zhao	University of California, Riverside
Keren Zhou	George Mason University

ICS 2025 External Reviewers

Aamer Jaleel	Nvidia
Aparna Chandramowlishwaran	UC Irvine
Bilge Acun	Meta
Boris Grot	University of Edinburgh
Frank Mueller	North Carolina State University
Jiesong Liu	North Carolina State University
Michela Becchi	North Carolina State University
Nael Abu-Ghazaleh	University of California, Riverside
Natalie Enright Jerger	University of Toronto
Nengkun Yu	Stony Brook University
Nikos Hardavellas	Northwestern
Rajiv Gupta	University of California, Riverside
Robert Underwood	Argonne National Laboratory
Swamit Tannu	University of Wisconsin–Madison
Tony (Tong) Geng	University of Rochester
Wei Niu	University of Georgia
Zhenhua Liu	Stony Brook University

ICS 2025 Steering Committee

Alex Nicolau	University of California Irvine
Alex Veidenbaum	University of California Irvine
Ana Lucia Varbanescu	University of Twente
Avi Mendelson	Technion
Bill Gropp	University of Illinois
Chen Ding	University of Rochester
Constantine Polychronopoulos	Juniper Networks
Dimitrios Nikolopoulos (Chair)	VTU
Dionisios Pnevmatikatos	NTUA
Eduard Ayguade	Barcelona Supercomputing Center
Fran Cazorla	Technical University of Catalonia/BSC
Frank Mueller	NCSU
Fred Chong	University of Chicago
Harry Wijshoff	Leiden University
Huiyang Zhou	NCSU
James Goodman	University of Wisconsin–Madison
Jose Moreira	IBM
Kenji Kise	Tokyo Institute of Technology

Kirk Cameron	VTU
Kyle Gallivan	Florida State University
Lawrence Rauchwerger	University of Illinois
Mahmut Kandemir	Pennsylvania State University
Mateo Valero	Technical University of Catalonia/BSC
Michael Gschwind	Facebook
Murali Annavaram	University of Southern California
Pete Beckman	Argonne Natl Lab
Peter Hofstee	IBM and TU Delft
Rosa Badia	Barcelona Supercomputing Center
Rudi Eigenman	University of Delaware
Sally A. McKee	Clemson
Valentina Salapura	Google
John Sopka	Independent consultant
Wen-mei Hwu	Nvidia
Yoav Etsion	Technion
Zhiyuan Li	Purdue

Contents

Session: Sparse Linear Algebra

Session: Acceleration

Session: Applications

Session: GPU Scheduling

Melina Soysal (ETH Zurich), Konstantina Koliogeorgi (ETH Zurich), Can Firtina (ETH Zurich), Nika Mansouri Ghiasi (ETH Zurich), Rakesh Nadig (ETH Zurich), Haiyu Mao (ETH Zurich), Geraldo Francisco de Oliveira Junior (ETH Zurich), Yu Liang (ETH Zurich), Klea Zambaku (ETH Zurich Bilkent University), Mohammad Sadrosadati (ETH Zürich), Onur Mutlu (ETH Zurich)

Session: Efficiency

DALdex: A DPU-Accelerated Persistent Learned Index via Incremental Learning
Aoyang Tong (Huazhong University of Science and Technology), Yu Hua (Huazhong University of Science and Technology), Menglei Chen (Huazhong University of Science and Technology)

From Islands to Archipelago: Towards Collaborative and Adaptive Burst Buffer for HPC Systems
Mingtian Shao (National University of Defense Technology), Ruibo Wang (National University of Defense Technology), Wenzhe Zhang (National University of Defense Technology), Kai Lu (National University of Defense Technology), Yiqin Dai (National University of Defense Technology), Huijun Wu (National University of Defense Technology)

PIE: Enabling Fast and Scalable Incremental Evolving Graph Analytics on Persistent Memory
Yunmo Zhang (City University of Hong Kong), Jiacheng Huang (City University of Hong Kong), Xizhe Yin (University of California Riverside), Junqiao Qiu (City University of Hong Kong), Hong Xu (The Chinese University of Hong Kong), Chun Jason Xue (MBZUAI)

DEDUPKV: A Space-Efficient and High-Performance Key-Value Store via Fine-Grained Deduplication
Safdar Jamil (Sogang University), Awais Khan (Oak Ridge National Lab), Xubin He (Temple University), Youngjae Kim (Sogang University)

Session: Optimizing Compilation

ConTraPh: Contrastive Learning for Parallelization and Performance Optimization
Quazi Ishtiaque Mahmud (Iowa State University), Ali TehraniJamsaz (Iowa State University), Nesreen K. Ahmed (Cisco AI Research), Theodore L. Willke (DataStax), Ali Jannesari (Iowa State University)

UJOpt: Heuristic Approach for Applying Unroll-and-Jam Optimization and Loop Order Selection
Shilpa Babalad (Indian Institute of Science), Shirish K Shevade (Indian Institute of Science), Matthew Jacob Thazhuthaveetil (Indian Institute of Science), R Govindarajan (Indian Institute of Science)

Loop Fusion in Matrix Multiplications with Sparse Dependence
Mohammad Mehdi Salehi (McMaster University), Kazem Cheshmi (McMaster University)

ConCo: Optimizing Compilation of Concurrent Tensor Programs on Shared GPU
Jiamin Lu (University of Science and Technology of China), Jingwei Sun (University of Science and Technology of China), Yunlong Xu (Independent Researcher), Peng Sun (Independent Researcher), Guangzhong Sun (University of Science and Technology of China)

Session: Best Papers

Pushing the Limits of GPU Lossy Compression: A Hierarchical Delta Approach
Boyuan Zhang (Indiana University), Yafan Huang (University of Iowa), Sheng Di (Argonne National Laboratory), Fengguang Song (Indiana University), Guanpeng Li (University of Iowa), Franck Cappello (Argonne National Laboratory)

Session: Performance Analysis

Session: Heterogeneity

Understanding the Idiosyncrasies of Emerging BlueField DPUs
Arjun Kashyap (University of California, Merced), Yuke Li (University of California, Merced), Darren Ng (University of California, Merced), Xiaoyi Lu (University of California, Merced)

Multi-node Multi-GPU Datalog
Ahmedur Rahman Shovon (University of Illinois Chicago), Yihao Sun (Syracuse University), Kristopher Micinski (Syracure University), Thomas Gilray (Washington State University), Sidharth Kumar (University of Illinois Chicago)

SmartNIC-GPU-CPU Heterogeneous System for Large Machine Learning Model with Software-Hardware Codesign
Anqi Guo (Boston University), Yuchen Hao (Meta Platforms), Xiteng Yao (Boston University), Shining Yang (Boston University), Jianyu Huang (Meta Platforms), Tony (Tong) Geng (University of Rochester), Martin Herbordt (Boston University)

D-Rex: Heterogeneity-Aware Reliability Framework and Adaptive Algorithms for Distributed Storage
Maxime Gonthier (University of Chicago,Argonne National Laboratory), Dante D. Sanchez-Gallegos (University Carlos III of Madrid), Haochen Pan (University of Chicago), Bogdan Nicolae (Argonne National Laboratory), Sicheng Zhou (Southern University of Science and Technology), Hai Duc Nguyen (University of Chicago ,Argonne National Laboratory), Valerie Hayot-Sasson (University of Chicago,Argonne National Laboratory), J. Gregory Pauloski (University of Chicago), Jesus Carretero (University Carlos III of Madrid), Kyle Chard (University of Chicago,Argonne National Laboratory), Ian Foster (University of Chicago,Argonne National Laboratory)

Session: Resource Management

ORION: Optimizing OLAP Query Execution with Proactive Caching and Separate Operators
Zhixin Tong (Shanghai Jiao Tong University), Jiuchen Shi (Shanghai Jiao Tong University,The Hong Kong Polytechnic University), Quan Chen (Shanghai Jiao Tong University), Pu Pang (Shanghai Jiao Tong University), Shixuan Sun (Shanghai Jiao Tong University), Jie Meng (Huawei Cloud), Jiang Liu (Huawei Cloud), En Shao (Chinese Academy of Sciences), Minyi Guo (Shanghai Jiao Tong University)

ORA: Job Runtime Prediction for High-Performance Computing Platforms Using the Online Retrieval-Augmented Language Model
Hongyi Liu (Peking University), Yinping Ma (Peking University), Xiaosong Huang (Peking University), Lingzhe Zhang (Peking University), Tong Jia (Peking University,National Key Laboratory of Data Space Technology and System), Ying Li (Peking University)

Generating Microservice Graphs with Production Characteristics for Efficient Resource Scaling
Fanrong Du (Shanghai Jiao Tong University), Jiuchen Shi (Shanghai Jiao Tong University,The Hong Kong Polytechnic University), Quan Chen (Shanghai Jiao Tong University), Pu Pang (Shanghai Jiao Tong University), Li Li (Shanghai Jiao Tong University), Minyi Guo (Shanghai Jiao Tong University)

HARNESS: Holistic Resource Management for Diversely Scaled Edge Cloud Systems
Ismet Dagli (Colorado School of Mines), Justin Davis (Colorado School of Mines), Mehmet Esat Belviranli (Colorado School of Mines)

Session: Code Optimization

Yunzhuo Liu (Nanjing University), Chengyuan Huang (Nanjing University), Tao Sun (China Mobile), Xiaodong Duan (China Mobile), Guihai Chen (Nanjing University), Chen Tian (Nanjing University)

Auto-Healer: Self-Healing Hardware for Perception Stage Faults in Autonomous Driving Systems
Ali Suvizi (George Washington University), Guru Venkataramani (George Washington University)

OpaQue: Program Output Obfuscation for Quantum Software Circuits in Quantum Clouds
Tirthak Patel (Rice University), Aditya Ranjan (Northeastern University), Daniel Silver (Northeastern University), Harshitta Gandhi (QBit Solutions Research), William Cutler (Oxford University), Devesh Tiwari (Northeastern University)

JBSA: A Bit-Serial Accelerator for Deep Neural Networks Using Superconducting SFQ Logic
Yang Su (ShanghaiTech University), Sheng Li (ShanghaiTech University), Huilong Jiang (Chinese Academy of Sciences), Haofei Yin (ShanghaiTech University), Rongliang Fu (The Chinese University of Hong Kong), Junying Huang (Chinese Academy of Sciences), Xiaochun Ye (Chinese Academy of Sciences), Zhimin Zhang (Chinese Academy of Sciences), Jie Ren (Chinese Academy of Sciences), Xiaoping Gao (Chinese Academy of Sciences), Tsung-Yi Ho (The Chinese University of Hong Kong), Dongrui Fan (Chinese Academy of Sciences)

Session: Graph Algorithms

YH-Light: Yielding Hierarchy-aware Partitioner for Large-scale Graph Processing
Xinbiao Gan (National University of Defense Technology), Tiejun Li (National University of Defense Technology), Chunye Gong (National University of Defense Technology), Jie Liu (National University of Defense Technology), Kai Lu (National University of Defense Technology)

MG-αGCD: Accelerating Graph Community Detection on Multi-GPU Platforms
Shuai Yang (Chinese Academy of Sciences), Changyou Zhang (Chinese Academy of Sciences)

GraCFL: A Versatile Vertex-Centric Graph System for High-Performance CFL Reachability Analysis
Sakib Fuad (University of California, Riverside), Amir Hossein Nodehi Sabet (University of California, Riverside), Umar Farooq (University of California, Riverside), Zhijia Zhao (University of California, Riverside)

OPMOS: Ordered Parallel Algorithm for Multi-Objective Shortest-Paths
Leo Gold (University of Connecticut), Adam Bienkowski (University of Connecticut), David Sidoti (US Naval Research Laboratory), Krishna Pattipati (University of Connecticut), Omer Khan (University of Connecticut)

A Multi-GPU Algorithm for Computing Maximal Independent Sets in Large Graphs
Anju Mongandampulath Akathoott (Texas State University), Benila Virgin Jerald Xavier (Texas State University), Martin Burtscher (Texas State University)

Session: Memory Systems

A Cost-Effective Dueling Framework for Set-Associative Cache Indexing
Kevin Weston (Texas A&M University), Vahid Janfaza (Texas A&M University), Avery Johnson (Texas A&M University), Abdullah Muzahid (Texas A&M University)

DREAM: Device-Driven Efficient Access to Virtual Memory
Nurlan Nazaraliyev (University of California, Riverside), Elaheh Sadredini (UC Riverside), Nael Abu-Ghazaleh (Computer Science and Engineering , University of California, Riverside)