

ICS 2019 Program

Wednesday Lunch 12:30pm to 2pm

ICS 2019 Welcome and Opening Remarks (2pm, Wednesday, Room 101ABC)

HPC Applications (Plenary Session 1: 2:05pm to 3:45pm, Wednesday, Room 101ABC)

Chair: Rudolf Eigenmann, University of Delaware

A scalable framework for Adaptive Computational General Relativity on Heterogeneous Clusters	Milinda Fernando (University of Utah); David Neilsen (Brigham Young University); Hari Sundar (University of Utah)
Parallelizing Cryo-EM 3D Reconstruction on GPU Cluster with A Partitioned and Streamed Model	Kunpeng Wang, Shizhen Xu, Haohuan Fu, Hongkun Yu, Wenlai Zhao and Guangwen Yang (Tsinghua University)
Efficient GPU Tree Walks for Effective Distributed N-Body Simulations	Jianqiao Liu (School of Electrical and Computer Engineering, Purdue University); Michael Robson (Department of Computer Science, University of Illinois Urbana-Champaign); Thomas Quinn (Department of Astronomy, University of Washington); Milind Kulkarni (School of Electrical and Computer Engineering, Purdue University)
Hybrid CPU/GPU Clustering in Shared Memory on the Billion Point Scale	Michael Gowanlock (Northern Arizona University)

Accelerator Programming (Plenary Session 2: 4pm to 5:40pm, Wednesday, Room 101ABC)

Chair: Chen Ding, University of Rochester

Accelerating Reduction and Scan Using Tensor Core Units	Abdul Dakkak and Cheng Li (UIUC); Jinjun Xiong (IBM); Wen-Mei Hwu (UIUC); Isaac Gelado (NVIDIA)
Laius: Towards Latency Awareness and Improved Utilization of Spatial Multitasking Accelerators in Datacenters	Wei Zhang, Weihao Cui, Kaihua Fu, and Quan Chen (Shanghai Jiao Tong University); Daniel Edward Mawhirter and Bo Wu (Colorado School of Mines); Chao Li and Minyi Guo (Shanghai Jiao Tong University)
HYPHA: A Framework based on Separation of Parallelisms to Accelerate Persistent Homology Matrix Reduction	Simon Zhang, Mengbai Xiao, Chengxin Guo, Liang Geng, Hao Wang, and Xiaodong Zhang (Ohio State University)
SDC: A Software Defined Cache for Efficient Data Indexing	Fan Ni, Song Jiang, and Hong Jiang (University of Texas, Arlington); Jian Huang (University of Illinois at Urbana-Champaign); Xingbo Wu (University of Illinois at Chicago)

Thursday Breakfast 7:45am to 8am

HPC Algorithms: Linear Algebra and Solvers (Parallel Session 3A: 8:30am to 11am, Thursday, Room 225A)

Chair: Hari Sundar, University of Utah

IA-SpGEMM: an Input-aware Auto-tuning Framework for Parallel Sparse Matrix-Matrix Multiplication	Zhen Xie and Guangming Tan (Institute of Computing Technology, CAS); Weifeng Liu (China University of Petroleum, Beijing); Ninghui Sun (State Key Laboratory of Computer Architecture, Institute of Computing Technology, CAS)
TSM2: Optimizing Tall-and-Skinny Matrix-Matrix Multiplication on GPUs	Jieyang Chen, Nan Xiong, and Xin Liang (University of California, Riverside); Dingwen Tao (University of Alabama); Sihuan Li, Kaiming Ouyang, and Kai Zhao (University of California, Riverside); Nathan DeBardeleben (Los Alamos National Laboratory); Qiang Guan (Kent State University); Zizhong Chen (University of California, Riverside)
Least Squares Solvers for Distributed Memory Machines with GPU Accelerators	Jakub Kurzak, Mark Gates, Ali Charara, Asim YarKhan, and Jack Dongarra (University of Tennessee)
A communication-avoiding 3D sparse triangular solver	Piyush Sao and Ramakrishana Kannan (ORNL); Xiaoye Li (LBNL); Richard Vuduc (Georgia Tech)
Using Performance Models to Understand Scalable Krylov Solver Performance at Scale for Structured Grid Problems	Paul Eller (University of Illinois at Urbana-Champaign); Torsten Hoefer (ETH Zurich); William Gropp (University of Illinois at Urbana-Champaign)
Performance Optimization of Reactive Molecular Dynamics Simulations With Dynamic Charge Distribution Models on Distributed Memory Platforms	Kurt A. O'Hearn, Hasan Metin Aktulga, and Abdullah Alperen (Michigan State University)

HPC Computer Architectures / Accelerators (Parallel Session 3B: 8:30am to 11am, Thursday, Room 225B)

Chair: Milind Kulkarni, Purdue University

AMPT-GA: Automatic Mixed Precision Floating Point Tuning for GPU Applications	Pradeep Venkata Kotipalli and Ranvijay Singh (Purdue University); Paul Wood (Johns Hopkins University); Ignacio Laguna (LLNL); Saurabh Bagchi (Purdue University)
GPU Snapshot: Checkpoint Offloading for GPU-Dense Systems	Kyushick Lee (University of Texas at Austin); Michael Sullivan, Siva Kumar Sastry Hari, Timothy Tsai, and Stephen W. Keckler (NVIDIA); Mattan Erez (University of Texas at Austin)
Address-Stride Assisted Approximate Load Value Prediction in GPUs	Haonan Wang and Mohamed Ibrahim (College of William & Mary); Sparsh Mittal (IIT Hyderabad); Adwait Jog (College of William & Mary)
Diligent TLBs: A Mechanism for Exploiting Heterogeneity in TLB Miss Behavior	Hussein Elnawawy (North Carolina State University); Rangeen Basu Roy Chowdhury (Intel Corporation); Amro Awad (University of Central Florida); Gregory Byrd (North Carolina State University)
QoSMT: Supporting Precise Performance Control for Simultaneous multithreading Architecture	Xin Jin, YaoYang Zhou, Bowen Huang, Zihao Yu, Xusheng Zhan, Huizhe Wang, and Sa Wang (Institute of Computing Technology, Chinese Academy of Sciences); Ningmei Yu (Xi'an University of Technology); Ninghui Sun and Yungang Bao (Institute of Computing Technology, Chinese Academy of Sciences)
An Online Quality Management Framework for Approximate Communication in Network-on-Chips	Yuechen Chen and Ahmed Louri (The George Washington University)

Thursday Lunch 12:30pm to 2pm

HPC Algorithms: Graphs and Tensors (Parallel Session 4A: 2pm to 3:40pm, Thursday, Room 225A)

Chair: Sreepathi Pai, University of Rochester

Efficient and Effective Sparse Tensor Reordering	Jijia Li (Pacific Northwest National Laboratory); Bora Ucar (UMR5668, CNRS and ENS Lyon); Umit Catalyurek and Jimeng Sun (Georgia Institute of Technology); Kevin Barker (Pacific Northwest National Laboratory); Richard Vuduc (Georgia Institute of Technology)
On optimizing distributed non-negative Tucker decomposition	Venkatesan T. Chakaravarthy and Shivmaran S. Pandian (IBM Research - India); Saurabh Raje (BITS Pilani, India); Yogish Sabharwal (IBM Research - India)
GPU Road Network Graph Contraction and SSSP Query	Roopbeh Karimi and David M. Koppelman (Louisiana State University, Electrical & Computer Engineering); Chris J. Michael (Naval Research Laboratory)
Multi-criteria partitioning of multi-block structured grids	Hengjie Wang and Aparna Chandramowlishwaran (University of California, Irvine)

Modeling / Resource Management (Parallel Session 4B: 2pm to 3:40pm, Thursday, Room 225B)

Chair: Xiaodong Zhang, Ohio State University

Avalon: Towards QoS Awareness and Improved Utilization through Multi-Resource Management in Datacenters	Quan Chen (Shanghai Jiao Tong University); Zhenning Wang (Huawei Inc.); Jingwen Leng, Chao Li, Wenli Zheng, and Minyi Guo (Shanghai Jiao Tong University)
Can We Trust Profiling Results? Understanding and Fixing the Inaccuracy in Modern Profilers	Hao Xu and Qingsen Wang (College of William and Mary); Shuang Song and Lizy Kurian John (The University of Texas at Austin); Xu Liu (College of William and Mary)
Power Efficient Job Scheduling by Predicting the Impact of Processor Manufacturing Variability	Dimitrios Chasapis, Marc Casas, and Miquel Moretó (Barcelona Supercomputing Center); Martin Schultz (Technical University of Munich); Barry Rountree (Lawrence Livermore National Laboratory); Mateo Valero (Barcelona Supercomputing Center)
GreenMM: Energy-Efficient GPU Matrix Multiplication Through Undervolting	Hadi Zamani Sabzi, Yuanlai Liu, Devashree Tripathy, Laxmi N. Bhuyan, and Zizhong Chen (University of California, Riverside)

Parallel Programming (Parallel Session 5A: 4pm to 5:40pm, Thursday, Room 225A)

Chair: Jiang Song, University of Texas at Arlington

WCCV: Improving Vectorization of IF-statements with Warp-Coherent Conditions	Huihui Sun and Florian Fey (University of Münster); Jie Zhao (State key Laboratory of Mathematical Engineering and Advanced Computing); Sergei Gorlatch (University of Münster)
Automatic Construct Selection and Variable Classification in OpenMP	Mohammad Norouzi, Felix Wolf (Technische Universitaet Darmstadt); and Ali Jannesari (Iowa State University)
Efficient Thread/Page/Parallelism Autotuning for NUMA Systems	Mihail Popov, Alexandra Jimborean, and David Black-Schaffer (Uppsala University)
Efficient Hierarchical Online-Autotuning	Philip Pfaffe (Karlsruhe Institute of Technology); Tobias Grosser (ETH Zurich); Martin Tillmann (Karlsruhe Institute of Technology)

Distributed Systems (Parallel Session 5B: 4pm to 5:40pm, Thursday, Room 225B)

Chair: Bronis R. de Supinski, Lawrence Livermore National Lab

Software Combining to Mitigate Multithreaded MPI Contention	Abdelhalim Amer (Argonne National Laboratory); Charles Archer, Michael Blocksome, Chongxiao Cao, Michael Chuvelev, Hajime Fujita, and Maria Garzaran (Intel Corporation); Yanfei Guo (Argonne National Laboratory); Jeff R. Hammond (Intel Corporation); Shintaro Iwasaki (The University of Tokyo); Kenneth J. Raffanetti (Argonne National Laboratory); Mikhail Shiryaev (Intel Corporation); Min Si (Argonne National Laboratory); Kenjiro Taura (The University of Tokyo); Sagar Thapaliya (Intel Corporation); Pavan Balaji (Argonne National Laboratory)
Optimizing Computation-Communication Overlap in Asynchronous Task-Based Programs	Emilio Castillo (Barcelona Supercomputing Center); Nikhil Jain (Lawrence Livermore National Laboratory); Marc Casas (Barcelona Supercomputing Center); Miquel Moreto (Technical Univeristy of Catalonia); Martin Schulz (Technical University of Munich); Ramon Beivide (University of Cantabria); Mateo Valero (Barcelona Supercomputing Center); Abhinav Bhatele (Lawrence Livermore National Laboratory)
Henosis: Workload-driven small array consolidation and placement for HDF5 applications on heterogeneous data stores	Donghe Kang, Vedang Patel, Ashwati Nair, Spyros Blanas, Yang Wang, Srinivasan Parthasarathy (The Ohio State University)
DeepHiR : Improving High-radix Router Throughput with Deep Hybrid Memory Buffer Microarchitecture	Cunlu Li, Dezun Dong, Xiangke Liao (NUDT); John Kim, Changhyun Kim (KAIST)

Chairs' Report (6pm to 6:30pm, Thursday, Room 105BC)

Reception (6pm to 8pm, Thursday, Room 105BC)

Friday Breakfast 7:45am to 8am

Machine Learning Acceleration (Plenary Session 6: 8:30am to 11am, Friday, Room 225AB)

Chair: Jim Dehnert, Retired

The Anatomy of Efficient FFT and Winograd Convolutions on Modern CPUs	Aleksandar Zlateski (MIT); Zhen Jia and Kai Li (Princeton University); Fredo Durand (MIT)
Optimizing the Linear Fascicle Evaluation Algorithm for Many-Core Systems	Karan Aggarwal and Uday Bondhugula (Indian Institute of Science)
Deep Reuse: Streamline CNN Inference On the Fly via Coarse-Grained Computation Reuse	Lin Ning and Xipeng Shen (North Carolina State University)
Full-stack Optimization for Accelerating CNNs Using Powers-of-Two Weights with FPGA Validation	Bradley McDanel, Sai Qian Zhang, HT Kung, and Xin Dong (Harvard University)
O3BNN: An Out-Of-Order Architecture for High-Performance Binarized Neural Network Inference with Fine-Grained Pruning	Tong Geng, Tianqi Wang, Chunshu Wu, and Chen Yang (Boston University); Wei Wu (Los Alamos National Laboratory); Ang Li (Pacific Northwest National Laboratory); Martin Herbordt (Boston University)
RFAcc: A 3D ReRAM Associative Array based Random Forest Accelerator	Lei Zhao (University of Pittsburgh); Quan Deng (National University of Defense Technology); Youtao Zhang and Jun Yang (University of Pittsburgh)

Friday Lunch 12:30pm to 2pm

Correctness, Efficiency and Security (Plenary Session 7: 2pm to 3:15pm, Friday, Room 225AB)

Chair: Zhen Jia, Princeton University

BonVoision: Leveraging Spatial Data Smoothness For Recovery From Memory Soft Errors	Bo Fang, Hassan Halawa, Karthik Pattabiraman, and Matei Ripeanu (UBC); Sriram Krishnamoorthy (Pacific Northwest National Lab)
GPUGuard: Mitigating Contention Based Side and Covert Channel Attacks on GPUs	Qiumin Xu (Google); Hoda Naghibijouybari (University of California, Riverside); Shibo Wang (Google); Nael Abu-Ghazaleh (University of California, Riverside); Murali Annavaram (USC)
Dynamically Linked MSHRs for Adaptive Miss Handling in GPUs	Yongbin Gu and Lizhong Chen (Oregon State University)

End of Conference (3:15pm, Friday)