2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS)

IPDPS 2021

Table of Contents

Message from the 2021 General Co-Chairs xx
Message from the 2021 Program Chair xxii
IPDPS 2021 Technical Program xxiii
IPDPS 2021 Organization xxx

Keynote Address 1

A Tale of Two C’s: Convergence and Composability 1
İlkay Altintas (San Diego Supercomputer Center, USA)

Session 1: Performance

Correlation-wise Smoothing: Lightweight Knowledge Extraction for HPC Monitoring Data 2
Alessio Netti (Leibniz Supercomputing Centre, Germany), Daniele Tafani (Fujitsu Enabling Software Technology GmbH, Germany), Michael Ott (Leibniz Supercomputing Centre, Germany), and Martin Schulz (Technical University of Munich, Germany)

Dancing in the Dark: Profiling for Tiered Memory 13
Jinyoung Choi (University of California, Riverside, USA), Sergey Blagodurov (Advanced Micro Devices, Inc., USA), and Hung-Wei Tseng (University of California, Riverside, USA)

Noise-Resilient Empirical Performance Modeling with Deep Neural Networks 23
Marcus Ritter (Technical University of Darmstadt, Germany), Alexander Geiß (Technical University of Darmstadt, Germany), Johannes Wehrstein (Technical University of Darmstadt, Germany), Alexandru Calotoi (ETH Zürich, Switzerland), Thorsten Reimann (Technical University of Darmstadt, Germany), Torsten Hoefer (ETH Zürich, Switzerland), and Felix Wolf (Technical University of Darmstadt, Germany)
Session 2: Linear Algebra

Optimizing Memory-Compute Colocation for Irregular Applications on a Migratory Thread Architecture 58
Thomas B. Rolinger (University of Maryland, USA), Christopher D. Krieger (Laboratory for Physical Sciences, USA), and Alan Sussman (University of Maryland, USA)

TileSpMV: A Tiled Algorithm for Sparse Matrix-Vector Multiplication on GPUs 68
Yu Yao Niu (Super Scientific Software Laboratory, China University of Petroleum-Beijing, China), Zhengyang Lu (Super Scientific Software Laboratory, China University of Petroleum-Beijing, China), Meichen Dong (Super Scientific Software Laboratory, China University of Petroleum-Beijing, China), Zhou Jin (Super Scientific Software Laboratory, China University of Petroleum-Beijing, China), Weifeng Liu (Super Scientific Software Laboratory, China University of Petroleum-Beijing, China), and Guangming Tan (State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, China)

Leveraging PaRSEC Runtime Support to Tackle Challenging 3D Data-Sparse Matrix Problems 79
Qinglei Cao (University of Tennessee, USA), Yu Pei (University of Tennessee, USA), Kadir Akbudak (ASELSAN Research Center, Turkey), George Bosilca (University of Tennessee, USA), Hatem Ltaief (King Abdullah University of Science and Technology (KAUST), Saudi Arabia), David Keyes (King Abdullah University of Science and Technology (KAUST), Saudi Arabia), and Jack Dongarra (University of Tennessee, USA)

Communication-Avoiding and Memory-Constrained Sparse Matrix-Matrix Multiplication at Extreme Scale 90
Md Taufique Hussain (Indiana University Bloomington, USA), Oguz Selvitopi (Lawrence Berkeley National Laboratory, USA), Aydin Buluc (Lawrence Berkeley National Lab, USA), and Ariful Azad (Indiana University Bloomington, USA)

Characterizing Small-Scale Matrix Multiplications on ARMv8-Based Many-Core Architectures 101
Weiling Yang (National University of Defense Technology, China), Jianbin Fang (National University of Defense Technology, China), and Dezun Dong (National University of Defense Technology, China)
Session 3: Scheduling

DAG-Based Scheduling with Resource Sharing for Multi-task Applications in a Polyglot GPU Runtime 111

Alberto Parravicini (Politecnico di Milano, Italy), Arnaud Delamare (Oracle Labs, Switzerland), Marco Arnaboldi (Oracle Labs, Switzerland), and Marco Santambrogio (Politecnico di Milano, Italy)

CTXBack: Enabling Low Latency GPU Context Switching via Context Flashback 121

Zhuoran Ji (The University of Hong Kong, China) and Cho-Li Wang (The University of Hong Kong, China)

Transparent I/O-Aware GPU Virtualization for Efficient Resource Consolidation 131

Nelson Mimura Gonzalez (IBM Thomas J. Watson Research Center, USA) and Tonia Elengikal (IBM Thomas J. Watson Research Center, USA)

Demystifying GPU UVM Cost with Deep Runtime and Workload Analysis 141

Tyler Allen (Clemson University, USA) and Rong Ge (Clemson University, USA)

DUET: A Compiler-Runtime Subgraph Scheduling Approach for Tensor Programs on a Coupled CPU-GPU Architecture 151

Minjia Zhang (Microsoft, USA), Zehua Hu (Beijing University, China), and Mingqin Li (Microsoft, USA)

Session 4: Architecture 1

CAGC: A Content-Aware Garbage Collection Scheme for Ultra-Low Latency Flash-Based SSDs 162

Suzhen Wu (Xiamen University, China), Chunfeng Du (Xiamen University, China), Haijun Li (Xiamen University, China), Hong Jiang (University of Texas-Arlington, USA), Zhirong Shen (Xiamen University, China), and Bo Mao (Xiamen University, China)

NVMe-CR: A Scalable Ephemeral Storage Runtime for Checkpoint/Restart with NVMe-over-Fabrics 172

Shashank Gugnani (The Ohio State University, USA), Tianxi Li (The Ohio State University, USA), and Xiaoyi Lu (University of California, Merced, USA)

Virtual-Link: A Scalable Multi-producer, Multi-consumer Message Queue Architecture for Cross-Core Communication 182

Qinzhe Wu (University of Texas at Austin, USA), Jonathan Beard (Arm Inc., USA), Ashen Ekanayake (University of Texas at Austin, USA), Andreas Gerstlauer (University of Texas at Austin, USA), and Lizy K. John (University of Texas at Austin, USA)

High-Level Synthesis of Parallel Specifications Coupling Static and Dynamic Controllers 192

Vito Giovanni Castellana (Pacific Northwest National Laboratory, USA), Antonino Tumeo (Pacific Northwest National Laboratory, USA), and Fabrizio Ferrandi (Politecnico di Milano, Italy)

RVMA: Remote Virtual Memory Access 203

Ryan E. Grant (Sandia National Laboratories, USA), Michael J. Levenhagen (Sandia National Laboratories, USA), Matthew G.F. Dosanjh (Sandia National Laboratories, USA), and Patrick M. Widener (Sandia National Laboratories, USA)
Session 5: Graph Algorithms

Performance-Portable Graph Coarsening for Efficient Multilevel Graph Analysis 213
Michael S. Gilbert (The Pennsylvania State University, USA), Seher Acer (Sandia National Laboratories, USA), Erik G. Boman (Sandia National Laboratories, USA), Kamesh Madduri (The Pennsylvania State University, USA), and Sivasankaran Rajamanickam (Sandia National Laboratories, USA)

Efficient Distributed Algorithms in the k-Machine Model via PRAM Simulations 223
John Augustine (IIT Madras, India), Kishore Kothapalli (IIIT Hyderabad, India), and Gopal Pandurangan (University of Houston, USA)

Euler Meets GPU: Practical Graph Algorithms with Theoretical Guarantees 233
Adam Polak (Jagiellonian University, Poland), Adrian Siwiec (Jagiellonian University, Poland), and Michal Stobierski (Jagiellonian University, Poland)

MultiLogVC: Efficient Out-of-Core Graph Processing Framework for Flash Storage 245
Kiran Kumar Matam (Facebook Inc., USA), Hanieh Hashemi (University of Southern California, USA), and Murali Annanaram (University of Southern California, USA)

FusedMM: A Unified SDDMM-SpMM Kernel for Graph Embedding and Graph Neural Networks 256
Md. Khaledur Rahman (Indiana University Bloomington), Majedul Haque Sujon (Indiana University Bloomington), and Ariful Azad (Indiana University Bloomington)

Session 6: Resilience

Systemic Assessment of Node Failures in HPC Production Platforms 267
Anwesha Das (North Carolina State University, USA), Frank Mueller (North Carolina State University, USA), and Barry Rountree (Lawrence Livermore National Laboratory, USA)

Combining XOR and Partner Checkpointing for Resilient Multilevel Checkpoint/Restart 277
Masoud Gholami (Zuse Institute Berlin, Germany) and Florian Schintke (Zuse Institute Berlin, Germany)

Demystifying GPU Reliability: Comparing and Combining Beam Experiments, Fault Simulation, and Profiling 289
Fernando Fernandes dos Santos (Universidade Federal do Rio Grande do Sul, Brazil), Siva Kumar Sastry Hari (NVIDIA, USA), Pedro Martins Basso (Universidade Federal do Rio Grande do Sul, Brazil), Luigi Carro (Universidade Federal do Rio Grande do Sul, Brazil), and Paolo Rech (Politecnico di Torino, Italy)

Improving Checkpointing Intervals by Considering Individual Job Failure Probabilities 299
Alvaro Frank (Johannes Gutenberg University Mainz, Germany), Manuel Baumgartner (Johannes Gutenberg University Mainz, Germany), Reza Salkhordeh (Johannes Gutenberg University Mainz, Germany), and André Brinkmann (Johannes Gutenberg University Mainz, Germany)
Covirt: Lightweight Fault Isolation and Resource Protection for Co-Kernels
Nicholas Gordon (University of Pittsburgh, USA) and John Lange
(University of Pittsburgh, USA)

Session 7: Systems 1

Introducing Application Awareness Into a Unified Power Management Stack
Daniel C. Wilson (Boston University, USA), Siddhartha Jana (Intel Corporation, USA), Amritaddha Marathe (Lawrence Livermore National Laboratory, USA), Stephanie Brink (Lawrence Livermore National Laboratory, USA), Christopher M. Cantonu (Intel Corporation, USA), Diana R. Guttmann (Intel Corporation, USA), Brad Geltz (Intel Corporation, USA), Lauren H. Lawson (Intel Corporation, USA), Asma H. Al-rawi (Intel Corporation, USA), Ali Mohammad (Intel Corporation, USA), Fuat Keceli (Intel Corporation, USA), Federico Ardanaz (Intel Corporation, USA), Jonathan M. Eastep (Intel Corporation, USA), and Ayse K. Coskun (Boston University, USA)

PALM: Progress- and Locality-Aware Adaptive Task Migration for Efficient Thread Packing
Jinsu Park (UNIST, Republic of Korea), Seongbeom Park (UNIST, Republic of Korea), Myeonggyun Han (UNIST, Republic of Korea), and Woongki Baek (UNIST, Republic of Korea)

Performance Evaluation of Adaptive Routing on Dragonfly-Based Production Systems
Sudheer Chunduri (Argonne National Laboratory, USA), Kevin Harms (Argonne National Laboratory, USA), Taylor Groves (Lawrence Berkeley National Laboratory, USA), Peter Mendygral (Hewlett Packard Enterprise, USA), Justs Zarins (The University of Edinburgh, Scotland), Michele Weiland (The University of Edinburgh, Scotland), and Yasaman Ghadar (Argonne National Laboratory, USA)

Cori: Dancing to the Right Beat of Periodic Data Movements over Hybrid Memory Systems
Thaleia Dimitra Doudali (Georgia Institute of Technology, USA), Daniel Zahka (Georgia Institute of Technology, USA), and Ada Gavrilovska (Georgia Institute of Technology)

Nowa: A Wait-Free Continuation-Stealing Concurrency Platform
Florian Schmaus (Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany), Nicolas Pfeiffer (Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany), Wolfgang Schröder-Preikschat (Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany), Timo Hönig (Ruhr University Bochum (RUB), Germany), and Jörg Nolte (Brandenburg University of Technology Cottbus-Senftenberg (BTU), Germany)

Session 8: Algorithms 1

Efficient Algorithms for Encrypted All-Gather Operation
Mehran Sadeghi Lahijani (Florida State University, USA), Abu Naser (Florida State University, USA), Cong Wu (Florida State University, USA), Mohsen Gavahi (Florida State University, USA), Viet Tung Hoang (Florida State University, USA), Zhi Wang (Florida State University, USA), and Xin Yuan (Florida State University, USA)
CBNet: Minimizing Adjustments in Concurrent Demand-Aware Tree Networks 382
Otavio Augusto de Oliveira Souza (Universidade Federal de Minas Gerais, Brazil), Olga Goussevskaia (Universidade Federal de Minas Gerais, Brazil), and Stefan Schmid (University of Vienna, Austria)

Scaling Sparse Matrix Multiplication on CPU-GPU Nodes 392
Yang Xia (Ohio State University, USA), Peng Jiang (University of Iowa, USA), Gagan Agrawal (Augusta University, USA), and Rajiv Ramnath (Ohio State University, USA)

zMesh: Exploring Application Characteristics to Improve Lossy Compression Ratio for Adaptive Mesh Refinement 402
Huizhang Luo (New Jersey Institute of Technology, USA), Junqi Wang (Rutgers University-Newark, USA), Qing Liu (New Jersey Institute of Technology, USA), Jieyang Chen (Oak Ridge National Laboratory, USA), Scott Klasky (Oak Ridge National Laboratory, USA), and Norbert Podhorszki (Oak Ridge National Laboratory, USA)

Efficient Parallel CP Decomposition with Pairwise Perturbation and Multi-sweep Dimension Tree 412
Linjian Ma (University of Illinois at Urbana Champaign, USA) and Edgar Solomonik (University of Illinois at Urbana Champaign, USA)

Keynote Address 2
12 Ways to Fool the Masses with Irreproducible Results 422
Lorena Barba (George Washington University, USA)

Best Papers - Plenary
Consistent Lock-free Parallel Stochastic Gradient Descent for Fast and Stable Convergence 423
Karl Bäckström (Chalmers University of Technology, Sweden), Ivan Waluuya (Chalmers University of Technology, Sweden), Marina Papatriantafilou (Chalmers University of Technology, Sweden), and Philippas Tsigas (Chalmers University of Technology, Sweden)

Redesigning Peridigm on SIMT Accelerators for High-Performance Peridynamics Simulations 433
Xinyuan Li (Computer Information Network Center, CAS; University of Chinese Academy of Sciences, China), Huang Ye (Computer Information Network Center, CAS, China), and Jian Zhang (Computer Information Network Center, CAS, China)

Designing High-Performance MPI Libraries with On-the-Fly Compression for Modern GPU Clusters 444
Qinghua Zhou (The Ohio State University, USA), Ching-Hsiang Chu (The Ohio State University, USA), Nithin Senthil Kumar (The Ohio State University, USA), Pouya Kousha (The Ohio State University, USA), Seyedeh Mahdieh Ghazimirsaeed (The Ohio State University, USA), Hari Subramoni (The Ohio State University, USA), and Dhabaleswar K. Panda (The Ohio State University, USA)
Session 9: Programming Models & Compilers

ARBALEST: Dynamic Detection of Data Mapping Issues in Heterogeneous OpenMP Applications 464
Lechen Yu (Georgia Institute of Technology, USA), Joachim Protze (RWTH Aachen University, Germany), Oscar Hernandez (Oak Ridge National Laboratory, USA), and Vivek Sarkar (Georgia Institute of Technology, USA)

Spray: Sparse Reductions of Arrays in OpenMP 475
Jan Hückelheim (Argonne National Laboratory, USA) and Johannes Doerfert (Argonne National Laboratory, USA)

Code Generation for Room Acoustics Simulations with Complex Boundary Conditions 485
Larisa Stoltzfus (University of Edinburgh, United Kingdom), Brian Hamilton (University of Edinburgh, United Kingdom), Michel Steuwer (University of Edinburgh, United Kingdom), Lu Li (University of Edinburgh, United Kingdom), and Christophe Dubach (McGill University, Canada)

Temporal Blocking of Finite-Difference Stencil Operators with Sparse “off-the-grid” Sources 497
George Bisbas (Imperial College London, UK), Fabio Luporini (Devito Codes, UK), Mathias Louboutin (Georgia Institute of Technology, USA), Rhodri Nelson (Imperial College London, UK), Gerard J. Gorman (Imperial College London, UK), and Paul H.J. Kelly (Imperial College London, UK)

Session 10: Algorithms 2

Accelerating non-Power-of-2 size Fourier Transforms with GPU Tensor Cores 507
Louis Pisha (NVIDIA Corporation, USA) and Lukasz Ligowski (NVIDIA Corporation, USA)

Parallel String Graph Construction and Transitive Reduction for De Novo Genome Assembly 517
Giulia Guidi (University of California at Berkeley, USA), Oguz Selvitopi (Lawrence Berkeley National Laboratory, USA), Marquita Ellis (University of California at Berkeley, USA), Leonid Oliker (Lawrence Berkeley National Laboratory, USA), Katherine Yelick (University of California at Berkeley, USA), and Aydin Buluc (University of California at Berkeley, USA)
Distributed-Memory k-mer Counting on GPUs  
Israt Nisa (Lawrence Berkeley National Laboratory, USA), Prashant Pandey (Lawrence Berkeley National Laboratory, USA; University of California, Berkeley), Marquita Ellis (Lawrence Berkeley National Laboratory, USA; University of California, Berkeley), Leonid Oliker (Lawrence Berkeley National Laboratory, USA), Aydin Buluç (Lawrence Berkeley National Laboratory, USA; University of California, Berkeley), and Katherine Yelick (Lawrence Berkeley National Laboratory, USA; University of California, Berkeley)

Distributed-Memory Multi-GPU Block-Sparse Tensor Contraction for Electronic Structure  
Thomas Herault (Innovative Computing Laboratory, the University of Tennessee, USA), Yves Robert (Innovative Computing Laboratory, the University of Tennessee, USA; ENS Lyon, France), George Bosilca (Innovative Computing Laboratory, the University of Tennessee, USA), Robert J. Harrison (IACS, Stony Brook University, USA), Cannada A. Lewis (Sandia National Laboratory, USA), Edward F. Valeev (Department of Chemistry, Virginia Tech, USA), and Jack J. Dongarra (Innovative Computing Laboratory, the University of Tennessee, USA)

Session 11: Systems 2

Adaptive Spatially Aware I/O for Multiresolution Particle Data Layouts  
Will Usher (SCI Institute, University of Utah, USA), Xuan Huang (SCI Institute, University of Utah, USA), Steve Petruzza (Utah State University, USA), Sidharth Kumar (University of Alabama, Birmingham, USA), Stuart R. Slattery (Oak Ridge National Laboratory, USA), Sam T. Reeve (Lawrence Livermore National Laboratory, USA), Feng Wang (SCI Institute, University of Utah, USA), and Valerio Pascucci (SCI Institute, University of Utah, USA)

Interpreting Write Performance of Supercomputer I/O Systems with Regression Models  
Bing Xie (Oak Ridge National Laboratory, USA), Zilong Tan (Carnegie Mellon University, USA), Philip Carns (Argonne National Laboratory, USA), Jeff Chase (Duke University, USA), Kevin Harms (Argonne National Laboratory, USA), Jay Lofstead (Sandia National Laboratories, USA), Sarp Oral (Oak Ridge National Laboratory, USA), Sudharshan S. Vazhkudai (Micron Technology, USA), and Feiyi Wang (Oak Ridge National Laboratory, USA)

Finer-LRU: A Scalable Page Management Scheme for HPC Manycore Architectures  
Jiwoo Bang (Seoul National University, Republic of Korea), Chungyong Kim (Seoul National University, Republic of Korea), Sunggon Kim (Seoul National University, Republic of Korea), Qichen Chen (Seoul National University, Republic of Korea), Cheongjun Lee (Korea Aerospace University, Korea), Eun-Kyu Byun (Korea Institute of Science and Technology Information, Republic of Korea), Jaehwan Lee (Korea Aerospace University, Korea), and Hyeonsang Eom (Seoul National University, Republic of Korea)
Arbitration Policies for On-Demand User-Level I/O Forwarding on HPC Platforms 577
Jean Luca Bez (Federal University of Rio Grande do Sul, Brazil),
Alberto Miranda (Barcelona Supercomputing Center, Spain), Ramon Nou
(Barcelona Supercomputing Center, Spain), Francieli Zanon Boito
(LaBRI, University of Bordeaux, Inria, CNRS, Bordeaux-INP, France),
Toni Cortes (Polytechnic University of Catalonia, Barcelona
Supercomputing Center, Spain), and Philippe O. A. Navaux (Federal
University of Rio Grande do Sul, Brazil)

A Hybrid Scheduling Scheme for Parallel Loops 587
Aaron Handleman (Washington University in St. Louis, USA), Arthur G.
Rattew (Washington University in St. Louis, USA), I-Ting Angelina Lee
(Washington University in St. Louis, USA), and Tao B. Schardl
(Massachusetts Institute of Technology, USA)

Session 12: Neural Networks

EAGLE: Expedited Device Placement with Automatic Grouping for Large Models 599
Hao Lan (University of Toronto, Canada), Li Chen (University of
Louisiana at Lafayette, USA), and Bauchun Li (University of Toronto,
Canada)

BiPS: Hotness-Aware Bi-tier Parameter Synchronization for Recommendation Models 609
Qiming Zheng (Shanghai Jiao Tong University, China), Quan Chen
(Shanghai Jiao Tong University, China), Kaihao Bai (Shanghai Jiao Tong
University, China), Hufeng Guo (Huawei Technologies Ltd, China), Yong
Gao (Huawei Technologies Ltd, China), Xiuzhiang He (Huawei Technologies
Ltd, China), and Minyi Guo (Shanghai Jiao Tong University, China)

DSXplore: Optimizing Convolutional Neural Networks via Sliding-Channel Convolutions 619
Yuke Wang (University of California, Santa Barbara, USA), Boyuan Feng
(University of California, Santa Barbara, USA), and Yufei Ding
(University of California, Santa Barbara, USA)

SUPER: SUb-Graph Parallelism for TransformERs 629
Arpan Jain (The Ohio State University, USA), Tim Moon (Lawrence
Livermore National Laboratory, USA), Tom Benson (Lawrence Livermore
National Laboratory, USA), Hari Subramoni (The Ohio State University,
USA), Sam Adé Jacobs (Lawrence Livermore National Laboratory, USA),
Dhabaleswar K. Panda (The Ohio State University, USA), and Brian Van
Essen (Lawrence Livermore National Laboratory, USA)

Session 13: Federated Learning and Science

Scalable Epidemiological Workflows to Support COVID-19 Planning and Response 639
Dustin Machi (University of Virginia, USA), Parantapa Bhatacharya
(University of Virginia, USA), Stefan Hoops (University of Virginia,
USA), Jiangzhuo Chen (University of Virginia, USA), Henning Mortveit
(University of Virginia, USA), Srinivasan Venkatramanan (University of
Virginia, USA), Bryan Lewis (University of Virginia, USA), Mandy
Wilson (University of Virginia, USA), Arindam Fadikar (Argonne
National Laboratory, USA), Tom Maiden (Pittsburgh Supercomputing
Center, USA), Christopher L. Barrett (University of Virginia, USA),
and Madhav V. Marathe (University of Virginia, USA)
Facilitating Data Discovery for Large-Scale Science Facilities using Knowledge Networks  651
Yubo Qin (Rutgers University, USA), Ivan Rodero (Rutgers University, USA), and Manish Parashar (Rutgers University, USA; University of Utah, Salt Lake City, USA)

Optimal Task Assignment for Heterogeneous Federated Learning Devices  661
Laércio Lima Pilla (Univ. Paris-Saclay, CNRS, Laboratoire de Recherche en Informatique (LRI), France)

Detecting Malicious Model Updates from Federated Learning on Conditional Variational Autoencoder  671
Zhipin Gu (National University of Defense Technology, China) and Yuexiang Yang (National University of Defense Technology, China)

Keynote Address 3

Is Asymptotic Cost Analysis Useful in Developing Practical Parallel Algorithms  681
Guy Blelloch (Carnegie Mellon University, USA)

Keynote Address 4

From Parallelization to Customization – Challenges and Opportunities  682
Jason Cong (University of California, Los Angeles, USA)

Session 14: Algorithms 3

High Performance Streaming Tensor Decomposition  683
Yongseok Soh (University of Oregon, USA), Patrick Flick (Google, Facebook, Microsoft), Xing Liu (Google, Facebook, Microsoft), Shaden Smith (Google, Facebook, Microsoft), Fabio Checconi (Parallel Computing Lab, Intel, USA), Fabrizio Petrini (Parallel Computing Lab, Intel, USA), and Jee Choi (University of Oregon, USA)

Plex: Scaling Parallel Lexing with Backtrack-free Prescanning  693
Le Li (The University of Tokyo, Japan), Shigeyuki Sato (The University of Tokyo, Japan), Qiheng Liu (The University of Tokyo, Japan), and Kenjiro Taura (The University of Tokyo, Japan)

Speculative Parallel Reverse Cuthill-McKee Reordering on Multi-and Many-core Architectures  703
Daniel Mlakar (Graz University of Technology, Austria), Martin Winter (Graz University of Technology, Austria), Mathias Parger (Graz University of Technology, Austria), and Markus Steinberger (Graz University of Technology, Austria)

Jigsaw: A Slice-and-Dice Approach to Non-Uniform FFT Acceleration for MRI Image Reconstruction  714
Brendan L. West (University of Michigan, USA), Jeffrey A. Fessler (University of Michigan, USA), and Thomas F. Wenisch (University of Michigan, USA)
Rank Position Forecasting in Car Racing 724
Bo Peng (Indiana University, USA), Jiayu Li (Indiana University, USA), Selahattin Akkas (Indiana University, USA), Taku Araki (NEC Corporation, Japan), Ohno Yoshiyuki (NEC Corporation, Japan), and Judy Qiu (Indiana University, USA)

Session 15: Cloud Performance

Towards Practical Cloud Offloading for Low-Cost Ground Vehicle Workloads 734
Yuan Xu (Institute of Computing Technology; University of Chinese Academy of Sciences; Peng Cheng Laboratory, China), Tianwei Zhang (Nanyang Technological University, Singapore), Jimin Han (Institute of Computing Technology; University of Chinese Academy of Sciences; Peng Cheng Laboratory, China), Sa Wang (Institute of Computing Technology; University of Chinese Academy of Sciences; Peng Cheng Laboratory, China), and Yungang Bao (Institute of Computing Technology; University of Chinese Academy of Sciences; Peng Cheng Laboratory, China)

Towards Internet-Scale Convolutional Root-Cause Analysis with DiagNet 746
Loïck Bonniot (InterDigital, Univ Rennes, Inria, CNRS, IRISA), Christoph Neumann (InterDigital), and François Tàüani (Univ Rennes, Inria, CNRS, IRISA)

Astra: Autonomous Serverless Analytics with Cost-Efficiency and QoS-Awareness 756
Jananie Jarachanthan (University of Louisiana at Lafayette, USA), Li Chen (University of Louisiana at Lafayette, USA), Fei Xu (East China Normal University, China), and Bo Li (Hong Kong University of Science and Technology, Hong Kong)

Max-Stretch Minimization on an Edge-Cloud Platform 766
Anne Benoit (ENS Lyon, France), Redouane Elghazi (ENS Lyon and Université Franche-Comté, France), and Yves Robert (ENS Lyon, France and University Tennessee Knoxville, USA)

Decentralized Low-Latency Task Scheduling for Ad-Hoc Computing 776
Janick Edinger (University of Hamburg, Germany), Martin Breitbach (University of Mannheim, Germany), Niklas Gabrisch (University of Mannheim, Germany), Dominik Schäfer (University of Mannheim, Germany), Christian Becker (University of Mannheim, Germany), and Amr Rizk (University of Duisburg-Essen, Germany)

Session 16: Systems 3

Lightweight Function Monitors for Fine-Grained Management in Large Scale Python Applications 786
Tim Shaffer (University of Notre Dame, USA), Zhuozhao Li (University of Chicago, USA), Ben Tovar (University of Notre Dame, USA), Yadu Babuji (University of Chicago, USA), TJ Dasso (University of Notre Dame, USA), Zoe Surma (University of Notre Dame, USA), Kyle Chard (University of Chicago; Argonne National Laboratory, USA), Ian Foster (University of Chicago; Argonne National Laboratory, USA), and Douglas Thain (University of Notre Dame, USA)
AlphaR: Learning-Powered Resource Management for Irregular, Dynamic Microservice Graph  .797
Xiaofeng Hou (Shanghai Jiao Tong University), Chao Li (Shanghai Jiao Tong University), Jiacheng Liu (Shanghai Jiao Tong University), Lu Zhang (Shanghai Jiao Tong University), Shaolei Ren (University of California, Riverside), Jingwen Leng (Shanghai Jiao Tong University), Quan Chen (Shanghai Jiao Tong University), and Minyi Guo (Shanghai Jiao Tong University)

Deep Reinforcement Agent for Scheduling in HPC .807.................................................................
Yuping Fan (Illinois Institute of Technology, USA), Zhiling Lan (Illinois Institute of Technology, USA), Taylor Childers (Argonne National Laboratory, USA), Paul Rich (Argonne National Laboratory, USA), William Allcock (Argonne National Laboratory, USA), and Michael E. Papka (Argonne National Laboratory, USA)

F-Write: Fast RDMA-Supported Writes in Erasure-Coded In-Memory Clusters .817......................
Bin Xu (Huazhong University of Sci.& Tech, China), Jianzhong Huang (Huazhong University of Sci.& Tech, China), Qiang Cao (Huazhong University of Sci.& Tech, China), Xiao Qin (Auburn University, USA), and Ping Xie (Qinghai Normal University, China)

Argus: Efficient Job Scheduling in RDMA-Assisted Big Data Processing .827..............................
Sijie Wu (Huazhong University of Science and Technology, China), Hanhua Chen (Huazhong University of Science and Technology, China), Yonghui Wang (Huazhong University of Science and Technology, China), and Hai Jin (Huazhong University of Science and Technology, China)

Session 17: GPU Computing

Scaling Out a Combinatorial Algorithm for Discovering Carcinogenic Gene Combinations to Thousands of GPUs .837..............................................................................................................
Sajal Dash (Oak Ridge National Laboratory, USA), Qais Al-Hajri (Virginia Tech, USA), Wu-chun Feng (Virginia Tech, USA), Harold R. Garrier (Edward Via College of Osteopathic Medicine, USA), and Ramu Anandakrishnan (Edward Via College of Osteopathic Medicine, USA)

A Multi-GPU Design for Large Size Cryo-EM 3D Reconstruction .847........................................
Zihao Wang (Institute of Computing Technology, Chinese Academy of Sciences, China; University of Chinese Academy of Sciences, China), Xiaohua Wan (Institute of Computing Technology, Chinese Academy of Sciences, China), Zhiyong Liu (Institute of Computing Technology, Chinese Academy of Sciences, China), Qianshuo Fan (Huazhong University of Science and Technology, China), Fa Zhang (Institute of Computing Technology, Chinese Academy of Sciences, China), and Guangming Tan (Institute of Computing Technology, Chinese Academy of Sciences, China)
Accelerating Multigrid-Based Hierarchical Scientific Data Refactoring on GPUs

Jieyang Chen (Oak Ridge National Laboratory, USA), Lipeng Wan (Oak Ridge National Laboratory, USA), Xin Liang (Missouri University of Science and Technology, USA), Ben Whitney (Oak Ridge National Laboratory, USA), Qing Liu (New Jersey Institute of Technology, USA), David Pugmire (Oak Ridge National Laboratory, USA), Nicholas Thompson (Oak Ridge National Laboratory, USA), Jong Youl Choi (Oak Ridge National Laboratory, USA), Matthew Wolf (Oak Ridge National Laboratory, USA), Todd Munson (Argonne National Laboratory, USA), Ian Foster (Argonne National Laboratory, USA), and Scott Klasky (Oak Ridge National Laboratory, USA)

Extremely Fast and Energy Efficient One-way Wave Equation Migration on GPU-Based Heterogeneous Architecture

Long Qu (Total, France), Loris Lucido (Eolen, France), Marie Bonnasse-Gahot (Total, France), Pascal Vezolle (IBM, France), and Diego Klahr (Total, USA)

Revisiting Huffman Coding: Toward Extreme Performance on Modern GPU Architectures

Jiannan Tian (Washington State University, USA), Cody Rivera (The University of Alabama, USA), Sheng Di (Argonne National Laboratory, USA), Jieyang Chen (Oak Ridge National Laboratory, USA), Xin Liang (Oak Ridge National Laboratory, USA), Dingwen Tao (Washington State University, USA), and Franck Cappello (Argonne National Laboratory, USA; University of Illinois at Urbana-Champaign, USA)

Session 18: Systems 4

Rack-Scaling: An Efficient Rack-Based Redistribution Method to Accelerate the Scaling of Cloud Disk Arrays

Zhehan Lin (Shanghai Jiao Tong University, China), Hanchen Guo (Shanghai Jiao Tong University, China), Chentao Wu (Shanghai Jiao Tong University, China), Jie Li (Shanghai Jiao Tong University, China), Guangtao Xue (Shanghai Jiao Tong University, China), and Minyi Guo (Shanghai Jiao Tong University, China)

Optimizing Performance for Open-Channel SSDs in Cloud Storage System

Xiaoyi Zhang (Alibaba Group, China), Feng Zhu (Alibaba Group, China), Shu Li (Alibaba Group, China), Kun Wang (Alibaba Group, China), Wei Xu (Alibaba Group, China), and Dengcai Xu (Alibaba Group, China)

AuTraScale: An Automated and Transfer Learning Solution for Streaming System Auto-Scaling

Liang Zhang (Shanghai Jiao Tong University, China), Wenli Zheng (Shanghai Jiao Tong University, China), Chao Li (Shanghai Jiao Tong University, China), Yao Shen (Shanghai Jiao Tong University, China), and Minyi Guo (Shanghai Jiao Tong University, China)

SNOW Revisited: Understanding When Ideal READ Transactions Are Possible

Kishori M. Konwar (RLE, MIT), Wyatt Lloyd (Princeton University, USA), Haonan Lu (Princeton University, USA), and Nancy Lynch (CSAIL, MIT, USA)
Session 19: Algorithms 4

Byzantine Dispersion on Graphs 942
Anisur Rahaman Molla (Indian Statistical Institute Kolkata, India), Kaushik Mondal (Indian Institute of Technology Ropar, India), and William K. Moses Jr. (University of Houston, USA)

Byzantine Agreement with Unknown Participants and Failures 952
Pankaj Khanchandani (ETH Zurich, Switzerland) and Roger Wattenhofer (ETH Zurich, Switzerland)

QPR: Quantizing PageRank with Coherent Shared Memory Accelerators 962
Abdullah T. Mughrabi (North Carolina State University, USA), Mohannad Ibrahim (North Carolina State University, USA), and Gregory T. Byrd (North Carolina State University, USA)

Distributed Training of Embeddings using Graph Analytics 973
Gurbinder Gill (Katana Graph Inc., USA), Roshan Dathathri (Katana Graph Inc., USA), Saeed Maleki (Microsoft Research, USA), Madan Musuvathi (Microsoft Research, USA), Todd Mytkowicz (Microsoft Research, USA), and Olli Saarikivi (Microsoft Research, USA)

Multiplicative Weights Algorithms for Parallel Automated Software Repair 984
Joseph Renzullo (Arizona State University, USA), Westley Weimer (University of Michigan, USA), and Stephanie Forrest (Arizona State University, USA)

Session 20: Deep Neural Networks and Learning

An In-Depth Analysis of Distributed Training of Deep Neural Networks 994
Yun Young Ko (Hanyang University, Korea), Kibong Choi (Hanyang University, Korea), Jiwon Seo (Hanyang University, Korea), and Sang-Wook Kim (Hanyang University, Korea)

Automatic Graph Partitioning for Very Large-Scale Deep Learning 1004
Masahiro Tanaka (National Institute of Information and Communications Technology (NICT), Japan), Kenjiro Taura (University of Tokyo, Japan), Toshihiro Hanawa (University of Tokyo, Japan), and Kentaro Torisawa (National Institute of Information and Communications Technology (NICT), Japan)
Session 21: Architecture 2

SRNoC: A Statically-Scheduled Circuit-Switched Superconducting Race Logic NoC ................................. 1046
George Michelogiannakis (Lawrence Berkeley National Laboratory, USA),
Darren Lyles (Lawrence Berkeley National Laboratory, USA), Patricia
Gonzalez-Guerrer (Lawrence Berkeley National Laboratory, USA), Meriam
Bautista (Lawrence Berkeley National Laboratory, USA), Dilip Vasudevan
(Lawrence Berkeley National Laboratory, USA), and Anastasiia Butko
(Lawrence Berkeley National Laboratory, USA)

Matrix Engines for High Performance Computing: A Paragon of Performance or Grasping at
Straws? ............................................................................................................................................ 1056
Jens Domke (RIKEN Center for Computational Science (R-CCS), Japan),
Emil Vatai (RIKEN Center for Computational Science (R-CCS), Japan),
Aleksandr Drozd (RIKEN Center for Computational Science (R-CCS),
Japan), Peng Chen (National Institute of Advanced Industrial Science
and Technology, Japan), Yosuke Oyama (Tokyo Institute of Technology,
Japan), Lingqi Zhang (Tokyo Institute of Technology, Japan), Shweta
Salaria (RIKEN Center for Computational Science (R-CCS), Japan),
Daichi Mukunoki (RIKEN Center for Computational Science (R-CCS),
Japan), Artur Podobas (KTH Royal Institute of Technology, Sweden),
Mohamed Wahib (National Institute of Advanced Industrial Science and
Technology, Japan), and Satoshi Matsuoka (RIKEN Center for
Computational Science (R-CCS), Japan)

Performance Analysis of Scientific Computing Workloads on General Purpose TEEs ................................. 1066
Ayaz Akram (University of California, Davis, USA), Anna Giannakou
(Lawrence Berkeley National Lab, USA), Venkatesh Akella (University of
California, Davis, USA), Jason Lowe-Power (University of California,
Davis, USA), and Sean Peisert (Lawrence Berkeley National Lab and
University of California, Davis, USA)

High-Performance Spectral Element Methods on Field-Programmable Gate Arrays:
Implementation, Evaluation, and Future Projection ............................................................. 1077
Martin Karp (KTH Royal Institute of Technology, Sweden), Artur Podobas
(KTH Royal Institute of Technology, Sweden), Niclas Jansson (KTH Royal
Institute of Technology, Sweden), Tobias Kenter (Paderborn University,
Germany), Christian Plessl (Paderborn University, Germany), Philipp
Schlatter (KTH Royal Institute of Technology, Sweden), and Stefano
Markidis (KTH Royal Institute of Technology, Sweden)

High-Level FPGA Accelerator Design for Structured-Mesh-Based Explicit Numerical Solvers ........ 1087
Kamalavasan Kamalakkannan (University of Warwick, UK), Gihan R.
Mudalige (University of Warwick, UK), Istvan Z. Reguly (Pazmany Peter
Catholic University, Hungary), and Suhail A. Fahmy (King Abdullah
University of Science and Technology (KAUST), Saudi Arabia)

Author Index