The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York, New York 10121-0710

Copyright © 2007 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: Publications Dept., ACM, Inc. Fax +1 (212) 869-0481 or <permissions@acm.org>.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that has been previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ISBN: 978-1-4503-4361-9

Additional copies may be ordered prepaid from:

ACM Order Department
General Post Office
P.O. Box 30777
New York, NY 10087-0777

Phone: 1-800-342-6626
(US and Canada)
+1-212-626-0500
(all other countries)
Fax: +1-212-944-1318
E-mail: acmhelp@acm.org

Table of Contents

Message from the General Chair ..... **vi**

Message from the Program Chairs ..... **vii**

ICS'16 Conference Committee ..... **ix**

ICS'16 Reviewers ..... **x**

Keynote I: Yale Patt (The University of Texas at Austin): Performance = Bandwidth divided by Latency **1**

Keynote II: Wen-mei Hwu (University of Illinois Urbana-Champaign): Innovative Applications and Technology Pivots – A Perfect Storm in Computing **1**

Session 1: Heterogeneous Systems

Polly-ACC: Transparent compilation to heterogeneous hardware **1-13**
Tobias Grosser and Torsten Hoefler

Hybrid CPU-GPU scheduling and execution of tree traversals **1-12**
Jianqiao Liu, Nikhil Hegde, and Milind Kulkarni

Exploiting Dynamic Reuse Probability to Manage Shared Last-level Caches in CPU-GPU Heterogeneous Processors **1-14**
Siddharth Rai, and Mainak Chaudhuri

Session 2: Power, Energy, Variation

AEQUITAS: Coordinated Energy Management Across Parallel Applications **1-12**
Haris Ribic and Yu David Liu

Runtime-Guided Mitigation of Manufacturing Variability in Power-Constrained Multi-Socket NUMA Nodes **1-12**
Dimitrios Chasapis, Marc Casas, and Miquel Moretó, Martin Schulz, and Eduard Ayguadé, Jesus Labarta, and Mateo Valero

Variation Among Processors Under Turbo Boost in HPC Systems **1-12**
Bilge Acun, Phil Miller, and Laxmikant V. Kale

Session 3: NVMs & Persistent Memory

Mini-Ckpts: Surviving OS Failures in Persistent Memory **1-14**
David Fiala and Frank Mueller, Kurt Ferreira, and Christian Engelmann

High Performance Design for HDFS with Byte-Addressability of NVM and RDMA **1-14**
Nusrat Sharmin Islam, Md. Wasi-ur- Rahman, Xiaoyi Lu, and Dhabaleswar K. (DK) Panda

Write-Aware Management of NVM-based Memory Extensions **1-12**
Amro Awad, Sergey Blagodurov, and Yan Solihin

Session 4: Data Centers

HOPE: Enabling Efficient Service Orchestration in Software-Defined Data Centers..... 1-12
Yang Hu, Chao Li, Longjun Liu, and Tao Li

Towards an Adaptive Multi-Power-Source Datacenter..... 1-11
Longjun Liu, Hongbin Sun, Chao Li, Yang Hu, Nanning Zheng and Tao Li

GreenGear: Leveraging and Managing Server Heterogeneity for Improving Energy Efficiency in Green Data Centers 1-14
Xu Zhou, Haoran Cai, Qiang Cao, Hong Jiang, Lei Tian and Changsheng Xie

Noise Aware Scheduling in Data Centers..... 1-14
Hameedah Sultan, Arpit Katiyar, and Smruti R. Sarangi

Session 5A: GPUs and SIMD

Coherence-Free Multiview: Enabling Reference-Discerning Data Placement on GPU..... 1-13
Guoyang Chen and Xipeng Shen

SFU-Driven Transparent Approximation Acceleration on GPUs..... 1-14
Ang Li, Shuaiwen Leon Song, Mark Wijtvliet, Akash Kumar and Henk Corporaal

Reusing Data Reorganization for Efficient SIMD Parallelization of Adaptive Irregular Applications..... 1-10
Peng Jiang, Linchuan Chen, and Gagan Agrawal

Session 5B: Communication and Coherence

SReplay: Deterministic Sub-Group Replay for One-Sided Communication 1-13
Xuehai Qian, Koushik Sen, Paul Hargrove and Costin Iancu

Lynx: Using OS and Hardware Support for Fast Fine-Grained Inter-Core Communication..... 1-12
Konstantina Mitropoulou, Vasileios Porpodas, Xiaochun Zhang and Timothy M. Jones

Efficient Timestamp-Based Cache Coherence Protocol for Many-Core Architectures..... 1-13
Yuan Yao, Guanhua Wang, Zhiguo Ge, Tulika Mitra , Wenzhi Chen, and Naxin Zhang

Session 6A: Tools and Libraries

BLASX: A High Performance Level-3 BLAS Library for Heterogeneous Multi-GPU Computing..... 1-11
Linnan wang, Wei Wu, Zenglin Xu, Jianxiong Xiao, and Yi Yang

Peruse and Profit: Estimating the Accelerability of Loops..... 1-13
Snehasish Kumar, Vijayalakshmi Srinivasan, and Amirali Sharifian, Nick Sumner, and Arrvindh Shriraman

Simulation and Analysis Engine for Scale-Out Workloads..... 1-13
Nadav Chachmon, Daniel Richins, Robert Cohn and Magnus Christensson, Wenzhi Cui, and Vijay Janapa Reddi

Session 6B: Potpourri

Proteus: Exploiting Numerical Precision Variability in Deep Neural Networks 1-12
Patrick Judd and Jorge Albericio, Tayler Hetherington and Tor Aamodt, and Natalie Enright Jerger and Andreas Moshovos

Galaxyfly: A Novel Family of Flexible-Radix Low-Diameter Topologies for Large-Scales Interconnection Networks 1-12
Fei Lei, Dezun Dong, Xiangke Liao, Xing Su, and Cunlu Li

Replichard: Towards Tradeoff between Consistency and Performance for Metadata 1-11
Zhiying Li, Ruini Xue, and Lixiang Ao

Session 7: Memory

TokenTLB: A Token-Based Page Classification Approach..... 1-13
Albert Esteve, Alberto Ros, and Antonio Robles, Maria Engracia Gómez, and José Duato

Exploiting Private Local Memories to Reduce the Opportunity Cost of Accelerator Integration..... 1-12
Emilio G. Cota, Paolo Mantovani, and Luca P. Carloni

GCaR: Garbage Collection aware Cache Management with Improved Performance for Flash-based SSDs..... 1-12
Suzhen Wu, Yanping Lin, and Bo Mao and Hong Jiang

Session 8: Scheduling

Fairness-oriented OS Scheduling Support for Multicore Systems..... 1-12
Changdae Kim and Jaehyuk Huh

Scheduling Tasks with Mixed Timing Constraints in GPU-Powered Real-Time Systems 1-13
Yunlong Xu, Rui Wang, Tao Li and Mingcong Song, Lan Gao and Zhongzhi Luan, and Depei Qian

CuMAS: Data Transfer Aware Multi-Application Scheduling for Shared GPUs 1-12
Mehmet E Belviranli, Farzad Khorasani, Laxmi N Bhuyan, and Rajiv Gupta

Session 9: Parallelism Issues

DSMR: A Parallel Algorithm for Single-Source Shortest Path Problem..... 1-14
Saeed Maleki, Donald Nguyen and Andrew Lenharth, and Marí´a Garzarán and David Padua

Parallel Transposition of Sparse Data Structures..... 1-13
Hao Wang, Weifeng Liu, and Kaixi Hou and Wu-chun Feng

SARVAVID: A Domain Specific Language for Developing Scalable Computational Genomics Applications..... 1-12
Kanak Mahadik, Christopher Wright, Jinyi Zhang, Milind Kulkarni, Saurabh Bagchi, and Somali Chaterji

Session 10: Multiplication

Fast Multiplication in Binary Fields on GPUs via Register Cache 1-12
Eli Ben Sasson, Matan Hamilis, and Mark Silberstein and Eran Tromer

Balanced Hashing and Efficient GPU Sparse General Matrix-Matrix Multiplication 1-12
Pham Nguyen Quang Anh, Rui Fan, and Yonggang Wen

Optimizing Sparse Matrix-Vector Multiplication for Large-Scale Data Analytics 1-12
Daniele Buono, Fabrizio Petrini, Fabio Checconi, Xing Liu, Xinyu Que, Chris Long, Tai-Ching Tuan

Session 11: Prefetching

TurboTiling: Leveraging prefetching to boost performance of tiled codes 1-12
Sanyam Mehta, Rajat Garg, Nishad Trivedi, and Pen-Chung Yew

Graph Prefetching Using Data Structure Knowledge 1-11
Sam Ainsworth and Timothy M. Jones

Prefetching techniques for near-memory throughput processors 1-14
Reena Panda, Yasuko Eckert, Nuwan Jayasena, Onur Kayiran, and Michael Boyer, and Lizy Kurian John

Session 12: GPU Architecture

Origami: Folding Warps for Energy Efficient GPUs 1-12
Mohammad Abdel-majeed, Daniel Wong, Justin Kuang, and Murali Annavaram

Barrier-Aware Warp Scheduling for Throughput Processors 1-12
Yuxi Liu, Zhibin Yu, Lieven Eeckhout, Vijay Janapa Reddi, Yingwei Luo, Xiaolin Wang, Zhenlin Wang, Chengzhong Xu

Tag-Split Cache for Efficient GPGPU Cache Utilization 1-12
Lingda Li and Ari B. Hayes, Shuaiwen Song, and Eddy Zheng Zhang

ICS 2016 General Chairs' Welcome

On behalf of the ICS organizing committee, we are very pleased to welcome you to Istanbul, Turkey, for the 30th International Conference on Supercomputing (ICS-2016) on June 1-3, 2016. Istanbul is a jewel of a city, with its unique historical depth and splendid natural beauty, blended into a modern metropolis. Each civilization that has made Istanbul its home has left its mark in sublime and splendid ways, and the result is a city that gives one the feeling of universal history at every step from the Roman era to the Byzantine and Ottoman eras.

Visitors will enjoy a wide variety of activities in a city where two continents meet on the blue waters of the Bosphorus to offer an abundance of unique natural, historical, cultural, and culinary experiences. Many of the historical sites including the Blue Mosque, Hagia (St.) Sophia, Hippodrome, Grand Covered Bazaar, Obelisk of Theodosius, Serpentine Column, The Maiden's Tower, Galata, and Basilica Cistern will be explored as part of our social activity. The conference banquet will be held at the Bosphorus, where we will enjoy a view of Asia and Europe simultaneously.

ICS is the premier international forum for the presentation of research results in high-performance computing systems and supercomputing. This year is no different thanks to the organizing committee, reviewers, contributing authors, and particularly the program committee and the chairs, Mahmut Kandemir and Onur Mutlu. The program committee selected an extremely interesting program. Several exciting tutorials will also be part of the conference. We would like to thank all members of the organizing committee and steering committee as well as the technical program committee and the external reviewers.

We would also like to thank the corporate supporters of ICS who have provided generous financial support: ARM and IBM – our primary supporters this year. Our thanks also go to ACM SIGARCH for sponsoring the event.

We hope you find the program exciting, as we have a packed program with 43 technical papers, two keynotes, and a lightning session.

Let us meet where continents meet!

Ozcan Ozturk
ICS 2016 General co-Chair
Bilkent University

Kemal Ebcioglu
ICS 2016 General co-Chair
Global Supercomputing

Message from ICS 2016 Program Chairs

It is our great pleasure to welcome you to the technical program of the 30th International Conference on Supercomputing – ICS 2016! This year's conference, hosted in Istanbul, continues and reinforces the long-standing ICS tradition of being a primary venue for publishing and presenting top-notch research contributions in supercomputing, spanning high-performance computing, programming models, architecture design, system software support, commercial/experimental systems, and much more.

This year's program consists of 43 technical papers and two keynote speeches, by Yale Patt and Wen-mei Hwu. We made two notable changes compared to past ICS meetings. First, the technical program this year includes a lightning session, during which each paper's key idea is to be presented for at most 60 seconds. We believe the benefits of the lightning session are many: it enables all attendees to get an overview of the conference quickly, e.g., to decide which talks to attend; it enables the attendees to get a quick grasp of the key contributions of each paper; and it provides a fair chance to each paper to be presented at the same time when almost all conference attendees are likely to be present in a single session. Second, we reduced the length of each presentation to 16 minutes (plus 4 minutes for questions) so that we could keep the conference mostly as single track.

The 43 papers in this year's program were selected out of 180 submissions (12.5% more than last year). This year shares the third place in number of submissions along with ICS 1999 and ICS 2010. Out of these submissions, 25 were co-authored by PC members.

The Program Committee (PC), consisting of 50 experts, of diverse backgrounds, selected the papers to be presented at the conference. The committee was aided by 29 external reviewers, who provided expert opinions on most papers. In total, 79 reviewers wrote 952 reviews, 899 of which were by PC members. The average number of reviews per paper was 5.3. All submissions received at least 4 reviews, 112 received 5 reviews, and 61 received at least 6 reviews. We, as PC chairs, assigned all reviewers for all papers, except for two, for which we both had conflicts with. The review and decision process for these two submissions were handled by Bruce Childers, to whom our special thanks go.

We briefly describe the paper review and discussion process we used. The authors were given a chance to respond to initial reviews during a 4-day rebuttal period. After we received the initial reviews, we started the online discussion among reviewers to start building consensus. For submissions that received low scores or had low expertise levels during this time, we assigned additional PC reviewers. The online discussion among reviewers continued after the authors submitted rebuttals, which all PC members were advised to read and take into account, until the end of the PC meeting. We, as PC chairs, oversaw and moderated the online discussion of the papers, and divided them into three categories immediately before the PC meeting: online-accept (to be accepted without discussion at the PC meeting), PC-discuss (to be discussed at the PC meeting), likely-reject (likely to be rejected without discussion, unless a PC member requested discussion). 16 papers were accepted online before the PC meeting and did not require discussion, which enabled the PC to focus on papers that needed more discussion.

The Program Committee met in person at the Hilton Chicago O'Hare Airport on March 18, 2016, for a full day, to make accept/reject decisions on papers. 47 PC members were present at the meeting. Before the meeting, we (PC Chairs) used multiple metrics to divide the papers to be discussed into four categories (A1, A2, B, C), to aid the PC discussion, and assigned discussion leads who would start and lead the discussion of each paper. A total of 60 papers were discussed at the meeting. Most of the effort of the committee focused on the 40 papers in the A2 ad B categories, for which had the most diverging opinions among reviewers. 9 of the 10 papers in the A1 category were accepted; none of the 10 papers in the C category were accepted; and 18 of the 40 papers in the A2 and B categories were accepted. During the full-day PC meeting, all papers with a possible path to acceptance were discussed. As is customary, during the discussion of a paper, the PC members who had a conflict of interest left the room.

For each paper discussed during the PC meeting, the accept/reject decision was made through **building consensus across the entire PC**. We did not resort to voting in the accept/reject decision for any paper. In the end, the PC accepted 30 papers unconditionally (of which 16 were online accepts) and 13 papers conditionally. Each conditionally accepted paper was

assigned an anonymous shepherd, who worked with the authors, via the submission software, to ensure the revised paper adequately addressed the critical reviewer feedback. All conditionally-accepted papers were eventually accepted, resulting in an acceptance rate of 24%. The acceptance rates of the PC and non-PC authored papers were very similar.

At the end of the PC meeting, the discussion lead provided an author-visible summary of the PC meeting and online discussion for papers that were discussed but rejected. The goal of this summary is to provide insight and transparency into the reasons as to why the PC thought the paper was not acceptable. The summaries were provided solely for the benefit of the authors, and we believe they are important to keep the conference accept/reject decision process transparent.

We would like to thank the Program Committee members for their hard work, fairness, and collegiality during the entire process. We also would like to thank the external reviewers for their dedicated service. This program could not have been possible without their exceptional diligence and professionalism. We also thank the submission chair, Ashutosh Pattniak, for managing the submission process smoothly and Rachata Ausavarungnirun for helping us during the meeting. Their help has been invaluable.

The technical program of this year's ICS contains the papers accepted, two keynote talks by Yale Patt and Wen-mei Hwu, as well as three tutorials. We would like to thank our keynote speakers for accepting our invitation.

And, finally, we would like to thank all authors and attendees. We believe the quality of the submissions we received enabled us to form a strong and diverse technical program.

We hope your ICS attendance this year will be very rewarding. Please enjoy the technical program along with the amazing city of Istanbul!

Mahmut Kandemir, Pennsylvania State University
Onur Mutlu, ETH Zurich and Carnegie Mellon University
ICS 2016 Program Co-Chairs

ICS 2016 Conference Committee

General Co-Chair:

Dr. Kemal Ebcioglu, Global Supercomputing
Prof. Ozcan Ozturk, Bilkent University

Program Co-Chairs:

Prof. Mahmut Kandemir, Pennsylvania State University
Prof. Onur Mutlu, Carnegie Mellon University

Publicity Chairs:

Prof. Ulya Karpuzcu, University of Minnesota, Twin Cities

Workshops/Tutorials Co-Chair:

Prof. Smail Niar, University of Valenciennes, France
Dr. Alper Buyuktosunoglu, IBM
Dr. Ramon Bertran, IBM

Publications Chair:

Eren Kursun, JPM

Submission Chair:

Ashutosh Pattnaik, Pennsylvania State University

Program Committee:

Meenakshi Arunachalam, Intel
Yungang Bao, ICT-CAS
Marc Casas, BSC, Spain
Nil Chatterjee, Nvidia
Bruce Childers, U. of Pittsburgh
Chita Das, Penn State
Reetuparna Das, U. of Michigan
Luiz DeRose, Cray Inc
Evelyn Duesterwald, IBM Research
Lieven Eeckhout, Ghent U., Belgium
Oguz Ergin, TOBB U., Turkey
Alfredo Goldman, University of Sao Paulo
Ananth Grama, Purdue University
Sudhanva Gurumurthi, IBM/UVa
Hiroshi Inoue, IBM Research
Koji Inoue, Kyushu, Japan
Engin Ipek, U. of Rochester
Canturk Isci, IBM Research
Jose Joao, ARM
Adwait Jog, College of W&M
Myoungsoo Jung, Yonsei U., Korea
Ulya Karpuzcu, U. of Minnesota
Omer Khan, U. of Connecticut
Samira Khan, U. of Virginia
Mike Kozuch, Intel
Sriram Krishnamoorthy, PNNL
Richard Lethin, Reservoir Labs
Tao Li, U. of Florida
Yanjing Li, U. of Chicago
Mieszko Lis, UBC, Canada
Xiaosong Ma, Qatar CRI, Qatar
Scott Mahlke, U. of Michigan
Naoya Maruyama, RIKEn, Japan
Satoshi Matsuoka, Tokyo Tech, Japan
Trevor Mudge, U. of Michigan
Dmitry Ponomarev, U. of Binghamton
Gabriel Rodriguez, U. of Coruna, Spain
Soumyaroop Roy, AMD
P. Sadayappan, Ohio State U.
Xipeng Shen, NCSU
Anand Sivasubramaniam, Penn State U.
Lavanya Subramanian, Intel

Rajeev Thakur, Argonne National Laboratory
Osman Unsal, BSC, Spain
Jeff Vetter, ORNL
Yuan Xie, UCSB
Pen-Chung Yew, Minnesota U.
Zheng Zhang, Rutgers U.
Jishen Zhao, UCSC

Reviewers

Tor Aamodt
Junwhan Ahn
Can Alkan
Mohammad Arjomand
Meenakshi Arunachalam
Guillaume Aupy
Yungang Bao
Yu Cai
Marc Casas
Nil Chatterjee
Nachiappan Chidambaram
Nachiappan
Bruce Childers
Chita Das
Reetuparna Das
Luiz DeRose
Chen Ding
Evelyn Duesterwald
Eiman Ebrahimi
Lieven Eeckhout
Mattan Erez
Oguz Ergin
Alfredo Goldman
Ananth Grama
Sudhanva Gurumurthi
Torsten Hoefler
Hiroshi Inoue
Koji Inoue
Engin Ipek
Canturk Isci
Jose Joao
Adwait Jog
Myoungsoo Jung
Ulya Karpuzcu
Onur Kayiran
Omer Khan
Samira Khan
Hyesoon Kim
Nam Sung Kim
Mike Kozuch
Sriram Krishnamoorthy
Emre Kultursay
Donghyuk Lee
Richard Lethin
Tao Li
Yanjing Li
Mieszko Lis
Xiaosong Ma
Scott Mahlke
Naoya Maruyama
Satoshi Matsuoka
Asit Mishra
Trevor Mudge
Dmitry Ponomarev
Minsoo Rhu
Gabriel Rodriguez
Soumyaroop Roy
Olatunji Ruwase
P. Sadayappan
John Sampson
Vivek Seshadri
Xipeng Shen
Anand Sivasubramaniam
Santhosh Srinath
Lavanya Subramanian
Xiaobai Sun
Rajeev Thakur
Osman Unsal
Jeff Vetter
Yuan Xie
Pen-Chung Yew
Zheng Zhang
Jishen Zhao

Keynote I

Performance = Bandwidth divided by Latency

Dr. Yale Patt

Professor of Electrical and Computer Engineering,
Ernest Cockrell, Jr. Centennial Chair in Engineering, and
University Distinguished Teaching Professor,
The University of Texas at Austin

Abstract: Supercomputing has always required the ultimate in performance, and if you look at the evidence (e.g., the Famous Top 500), you would think "Performance" equals "Bandwidth." Unfortunately, it matters how long it takes to get something done. Otherwise, for example, it would be a win to do a minimal power 64 bit integer add by streaming the bits serially through a full adder and latch in 64 cycles. Clearly, latency matters. In this talk I would like to look at some of the opportunities for decreasing latency, and how the language designer, programmer, compiler writer, and microarchitect can all contribute to this aspect of Supercomputing performance.

Bio: Yale Patt is Professor of Electrical and Computer Engineering and the Ernest Cockrell, Jr. Centennial Chair in Engineering at The University of Texas at Austin. He enjoys equally teaching freshmen, teaching graduate students, and directing the research of six PhD students in high performance computer implementation. He has, for more than 50 years, combined an active research program with extensive consulting and a strong commitment to teaching. The focus of his research is generally five to ten years beyond what industry provides at that point in time. His rationale has always been that he does not do revenue shipments, preferring to produce knowledge that will be useful to future revenue shipments and, more importantly, graduates who will design those future products.

Copyright is held by the author/owner(s).
ICS '16, June 01-03, 2016, Istanbul, Turkey

Keynote II

Innovative Applications and Technology Pivots – A Perfect Storm in Computing

Dr. Wen-mei W. Hwu

Professor and Sanders-AMD chair
University of Illinois, Urbana-Champaign

Abstract: Since early 2000, we have been experiencing two very important developments in computing. One is that a tremendous amount of resources have been invested into innovative applications such as first-principle based models, deep learning and cognitive computing. The other part is that the industry has been taking a technological path where application performance and power efficiency vary by more than two orders of magnitude depending on their parallelism, heterogeneity, and locality. Since then, most of the top supercomputers in the world are heterogeneous parallel computing systems. New standards such as the Heterogeneous Systems Architecture (HSA) are emerging to facilitate software development. Much has been and needs to be learned about of algorithms, languages, compilers and hardware architecture in this movement. What are the applications that continue to drive the technology development? How hard is it to program these systems today? How will we programming these systems in the future? How will innovations in memory devices present further opportunities and challenges? What is the impact on long-term software engineering cost on applications? In this talk, I will present some research opportunities and challenges that are brought about by this perfect storm.

Bio: Wen-mei W. Hwu is a Professor and holds the Sanders-AMD Endowed Chair in the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. He is also CTO of MulticoreWare Inc., chief scientist of UIUC Parallel Computing Institute and director of the IMPACT research group (www.crhc.uiuc.edu/Impact). He directs the UIUC CUDA Center of Excellence and serves as one of the principal investigators of the NSF Blue Waters Petascale supercomputer. For his contributions, he received the ACM SigArch Maurice Wilkes Award, the ACM Grace Murray Hopper Award, the ISCA Influential Paper Award, the IEEE Computer Society B. R. Rau Award and the Distinguished Alumni Award in Computer Science of the University of California, Berkeley. He is a fellow of IEEE and ACM. Dr. Hwu received his Ph.D. degree in Computer Science from the University of California, Berkeley.

Copyright is held by the author/owner(s).
ICS '16, June 01-03, 2016, Istanbul, Turkey