Guest Editors’ Introduction: Special Issue on Applications for the Heterogeneous Computing Era

Pavan Balaji¹ and Satoshi Matsuoka²

As we look beyond the petascale era, accelerators such as graphics processing units (GPUs), field-programmable gate arrays (FPGAs), and Intel many integrated cores (MIC), as well as upcoming integrated hybrid processing cores are expected to play a preeminent role in architecting the largest systems in the world. While there is significant interest in these architectures, much of this interest is an artifact of the hype associated with them. With the rapid proliferation of different forms and flavors of accelerators in the market, understanding their capabilities and shortcomings is an important part of their ability to impact the largest systems in the world.

Despite the importance of such architectures, research in gaining such understanding of these accelerators is still up and coming. This special issue focuses on understanding the implications of accelerators on the architectures and programming environments of future systems. It seeks to ground accelerator research through studies of application kernels or whole applications on such systems, as well as tools and libraries that improve the performance or productivity of applications trying to use these systems.

In this issue

This issue brings forward two papers along this focus area, one on the ability of GPUs in speeding up synthetic aperture radar (SAR) image change detection, and the second on modeling and performance prediction on GPU architectures.

In “Parallel unsupervised SAR image change detection on GPU”, Zhu et al. present a new approach to parallelize SAR image change detection on GPUs. They propose a two-step process that first generates the difference image based on a log-ratio operator, and then detects changes in the difference image using a modified fuzzy c-means clustering algorithm. PLog-FLCM is implemented on AMD Accelerated Parallel Processing SDK based on the Open Computing Language (OpenCL). The authors describe the parallelization characteristics and implementation details of the proposed PLog-FLICM algorithm and demonstrate through a thorough experimental evaluation on several SAR images that the proposed algorithm outperforms other algorithms, and the designed parallel algorithm can greatly reduce the computational time of change detection algorithm. Furthermore, the paper also investigates the performance portability of PLog-FLICM in the different CPU and GPU platforms.

In “Modeling and predicting performance of high performance computing applications on hardware accelerators”, Meswani et al. tackle the complexity associated with porting applications to accelerator architectures by providing a mechanism to predict the performance applications would achieve before porting them to such architectures. Specifically, the authors present a performance-modeling framework that predicts application performance rapidly and accurately for hybrid-core systems. The performance predictions of two full-scale HPC applications, HYCOM and Milc, are also presented in the paper for two forms of accelerators: GPUs and FPGAs.

We hope the articles in this special issue will provide relevant insights into the emerging trends in parallel programming models and systems software for HEC systems.

Pavan Balaji holds appointments as a Computer Scientist and Group Lead at the Argonne National Laboratory, as a Research Fellow of the Computation Institute at the University of Chicago, and as an Institute Fellow of the Northwestern-Argonne Institute of Science and Engineering at Northwestern University. His research interests include parallel programming models and runtime systems for communication and I/O, modern system architecture (multi-core, accelerators, complex memory subsystems, high-speed networks), cloud computing systems, and job scheduling and resource management. He has nearly 100 publications in these areas and has delivered nearly 120 talks and tutorials at various conferences and research institutes. He is a recipient of the U.S. Department of Energy’s Early Career Award. He has also received several other

¹ Argonne National Laboratory, Argonne, IL, USA
² Tokyo Institute of Technology, Tokyo, Japan
awards including the Director’s Technical Achievement award at Los Alamos National Laboratory, an Outstanding Researcher award at the Ohio State University, five best paper awards and various others. He serves as the worldwide chairperson for the IEEE Technical Committee on Scalable Computing (TCSC). He has also served as a chair or editor for nearly 50 journals, conferences and workshops, and as a technical program committee member in numerous conferences and workshops. He is a senior member of the IEEE and a professional member of the ACM. More details about Dr Balaji are available at http://www.mcs.anl.gov/~balaji. Contact him at balaji@mcs.anl.gov.

Satoshi Matsuoka is a Professor at the Global Scientific Information and Computing Center of Tokyo Institute of Technology (GSIC). He is the leader of TSUBAME series of supercomputers, which became the 4th fastest in the world on the Top500 and awarded the “Greenest Production Supercomputer in the World” by the Green 500 in November 2010 and June 2011. He has also co-lead the Japanese national grid project NAREGI during 2003–2007, and is currently leading various projects such as the JST-CREST Ultra Low Power HPC and JSPS Billion-Scale Supercomputer Resilience. He has authored over 500 papers according to Google Scholar, and has chaired many ACM/IEEE conferences, including the Technical Papers Chair, Community Chair, and the upcoming Program Chair for Supercomputing Conferences 2009, 2011 and 2013, respectively. He is a fellow of ACM and European ISC, and has won many awards including the JSPS Prize from the Japan Society for Promotion of Science in 2006, awarded by his Highness Prince Akishinomiya, the ACM Gordon Bell Prizes for 2011, and the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2012.