6-15-2020

Toward High Performance Computing Education

Rajendra K. Raj
Rochester Institute of Technology

Carol J. Romanowski
Rochester Institute of Technology

Sherif G. Aly
American University in Cairo

Brett A. Becker
University College Dublin

Juan Chen
National University of Defense Technology China

See next page for additional authors

Authors

Rajendra K. Raj, Carol J. Romanowski, Sherif G. Aly, Brett A. Becker, Juan Chen, Sheikh Ghafoor, Nasser Giacaman, Steven I. Gordon, Cruz Izu, Shahram Rahimi, Michael P. Robson, and Neena Thota

Toward High Performance Computing Education

Rajendra K. Raj*
Rochester Institute of Technology
Rochester, NY, USA
rkr@cs.rit.edu

Carol J. Romanowski†
Rochester Institute of Technology
Rochester, NY, USA
cjr@cs.rit.edu

Sherif G. Aly
The American University in Cairo
Cairo, Egypt
sgamal@aucegypt.edu

Brett A. Becker
University College Dublin
Dublin, Ireland
brett.becker@ucd.ie

Juan Chen
National University of Defense Technology
Changsha, Hunan, China
juanchen@nudt.edu.cn

Sheikh Ghafoor
Tennessee Tech
Cookeville, TN, USA
sghafoor@tntech.edu

Nasser Giacaman
The University of Auckland
Auckland, New Zealand
n.giacaman@auckland.ac.nz

Steven I. Gordon
The Ohio State University
Columbus, OH, USA
gordon.1@osu.edu

Cruz Izu
The University of Adelaide
Adelaide, SA, Australia
cruz.izu@adelaide.edu.au

Shahram Rahimi
Mississippi State University
Mississippi State, MS, USA
rahimi@cse.msstate.edu

Michael P. Robson
Villanova University
Villanova, PA, USA
michael.robson@villanova.edu

Neena Thota
University of Massachusetts Amherst
Amherst, MA, USA
nthota@cs.umass.edu

ABSTRACT

High Performance Computing (HPC) is the ability to process data and perform complex calculations at extremely high speeds. Current HPC platforms can achieve calculations on the order of quadrillions of calculations per second with quintillions on the horizon. The past three decades witnessed a vast increase in the use of HPC across different scientific, engineering and business communities, for example, sequencing the genome, predicting climate changes, designing modern aerodynamics, or establishing customer preferences. Although HPC has been well incorporated into science curricula such as bioinformatics, the same cannot be said for most computing programs.

This working group will explore how HPC can make inroads into computer science education, from the undergraduate to postgraduate levels. The group will address research questions designed to investigate topics such as identifying and handling barriers that inhibit the adoption of HPC in educational environments, how to incorporate HPC into various curricula, and how HPC can be leveraged to enhance applied critical thinking and problem solving skills. Four deliverables include: (1) a catalog of core HPC educational concepts, (2) HPC curricula for contemporary computing needs, such as in artificial intelligence, cyberanalytics, data science and engineering, or internet of things, (3) possible infrastructures for implementing HPC coursework, and (4) HPC-related feedback to the CC2020 project.

CCS CONCEPTS

• **Social and professional topics** → **Computer science education**; **Computing education**;

KEYWORDS

ITiCSE working group; high performance computing; HPC; high-performance computing curricula; contemporary computing education; computer science education.

ACM Reference Format:
Rajendra K. Raj, Carol J. Romanowski, Sherif G. Aly, Brett A. Becker, Juan Chen, Sheikh Ghafoor, Nasser Giacaman, Steven I. Gordon, Cruz Izu, Shahram Rahimi, Michael P. Robson, and Neena Thota. 2020. Toward High Performance Computing Education. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20), June 15–19, 2020, Trondheim, Norway.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3341525.3394989

1 MOTIVATION

Current High Performance Computing (HPC) resources provide the ability to process data and perform complex calculations at quadrillions of calculations per second, orders of magnitude faster than ordinary high-speed computers [9]. HPC can be performed on dedicated supercomputers typically containing thousands of compute nodes working together to complete one or more tasks in parallel, or in recent years, large numbers of inexpensive commodity computers configured in parallel or distributed settings, with

*Working Group Leader
†Working Group Co-Leader

ITiCSE '20, June 15–19, 2020, Trondheim, Norway

ACM ISBN 978-1-4503-6874-2/20/06.
https://doi.org/10.1145/3341525.3394989

the Hadoop ecosystem representing an open-source archetype [1]. Further, "virtual" supercomputers consisting of distributed, interconnected cloud resources have also been leveraged for extremely low-cost and on-demand HPC.

The exascale era, with quintillions of calculations per second, promises to revolutionize many aspects of not just computing, but science in general [13]. As Gray noted [5], scientific discovery has moved from the empirical and theoretical paradigms through computational modeling/simulation to the fourth paradigm of data-intensive exploration, for which HPC is the keystone. Only HPC strategies [7] make it possible to have modern climate prediction, healthcare, structural designs, aerodynamic designs, management of natural resources, arts and entertainment, social interaction, just to name a few. Many people, including many computing educators, don't realize how much HPC impacts their daily lives.

More specifically, HPC is essential for today's "hot" computing areas such as Artificial Intelligence (AI), Cyberanalytics (CA), Data Science and Engineering (DSE) [12], and the Internet of Things (IoT). The symbiosis between AI and HPC is demonstrated by the HPC community's support for the anticipated AI revolution and the concomitant expectations that AI will drive improvements in HPC hardware and algorithms [4]. Baz [2] makes a strong case, with examples, for HPC in IoT while Perrin et al. [11] make a similar case for HPC in data analytics. In fact, one could argue legitimately that deep learning would not be feasible without the availability of different HPC frameworks [14]. The use of HPC and AI for Cyberanalytics has also become commonplace [8].

HPC education presents unique challenges. Almost all HPC applications implement some form of parallelism that is an inherently difficult concept to grasp. Additionally, a substantial portion of HPC education seems ad hoc, and would benefit from a more structured approach. Finally, new approaches are needed to advance the education of disciplinary specialists [7]. To progress toward these goals, the working group will investigate how HPC education can be delivered to the next generation of faculty and students.

2 WORKING GROUP OBJECTIVES

There have been prior efforts in bringing HPC into computing education [6, 7]. An ITiCSE 2010 working group [3] also addressed questions such as what aspects of parallelism should students learn; how these concepts and practices could be incorporated into the computing curriculum; what resources might be needed; and what systemic obstacles exist to prevent adoption. HPC has been well-incorporated into many bioinformatics and computational biology curricula [10], but not uniformly or adequately into computing programs as yet.

This working group will extend earlier work [6, 7] by addressing the following: What efforts are needed to prepare computing and other STEM researchers, educators, and practitioners to include HPC coursework as part of the curriculum? How can educators scale-up or scale-out successful HPC education approaches? What obstacles and barriers inhibit HPC acceptance and how might educators overcome them? How can HPC help to enhance applied critical thinking and problems-solving skills of computing students?

Building on the background and motivation already mentioned, this working group seeks to:

(1) Catalog HPC elements such as multi- and many-core programming; distributed memory, shared memory, and hybrid models; accelerators including graphical processing units, FPGAs, Xeon Phi and Quantum processing units; parallel and distributed file systems, and supercomputers.
(2) For each of the contemporary application areas (such as such as AI, DSE or IoT), catalog best practices and examples of how educators can incorporate HPC.
(3) Explore infrastructures for implementing HPC coursework with a focus on simple and low-cost solutions.
(4) Develop HPC recommendations for modern computing curricula and provide feedback to the CC2020 project.

ACKNOWLEDGMENTS

John Impagliazzo contributed substantially to this paper. Raj acknowledges support provided by the National Science Foundation under Awards 1433736 and 1922169.

REFERENCES

[1] Apache Software Foundation. 2019. Apache Hadoop. https://hadoop.apache.org.
[2] Didier El Baz. 2014. IoT and the Need for High Performance Computing. In *2014 International Conference on Identification, Information and Knowledge in the Internet of Things*. IEEE, Beijing, China, 6. https://doi.org/10.1109/IIKI.2014.8
[3] Richard Brown, Elizabeth Shoop, Joel Adams, Curtis Clifton, Mark Gardner, Michael Haupt, and Peter Hinsbeeck. 2010. Strategies for Preparing Computer Science Students for the Multicore World. In *Proceedings of the 2010 ITiCSE Working Group Reports (ITiCSE-WGR '10)*. Association for Computing Machinery, New York, NY, USA, 97–115. https://doi.org/10.1145/1971681.1971689
[4] Rob Farber. 2017. *AI-HPC is Happening Now*. Technical Report. insideHPC, USA. https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/the-convergence-of-ai-and-hpc.pdf.
[5] Tony Hey, Stewart Tansley, and Kristin M. Tolle (Eds.). 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, Washington. http://research.microsoft.com/en-us/collaboration/fourthparadigm/
[6] Scott Lathrop. 2016. A Call to Action to Prepare the High-Performance Computing Workforce. *Computing in Science Engineering* 18, 6 (Nov 2016), 80–83. https://doi.org/10.1109/MCSE.2016.101
[7] Scott Lathrop and Thomas Murphy. 2008. High-Performance Computing Education. *Computing in Science Engineering* 10, 5 (Sep. 2008), 9–11. https://doi.org/10.1109/MCSE.2008.132
[8] Raffael Marty. 2013. Cyber Security: How Visual Analytics Unlock Insight. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. Association for Computing Machinery, New York, NY, USA, 1139. https://doi.org/10.1145/2487575.2491132
[9] NetApp, Inc. 2019. What Is High-Performance Computing? https://www.netapp.com/us/info/what-is-high-performance-computing.aspx.
[10] Esteban Perez-Wohlfeil, Oscar Torreno, Louisa J. Bellis, Pedro L. Fernandes, Brane Leskosek, and Oswaldo Trelles. 2018. Training Bioinformaticians in High Performance Computing. *Heliyon* 4, 12 (2018), 18. https://doi.org/10.1016/j.heliyon.2018.e01057
[11] Dimitri Perrin, Marija Bezbradica, Martin Crane, Heather J. Ruskin, and Christophe Duhamel. 2012. High-Performance Computing for Data Analytics. In *2012 IEEE/ACM 16th International Symposium on Distributed Simulation and Real Time Applications*. IEEE, New York, 234–242. https://doi.org/10.1109/DS-RT.2012.41
[12] Rajendra K. Raj, Allen Parrish, John Impagliazzo, Carol J. Romanowski, Sherif G. Aly, Casey C. Bennett, Karen C. Davis, Andrew McGettrick, Teresa Susana Mendes Pereira, and Lovisa Sundin. 2019. An Empirical Approach to Understanding Data Science and Engineering Education. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE-WGR '19)*. Association for Computing Machinery, New York, NY, USA, 73–87. https://doi.org/10.1145/3344429.3372503
[13] Daniel A. Reed and Jack Dongarra. 2015. Exascale Computing and Big Data. *Commun. ACM* 58, 7 (June 2015), 56–68. https://doi.org/10.1145/2699414
[14] Shayan Shams, Richard Platania, Kisung Lee, and Seung-Jong Park. 2017. Evaluation of Deep Learning Frameworks Over Different HPC Architectures. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, New York, 1389–1396. https://doi.org/10.1109/ICDCS.2017.259