Lecture Notes in Computer Science **10724**

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7407

Stephen Jarvis · Steven Wright
Simon Hammond (Eds.)

High Performance Computing Systems

Performance Modeling, Benchmarking, and Simulation

8th International Workshop, PMBS 2017
Denver, CO, USA, November 13, 2017
Proceedings

Springer

Editors
Stephen Jarvis
University of Warwick
Coventry
UK

Simon Hammond
Sandia National Laboratories
Albuquerque, NM
USA

Steven Wright
University of Warwick
Coventry
UK

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-72970-1 ISBN 978-3-319-72971-8 (eBook)
https://doi.org/10.1007/978-3-319-72971-8

Library of Congress Control Number: 2017962895

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Special Issue on the 8th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computing Systems (PMBS 2017)

This volume contains the 13 papers that were presented at the 8th International Workshop on Performance Modeling, Benchmarking, and Simulation of High Performance Computing Systems (PMBS 2017), which was held as part of the 29th ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC 2017) at the Colorado Convention Centre in Denver between 12–17 November 2017. SC offers a vibrant technical program, which includes technical papers, tutorials in advanced areas, Birds of a Feather sessions (BoFs), panel debates, a doctoral showcase, and a number of technical workshops in specialist areas (of which PMBS is one). The focus of PMBS is comparing high performance computing systems through performance modeling, benchmarking, or the use of tools such as simulators. Contributions are sought in areas including: performance modeling and analysis of applications and high performance computing systems; novel techniques and tools for performance evaluation and prediction; advanced simulation techniques and tools; micro-benchmarking, application benchmarking, and tracing; performance-driven code optimization and scalability analysis; verification and validation of performance models; benchmarking and performance analysis of novel hardware; performance concerns in software/hardware co-design; tuning and auto-tuning of HPC applications and algorithms; benchmark suites; performance visualization; real-world case studies; studies of novel hardware such as Intel's Knights Landing platform and NVIDIA Pascal GPUs.

The 8th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computing Systems (PMBS 2017) was held on November 13 as part of the 29th ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC 2017) at the Colorado Convention Center in Denver during November 12–17, 2017.

The SC conference is the premier international forum for high performance computing, networking, storage, and analysis. The conference is unique in that it hosts a wide range of international participants from academia, national laboratories, and industry; this year's conference attracted over 13,000 attendees and featured over 350 exhibitors in the industry's largest HPC technology fair.

This year's conference was themed "HPC Connects," encouraging academia and industry to come together to inspire new collaborations between different fields of science, with the goal of bringing about an impact on society and the changing nature of our world.

SC offers a vibrant technical program, which includes technical papers, tutorials in advanced areas, Birds of a Feather sessions (BoFs), panel debates, a doctoral showcase, and a number of technical workshops in specialist areas (of which PMBS is one).

The focus of the PMBS 2017 workshop was comparing high performance computing systems through performance modeling, benchmarking, or the use of tools such as simulators. We were particularly interested in receiving research papers that reported on the ability to measure and make trade-offs in hardware/software co-design to improve sustained application performance. We were also keen to capture the assessment of future systems, for example, through work that ensured continued application scalability through peta- and exa-scale systems.

Like SC 2017, the aim of the PMBS 2017 workshop was to bring together researchers from industry, national labs, and academia, who are concerned with the qualitative and quantitative evaluation and modeling of high performance computing systems. Authors were invited to submit novel research in all areas of performance modeling, benchmarking, and simulation, and we welcomed research that combined novel theory and practice. We also expressed an interest in submissions that included analysis of power consumption and reliability, and were receptive to performance modeling research that made use of analytical methods as well as those based on tracing tools and simulators.

Technical submissions were encouraged in areas including: performance modeling and analysis of applications and high performance computing systems; novel techniques and tools for performance evaluation and prediction; advanced simulation techniques and tools; micro-benchmarking, application benchmarking, and tracing; performance-driven code optimization and scalability analysis; verification and validation of performance models; benchmarking and performance analysis of novel hardware; performance concerns in software/hardware co-design; tuning and auto-tuning of HPC applications and algorithms; benchmark suites; performance visualization; real-world case studies; and studies of novel hardware such as the Intel's Knights Landing platform and NVIDIA Pascal GPUs.

PMBS 2017

We received a good number of submissions for this year's workshop. This meant that we were able to be selective in those papers that were chosen; the acceptance rate for papers was approximately 35%. The resulting papers show worldwide programs of research committed to understanding application and architecture performance to enable exascale computational science.

The workshop included contributions from Argonne National Laboratory, Brookhaven National Laboratory, Clemson University, École Normale Supérieure de Lyon, Edinburgh Parallel Computing Centre, ENS Lyon, Florida State University, Hewlett Packard Labs, Inria, Lawrence Berkley National Laboratory, Los Alamos National Laboratory, New Mexico State University, NVIDIA Corporation, Pacific Northwest National Laboratory, Pazmany Peter Catholic University, Universidade de Lisboa, University of Basel, University of Bristol, University at Buffalo, University of Cambridge, University of Chicago, University of Florida, University of Tennessee, University of Udine, University of Warwick, and Vanderbilt University.

Several of the papers are concerned with "Performance Evaluation and Analysis" (see Section A). The paper by Nathan Tallent et al. discusses the performance differences between PCIe- and NVLink-connected GPU devices on deep learning workloads. They demonstrate the performance advantage of NVLink over PCIe- connected GPUs. Balogh et al. provide a comprehensive survey of parallelization approaches, languages and compilers for unstructured mesh algorithms on GPU architectures. In particular, they show improvements in performance for CUDA codes when using the Clang compiler over NVIDIA's own nvcc. Guillaume Aupy and colleagues exploit the periodic nature of I/O in HPC applications to develop efficient scheduling strategies. Using their scheduling strategy they demonstrate a 32% increase in throughput on the Mira system. Finally, Romero et al. document their porting of the PWscf code to multi-core and GPU systems decreasing time-to-solution by 2–3×.

Section B of the proceedings collates papers concerned with "Performance Modeling and Simulation." Nicolas Denoyelle et al. present the cache-aware roofline model (CARM) and validate the model on a Xeon Phi Knights Landing platform. Similarly, Chennupati et al. document a scalable memory model to enable CPU performance prediction. Mollah et al. examine universal globally adaptive load-balanced routing algorithms on the Dragonfly topology. Their performance model is able to accurately predict the aggregate throughput for Dragonfly networks. Cavelan et al. apply algorithm-based focused recovery (ABFR) to N-body computations. They compare this approach with the classic checkpoint/restart strategy and show significant gains over the latter. Zhang et al. propose a multi-fidelity surrogate modeling approach, using a combination of low-fidelity models (mini-applications) and a small number of high fidelity models (production applications) to enable faster application/architecture co-design cycles. They demonstrate an improvement over using either low-fidelity models or high-fidelity models alone. Finally, Simakov and colleagues document their development of a simulator of the Slurm resource manager. Their simulation is able to use historical logs to simulate different scheduling algorithms to identify potential optimizations in the scheduler.

The final section of the proceedings, Section C, contains the three short papers presented at PMBS. The paper by Yoga et al. discusses their extension to the Gen-Z communication protocol in the structural simulation toolkit, enabling source-code attribution tagging in network packets. Tyler Allen and colleagues at the Lawrence Berkley National Laboratory, conduct a performance and energy survey for NERSC workloads on Intel KNL and Haswell architectures. The final paper in this volume, by Turner and McIntosh-Smith, presents a survey of application memory usage on the ARCHER national supercomputer.

The PMBS 2017 workshop was extremely well attended and we thank the participants for the lively discussion and positive feedback received throughout the workshop. We hope to be able to repeat this success in future years.

The SC conference series is sponsored by the IEEE Computer Society and the ACM (Association for Computing Machinery). We are extremely grateful for the support we received from the SC 2017 Steering Committee, and in particular from Almadena Chtchelkanova and Luiz DeRose, the workshop chair and vice chair.

The PMBS 2017 workshop was only possible thanks to significant input from AWE in the UK, and from Sandia National Laboratories and the Lawrence Livermore

National Laboratory in the USA. We acknowledge the support of the AWE Technical Outreach Program (project CDK0724).

We are also grateful to LNCS for their support, and to Alfred Hofmann and Anna Kramer for assisting with the production of this issue.

November 2017

Stephen A. Jarvis
Steven A. Wright
Simon D. Hammond

Organization

Workshop Chairs

Stephen Jarvis	University of Warwick, UK
Steven Wright	University of Warwick, UK
Simon Hammond	Sandia National Laboratories (NM), USA

Workshop Technical Program Committee

Reid Atcheson	Numerical Algorithms Group Ltd., UK
Pavan Balaji	Argonne National Laboratory, USA
Prasanna Balaprakash	Argonne National Laboratory, USA
David Beckingsale	Lawrence Livermore National Laboratory, USA
Abhinav Bhatele	Lawrence Livermore National Laboratory, USA
Robert Bird	Los Alamos National Laboratory, USA
Richard Bunt	ARM Ltd., UK
Cristopher Carothers	Rensselaer Polytechnic Institute, USA
Patrick Carribault	CEA, France
Aurélien Cavelan	University of Basel, Switzerland
Raphaël Couturier	L'université Bourgogne, Franche-Comté, France
Todd Gamblin	Lawrence Livermore National Laboratory, USA
Wayne Gaudin	NVIDIA, UK
Paddy Gillies	European Centre for Medium-Range Weather Forecasts, UK
Jeff Hammond	Intel Corporation, USA
Andreas Hansson	ARM Ltd., UK
Andy Herdman	UK Atomic Weapons Establishment, UK
Thomas Ilsche	Technische Universität Dresden, Germany
Nikhil Jain	Lawrence Livermore National Laboratory, USA
Guido Juckeland	Helmholtz-Zentrum Dresden-Rossendorf, Germany
Michael Klemm	Intel Corporation, Germany
Andrew Mallinson	Intel Corporation, UK
Satheesh Maheswaran	UK Atomic Weapons Establishment, UK
Simon McIntosh-Smith	Bristol University, UK
Branden Moore	Sandia National Laboratores (NM), USA
Misbah Mubarak	Argonne National Laboratory, USA
Gihan Mudalige	University of Warwick, UK
Elmar Peise	AICES, RWTH Aachen, Germany
John Pennycook	Intel Corporation, USA
Karthik Raman	Intel Corporation, USA
István Reguly	Pázmány Péter Catholic University, Hungary
Jose Cano Reyes	University of Edinburgh, UK

Yves Robert	ENS Lyon, France
Stephen Roberts	ARM Ltd., UK
Arun Rodrigues	Sandia National Laboratories (NM), USA
Fabio Schifano	Università di Ferrara, Italy
Andrey Semin	Intel Corporation, Germany
Govind Sreekar Shenoy	University of Edinburgh, UK
Thomas Steinke	Zuse Institute Berlin, Germany
Peter Strazdins	Australian National University, Australia
Christian Trott	Sandia National Laboratories (NM), USA
Alejandro Valero	University of Zaragoza, Spain
Yunquan Zhang	Chinese Academy of Sciences, China

Contents