Report from Dagstuhl Perspectives Workshop 15342

Power-Bounded HPC Performance Optimization

Edited by

Dieter Kranzlmüller[1] and Barry L. Rountree[2]

1 **Ludwig-Maximilians-Universität (LMU) & Leibniz Supercomputing Centre (LRZ), München, Germany,** `kranzlmueller@lrz.de`

2 **Lawrence Livermore National Laboratory (LLNL), Livermore, USA,** `rountree@llnl.gov`

Abstract

This report documents the program and the outcomes of Dagstuhl Perspectives Workshop 15342 "Power-Bounded HPC Performance Optimization". The workshop consists of two parts. In part one, our international panel of experts in facilities, schedulers, runtime systems, operating systems, processor architectures and applications provided thought-provoking and details insights into open problems in each of their fields with respect to the workshop topic. These problems must be resolved in order to achieve a useful power-constrainted exascale system, which operates at the highest performance within a given power bound. In part two, the participants split up in three groups, trying to address certain specific subtopics as identified during the expert plenaries. These subtopics have been discussed in more detail, followed by plenary sessions to compare and synthesize the findings into an overall picture. As a result, the workshop identified three major problems, which need to be solved on the way to power-bounded HPC performance optimization.

Perspectives Workshop August 16–21, 2015 – http://www.dagstuhl.de/15342

1998 ACM Subject Classification C.5.1 Super (Very Large) Computers, C.4 Performance of Systems, C.1.4 Parallel Architectures

Keywords and phrases Exascale computing, Energy efficiency, Power awareness, Scalability

Digital Object Identifier 10.4230/DagRep.5.8.1

1 Executive Summary

Barry L. Rountree
Dieter Kranzlmüller

The Dagstuhl Perspectives Workshop 15342 "Power-Bounded HPC Performance Optimization" has been an interesting experience, as in contrast to other workshops, we focused on the unknown characteristics of future exascale systems rather than on the state-of-the-art of todays petascale architectures. In order to do this, a large fraction of the workshop was spent on in-depth discussions in three working groups, while plenary sessions served to provide impulses on specific topics and to synthesize the findings of the breakout sessions. The key ingredient of this workshop has been the interaction between the particpants, leading to several new collaborations across vendors, national laboratories and academia.

The key findings of the workshop can be identified as follows:

- Power-bound performance optimization has different objectives according to the respective targets and operational goals. While infrastructure providers are often bound to a specific

spending, users want to utilize a resource at the maximum of its capabilities. As a result, any power-bound optimization must address multiple criteria, and the solution is rarely straight-forward but specific for a given setting.
- The currently available information on each layer of the computing environment is insufficient. Both, the availability of information with respect to its power characteristics, as well as the exchange between different layers, needs to be improved in order to optimize the operation of infrastructures and the execution of applications on a given system.
- Due to the number of dependencies, any optimization needs to find a good balance between "user happiness", total costs, and performance. These characteristics are imporant for both, providers and users, and a careful balancing strategy needs to be implemented without harming any interests of the actors too much.

The discussions at the Dagstuhl Perspectives Workshop have lead to the identification of a number of technical problems, which need to be addressed in the near future before achieving optimal results in a power-bound environment. As a conclusion, the participants agreed that a strategic and tactical agenda is needed, which identifies the individual problems and technologies as well as their interconnections, such that future systems can utilize this knowledge for new approaches of power-bound HPC performance optimization. The results of this investigations should be made available as a white book, which describes the strategy for future exascale systems.

2 Table of Contents

Executive Summary
Barry L. Rountree and Dieter Kranzlmüller 1

Overview of Talks

Introductory Remarks & Motivation
Barry L. Rountree 4

Musings on Power, Programming Models, and Applications
David Richards 4

Open Problems in Processor Architecture
Jonathan Eastep 4

Performance Optimization vs. Power – Experiences with Petascale Earthquake Simulations on SuperMUC
Michael Bader 5

HPC Data Center Infrastructure Challenges Under A Power Bound
Torsten Wilde 5

Future Directions in System Software on Power-Bounded Supercomputers
David K. Lowenthal 6

Plenary and Breakout Sessions 6

Open Problems – Future Research Direction 6

Participants 8

3 Overview of Talks

3.1 Introductory Remarks & Motivation

Barry L. Rountree (LLNL – Livermore, US)

The US Department of Energy and other supercomputing stakeholders believe that future high performance machines will be power limited, with a bound of 20 Megawatts suggested for the first exascale system. In this workshop we will explore the implications for power-limited computing, focusing primarily on optimization strategies. In particular, we will be making the distinction between energy-limited and power-limited systems, and discussing how hardware overprovisioning can increase performance for both.

3.2 Musings on Power, Programming Models, and Applications

David Richards (LLNL – Livermore, US)

In this talk we discuss how programming models for HPC applications might contribute to power optimization. Task-bases and other asynchronous programming models offer some hope in this regard as they expose concurrency as data dependencies in ways that might allow a runtime system to reorder or otherwise manage work to satisfy a power constraint. Unfortunately, no such models are production ready and it is unclear whether these models will be able to match the performance of more traditional HPC programming models. We discuss what developers might be willing to do so support power optimization. Finally, we hint at some of the challenges a runtime system might face in optimizing power by examining three examples of complex load imbalances that occur in real problems.

3.3 Open Problems in Processor Architecture

Jonathan Eastep (Intel – Hillsboro, US)

In this talk, I will discuss approaches for improving processor efficiency and tailoring processor architectures to work better with runtimes for optimizing system performance within a power bound. Approaches will include hardware-acceleration of basic building blocks of HPC codes, the addition of a 16-bit floating point format in SIMD units for the purpose of trading computational accuracy for additional performance, and optimization of the processor pipeline depth, transistor power-performance characteristics, and static power consumption to increase the utility of hardware over-provisioning strategies.

3.4 Performance Optimization vs. Power – Experiences with Petascale Earthquake Simulations on SuperMUC

Michael Bader (TU München, DE)

SeisSol is a high-order discontinuous-Galerkin software to simulate dynamic rupture and wave propagation processes during earthquakes. Working on unstructured meshes and applying static load distribution, it is representative for a large range of current simulation software. The high arithmetic intensity of its high-order discretization in space and time also make it an attractive candidate for peta- and maybe even exascale simulations. The talk specifically focused on power questions and issues that might arise with SeisSol simulations in the future:

- Measurements of the power consumption of SeisSol on the latest range of Intel CPU architectures revealed that optimising for time-to-solution implies improved energy-to-solution, already. Open questions include how to balance energy and time to solution in the choice of clock frequency and other hardware parameters, and how programmers could and should support runtime systems in this aspect.
- During the first petascale runs on the SuperMUC machine, SeisSol experienced machine crashes due to problems with the global power infrastructure, which were tracked down to strong variations and peaks in power consumption. As respective problems are likely to increase for exascale machines, will there be consequences for the software stack or even application programmers?
- Current processors already feature variations in their power consumption due to tolerances in the manufacturing process. Will such changes directly turn into performance variations in a power-bounded setup?

Will this make static load distribution, as currently applied in SeisSol, unfeasible on future machines? To conclude, the characterisation of the performance of simulation software will need to consider various quality numbers, especially time and energy to solution, and will open the question on how simulation software may interact with operating and runtime systems to mitigate power issues.

3.5 HPC Data Center Infrastructure Challenges Under A Power Bound

Torsten Wilde (LRZ – München, DE)

The era of energy efficient high performance computing (HPC) does not only create challenges for application developers and system software developers but also for the cooling infrastructure of HPC data centers. The move from air cooling to a mix of cooling technologies (air, indirect/direct water cooling, chiller supported and chiller-less cooling) in the data center coupled with the increasing dynamic power behavior of HPC systems makes the energy efficient operation of a data center nontrivial. This talk highlights current control challenges in the data center cooling infrastructure, using the LRZ data center as an example, and discusses how a power bound might help to improve the data center energy efficiency. We make the case that an adjustable (flexible) power bound might be beneficial in light of: the possibility of integrating renewable energy (mainly solar and wind power); changing electricity costs when buying energy at the energy market; and changing outside conditions that effect the coefficient of performance (COP) of the data center cooling infrastructure. We discuss how a power bound can affect the four pillars (data center infrastructure, HPC system hardware, HPC system software, HPC applications) of the "4 Pillar model for energy efficient HPC data centers" and show that some connecting between all four pillars might be required. Finally, a concrete example of the possible benefit of a power bound is shown using data of the LRZ data center.

3.6 Future Directions in System Software on Power-Bounded Supercomputers

David K. Lowenthal (University of Arizona – Tucson, US)

System software play a significant role in power-bounded supercomputers. This talk covers possible future directions in system software.

4 Plenary and Breakout Sessions

The workshop was divided into a plenary track with the above mentioned keynote speeches, each covering specific aspects of the problem domain, and a series of breakout sessions, where the participants discussed specific exercises more detailled in three groups. The groups were composed of a selected set of researchers, ensuring a good mixture of seniority and juniority, as well as a coverage of all the aspects required to address the problems at hand. The breakouts were designed as competitions between the group, whose results were evaluated in the follow-up plenary sessions.

The three group leaders were:
- David Lowenthal
- Frank Mueller
- Martin Schulz

The experience with these breakouts exceeded expectations by leading to new results and also extensive contributions to the discussions by all participants. As such, the structure of this workshop proved very useful and might be a good idea for other topics as well.

5 Open Problems – Future Research Direction

The workshop identified a series of major problems, each covering a number of technical issues as future research topics. The major problems are as follows:
- *Different groups have different optimization functions*: While the overall goal, efficient usage of power at the highest level of performance, is the central goal, the actual goals for each group depend on their respective layer in the computing stack. We identified different goals for the layers infrastructure, system software, algorithms, and applications. In addition, the goals may also differ between the computing centers corresponding to their respective targets and operational goals.

- *Information exchange between layers is insufficient*: In order to achieve optimal performance in a power-bound environment, improved information exchange between the above mentioned layers is needed. This requires corresponding tools and interfaces, such that the information available on one layer can be transferred to the layers above or below.
- *User happiness must be weighted against total costs against performance*: While solutions for one characteristic are possible, they have to be weighted against each other. Any solution needs to ensure that application users are happy enough with the operation of the machine, while providers are able to shoulder the costs, while the performance of the application offers a suitable time-to-solution.

Participants

- Michael Bader
TU München, DE
- Natalie Bates
Lawrence Livermore National Laboratory, US
- Pete Beckman
Argonne National Laboratory, US
- Jonathan Eastep
Intel – Hillsboro, US
- Neha Gholkar
North Carolina State University – Raleigh, US
- Joseph Greathouse
AMD – Austin, US
- Thomas Ilsche
TU Dresden, DE
- Dieter Kranzlmüller
LMU München, DE
- Stephanie Labasan
Univ. of Oregon – Eugene, US
- David K. Lowenthal
Univ. of Arizona – Tucson, US
- Matthias Maiterth
LMU München, DE
- Andres Marquez
Pacific Northwest National Lab. – Richland, US
- Yousri Mhedheb
KIT – Karlsruher Institut für Technologie, DE
- Shirley V. Moore
University of Texas – El Paso, US
- Frank Mueller
North Carolina State University – Raleigh, US
- Andreas Raabe
DFG – Bonn, DE
- David Richards
LLNL – Livermore, US
- Suzanne Rivoire
Sonoma State University – Rohnert Park, US
- Barry L. Rountree
LLNL – Livermore, US
- Martin Schulz
LLNL – Livermore, US
- Kathleen Sumiko Shoga
LLNL – Livermore, US
- Torsten Wilde
LRZ – München, DE