The HPC PowerStack: A Community-wide Collaboration Towards an Energy Efficient Software Stack

Siddhartha Jana (EEHPC-WG); Stephanie Brink (LLNL); Christopher Cantalupo (Intel); Jonathan Eastep (Intel); Masaaki Kondo (RIKEN, Univ. of Tokyo); Matthias Maiterth (LMU-Munich); Aniruddha Marathe (LLNL); Tapasya Patki (LLNL); Barry Rountree (LLNL); Ryuichi Sakamoto (Univ. of Tokyo); Martin Schulz (TU-Munich, LRZ); Carsten Trinitis (TU-Munich); Josef Weidendorfer (LRZ)

Abstract

This poster highlights an ongoing community-wide effort among vendors, labs, and academia, to incorporate power-awareness within system-stacks in upcoming Exascale machines. HPC PowerStack is the first-and-only community-driven vendor-neutral effort to identify what power optimization software actors are key within the modern-day stack; discuss their interoperability, and work towards gluing together existing open source projects to engineer cost-effective, but cohesive, portable implementations.

This poster disseminates key insights acquired in the project, provides prototyping status updates, highlights open questions, and solicits participation addressing the imminent exascale power challenge.

Motivation for the poster and its relevance to SC19

Community-interest in tackling Exascale power challenges is growing. While there exist several standalone efforts that attempt to tackle Exascale power challenges, the majority of the implemented techniques have been designed to meet site-specific needs or optimization goals. There is no consensus among the stakeholders in academia, research and industry on which software components of modern HPC stack should be deployed and how they should interoperate. Coordination among these components is critical towards maximizing a target metric (such as FLOPS per watt) while meeting operating constraints (such as energy).

This realization led to the formation of the PowerStack Community in 2016 (https://hpcpowerstack.github.io). The charter of this community includes (A) identifying the key software actors needed in a system power stack; (B) reaching a consensus on their roles and responsibilities; (C) designing communication protocols for bidirectional control and feedback signals among them for enabling scalable coordination at multiple granularities; (D) establishing a unified hierarchical communication model to access power monitor and control knobs in hardware and software; and (E) leveraging existing R&D prototypes and building a community that actively participates in the development and engineering efforts in this domain.

The vision of this poster is to provide an overview of the PowerStack initiative, give a glimpse of some initial prototyping results, list multiple collaborators, point to relevant literature published within the community, and highlight various working groups that the reader can contribute to based on their background and expertise. This will address all SC19 attendees from academia, research, and industry.

Recent Activities prior to this submission

In June 2018 and 2019, a group of 40+ senior researchers, developers, and leaders from vendors, labs, and academia around the globe have convened in Germany for a face-to-face seminar. The community

(representatives of all software stack layers), arrived at a consensus that (1) job/application-awareness is going to be critical for boosting system-wide optimization. This implies the need to drive interoperation between a job-level runtime and the job scheduler; (2) hierarchical control-systems provide a good model for scalable global optimization across the system, so the power-stack should be a hierarchical system with bidirectional control and feedback signals flowing between the actors; (3) rather than providing layered access to privileged hardware knobs, today's systems have an inefficiency in that they break this hierarchy model. And we as a community need to work towards fixing this. These were in accordance with the feedback from the SC18 PowerStack-BoF attendees.

The current list of contributors

- National labs: LLNL, LANL, Sandia, Argonne, Riken, STFC/Hartree, Cineca, LRZ, Grenoble
- System Integrators: Cray, Fujitsu, HPE, ATOS/Bull, IBM
- Chip Vendors: x86 (Intel, AMD), ARM, POWER (IBM)
- Job scheduler / Resource manager vendors: Altair, SchedMD, ALPS (Cray)
- Academia: TU-Munich, TU-Dresden, UniBo, SDU, Univ of Tokyo, LRZ,
- Facility and Operations: EEHPC-WG (Energy Efficient HPC Working Group)

In order to ensure that the work here is directly driven by state-of-the-art solutions adopted by HPC sites around the world, members of the committee (in collaboration with the Energy Efficiency HPC Working Group) surveyed multiple global centers including - CEA (Alternative Energies and Atomic Energy Commission), CINECA, JCAHPC (the University of Tsukuba, University of Tokyo), KAUST (King Abdullah University of Science and Technology), LRZ (Leibniz Supercomputing Centre), RIKEN, STFC (Science and Technology Facilities Council), Tokyo Institute of Technology, Trinity (Los Alamos and Sandia National Laboratories)

Publications and Tech Articles:

- Panel on Software Improvements for Power/Energy Measurement Capabilities - insideHPC, November 2018 (https://insidehpc.com/2019/01/dan-reed-panel-on-energy-efficient-computing-at-sc18/)
- The PowerStack Initiative (A Community-driven Effort) - EEHPC-WG Webinar Series, October 2018 (https://eehpcwg.llnl.gov/assets/091218_webinar_powerstack.pdf)
- OSTI Technical report, "A Strawman for an HPC PowerStack", C. Cantalupo, J. Eastep, S. Jana, M. Kondo, M., Maiterth, A. Marathe, T. Patki, B. Rountree, R. Sakamoto, M. Schulz, C. Trinitis, August 2018, (https://www.osti.gov/servlets/purl/1466153)
- Energy efficiency and the software stack - insideHPC, December 2017 (https://insidehpc.com/2017/12/sc17-energy-efficiency-software-stack-cross-community-efforts/)
- A global survey of HPC center energy and power-aware job scheduling and resource management, November 2017 (https://insidehpc.com/2017/12/first-global-survey-energy-power-aware-job-scheduling-resource-management/)