Energy Efficiency Workshop at PPAM

September 10, 2024, Ostrava

Title

Energy Efficient Operation of HPC Systems (EEOHPC)

Abstract

Energy and power challenges increase as High-Performance Computing (HPC) scales to meet growing demands from industry and research. These challenges include higher CO_2 emissions, increased energy costs, and putting a strain on the existing power infrastructure. In response, HPC centres look for ways to reduce energy consumption and enhance their energy efficiency, for instance by optimising resource utilisation, and manage workloads efficiently.

Efforts to improve energy efficiency have often focused on hardware advancements, including new microarchitectures, intra-core parallelism via vectorization, and use of accelerators for critical workloads. While these innovations have brought benefits such as reduced idle power and more efficient execution of instruction streams, they have also introduced new challenges such as swift power variations and programming complexities associated with exploiting parallelism.

Concurrently, advancements in data centre infrastructure, rack design, and cooling techniques have progressed. Liquid cooling, especially direct hot-water cooling, has gained traction for its cost-saving potential, albeit with challenges like the required changes in chip packaging and, board design to achieve even cooling distributions, and infrastructure upgrades to support rising power densities.

Despite hardware improvements, solely relying on them falls short of fully addressing the energy challenges due to their localized impact and limited adaptability to workloads. Complementary software solutions promise to provide an instantaneous global view of the system status and energy usage, support dynamic adaptation across the stack, enable long-term predictions of resource use, and deliver actionable insights on workload optimisations to users. Additionally, research on power-steering runtimes and monitoring tools has contributed to the development of user-facing analytics tools. The rapid progress of AI techniques opens up additional opportunities here.

This 1st **PPAM workshop on Energy Efficient Operation of HPC systems (EEOHPC)** brings together a number of invited speakers that tackle different dimensions of operating HPC systems in an energy-efficient manner.

Topics of the Workshop

The topics include:

- Monitoring infrastructures to accurately collect operational data with minimal overhead.
- Sharing of operational and monitoring data across system software components as well as between sites.
- Use of data analytics to understand and analyse monitoring data and predict future behaviour of workloads and of the HPC system.
- User-facing tools capturing and analysing the full breadth of operational, workload and monitoring information to provide actionable insights regarding optimisation, associated to the workload/code structure.

- Runtimes able to optimise system operation and workload mapping according to analytics results and resource utilisation predictions during runtime.
- Resource management systems able to act on the predictions by the data analytics framework, by adjusting the execution schedule, workload settings, workload mapping and system configuration.

Speakers

The following invited **speakers are confirmed:**

- **Thomas Ilsche (TU Dresden):** Monitoring and Data collection of Energy use in HPC systems
- **Oriol Vidal (BSC)**: Topic: Energy Efficient operation tuning (EAR)
- **Andrea Bartolini (UniBo)**: Topic: AI-based operational data analytics and prediction
- **Simon Pickartz (ParTec):** Topic: Dynamic scheduling
- **Ondrej Vysocky (IT4I@VSB)**: Topic: Resource management at the node level

Agenda

The proposed agenda is to be adapted to the general schedule to match the start/end and coffee/lunch break times. All **presentation slots are 20min + 10min questions**. The last hour is dedicated to an open discussion between speakers and the audience.

Time	**Topic**	**Abstract**
08:40 – 10:00	Keynotes from the conference	
10:00 – 10:30	*Coffee Break*	
10:30 – 10:40	Welcome (Chairs: Estela Suarez, Hans-Christian Hoppe, Lubomir Riha)	Welcome to the workshop and presentation of the agenda
10:40 – 11:10	Monitoring and Analysis of Energy Consumption in HPC systems (Thomas Ilsche, TU-Dresden)	A robust understanding of energy consumption is essential for efficiently operating High-Performance Computing (HPC) systems as well as data centre capacity planning. This talk discusses various instrumentation points as well as exemplary power measurement solutions and their accuracy and time resolution. Additionally, the talk will introduce a unified infrastructure for collection, storage, analysis and visualization.
11:10 – 11:40	Smart energy efficiency and management with EAR (Oriol Vidal, BSC)	EAR is a European open-source system software for energy efficiency and management in Data Centres. This talk will present the EAR architecture, emphasizing in the core components in charge of the energy efficiency and management in computational nodes. These components are the EAR Job Manager, the EAR Node Manager and the EAR Scheduler plugin. These three components share events, node telemetry, application significant performance and power metrics to implement smart energy optimization policies for HPC and AI workloads and node power cap guided by application activity phases.
11:40 – 12:10	Data-driven and AI-driven models for sustainable computing (Andrea Bartolini, University of Bologna)	The efficiency and sustainability of high-performance computing systems have never been so important for societal development. In this talk, I will cover the recent research results as well as

		lessons learnt in applying AI techniques for modelling and predicting key operational parameters essential for increasing the sustainability of large-scale high-performance computing installations.
12:10 – 12:30	Morning Wrap-up (Chairs)	(also to serve as buffer in case more questions are raised within the talks)
12:30 – 13:20	*Lunch Break*	
13:20 – 15:20	Keynotes from the conference	
15:20 – 15:50	*Coffee Break*	
15:50 – 16:20	Improving HPC system energy efficiency using MERIC runtime system (Ondrej Vysocky, IT4I@VSB)	An HPC system can be optimized for energy efficiency at several levels, while the highest level of dynamicity comes from the power management of computing components controlled at the job level. Complex parallel applications show different hardware requirements during their execution. This dynamic behaviour can be exploited for energy savings by changing the hardware power knobs to fit the configuration to the application's needs. Energy-efficient runtime systems provide administrator- and user-friendly ways to perform such hardware power management without requiring a deep understanding of the topic. EuroHPC Centre of Excellence Performance Optimisation and Productivity (POP) flagship code MERIC is a light-weight tool designed to provide a detailed analysis of application behaviour, identify the optimal hardware settings concerning energy consumption and runtime, and provide dynamic tuning during the application runtime. Thanks to complex execution time coverage by regions of interest, high tuning granularity starting at the level of ten milliseconds, and a large set of controlled power knobs, it pushes the achievable energy savings to the limit.
16:20 – 16:50	Towards Energy-efficient System-level Scheduling for Modular Supercomputers (Simon Pickartz, ParTec AG)	Today's HPC systems offer a variety of mechanisms to measure and control the energy and power consumption of the hardware. However, it is up to the system software to take advantage of these features to optimize system utilization in terms of throughput and energy efficiency. This is even more true for heterogeneous systems comprising modules with different capabilities to be matched with the diverse requirements of today's HPC workloads. Therefore, all levels of the system software stack, from the system level to the workflow and job level to the node level, have to be considered when designing the stack. The system level has a holistic view on the resource availability and requirements of all workloads. With the goal of globally optimizing system utilization, the Resource and Job Management System (RJMS) shall be able to dynamically adapt the schedule as jobs come and go. This talk provides an overview of state-of-the-art RJMSs and an analysis of how to evolve them into a production-quality solution for

		dynamic scheduling and resource management in HPC.
16:50 – 17:10	EuroHPC JU Call for an integrated, energy-optimised system SW stack (Hans-Christian Hoppe, FZJ	Presentation of EuroHPC initiative to develop an integrated monitoring, resource management, and scheduling stack for improving energy efficiency of EuroHPC JU HPC centre operation for their actual workloads.
17:10 – 17:40	Discussion	Panel and Open discussion with the audience
17:40	End of day	

Organizers

Organizer 1: Estela Suarez

Prof. Dr. Estela Suarez is Joint Lead of the department "Novel System Architecture Design" at the Jülich Supercomputing Centre, which she joined in 2010. Since 2022 she is also Associate Professor of High Performance Computing at the University of Bonn, and member of the RIAG (Research and Innovation Advisory Board from EuroHPC JU). Her research focuses on HPC system architecture and codesign. As leader of the DEEP project series she has driven the development of the Modular Supercomputing Architecture, including hardware, software and application implementation and validation. She also leads the codesign efforts within the European Processor Initiative. She holds a PhD in Physics from the University of Geneva (Switzerland) and a Master degree in Astrophysics from the University Complutense of Madrid (Spain).

Organizer 2: Hans-Christian Hoppe

Hans-Christian Hoppe has made significant contributions to HPC since the 1990s, with an impact on the MPI standard, first proof use of virtualization in Grid systems, a leading scalable performance analysis tool, and the co-definition of the Modular Supercomputing Architecture (MSA).

Since 2003, he worked as a Principal Engineer with Intel, responsible for the Intel Cluster tools, early research on Manycore, the Intel Visual Computing Institute, and the Intel Exascale Lab at JSC, and as co-lead for a Darpa project on disruptive Graph-analytics architectures.

In 2022, he joined Jülich Supercomputing Centre and ParTec AG, leading the DEEP-SEA project and contributing to several other EU-funded projects, including the RAISE CoE and EUPEX. His research interests include heterogeneous system architectures, dynamic use of resources in HPC, scalable performance analysis methods and tools, and integration of disruptive technologies (such as Quantum Computing) into HPC infrastructures.

Organizer 3: Lubomir Riha

Lubomir Riha, Ph.D. is the Head of the Infrastructure Research Lab at IT4Innovations National Supercomputing Center. Previously he was a research scientist in the High-Performance Computing Lab at George Washington University, ECE Department. He received his Ph.D. degree in Electrical Engineering from the Czech Technical University in Prague, Czech Republic, and a Ph.D. degree in Computer Science from Bowie State University, USA.

Currently, he is a local principal investigator of MAX3, SPACE and POP3 EuroHPC Centers of Excellence, as well as EPICURE, FALCON and EUPEX (designs a prototype of the European Exascale machine) EU projects. Previously he was a local PI of the POP2 CoE and SCALABLE and READEX projects.

His research interests are optimization of HPC applications, energy-efficient computing, acceleration of scientific and engineering applications using GPU and many-core accelerators, and parallel and distributed rendering.