HPC Power Measurements
Procurement Guidelines

Building Energy Efficient High Performance Computing:
4th Annual EE HPC WG Workshop

Daniel Hackenberg (daniel.hackenberg@tu-dresden.de)
Thomas Ilsche (thomas.ilsche@tu-dresden.de)
Robert Schöne (robert.schoene@tu-dresden.de)
Phase 1 Requirements and Accuracy Tests

- Requirements
  - Full system energy accounting
  - 2% accuracy
  - Min. 1 sample/s
  - Granularity: node level
  - Correct energy calculation based on power samples

- Reference measurements to evaluate accuracy of energy accounting
  - Use professional power meters (ZES LMG 450)
  - Use synthetic idle/busy-loop-benchmark to trigger worst case scenarios (find frequency with highest aliasing effects)
Phase 1 Requirements and Accuracy Tests

Requirements

- Full system energy accounting
- 2% accuracy
- Min. 1 sample/s
- Granularity: node level

Reference measurements to evaluate accuracy of energy accounting

- Use professional power meters (ZES LMG 450)
- Use synthetic idle/busy-loop-benchmark to trigger worst case scenarios (find frequency with highest aliasing effects)
Reference Measurement: 8 DC 54V Channels per Chassis
Energy Correctness Test with Synthetic Benchmark

Accuracy acceptance test:

- Submitting 5 identical jobs with 'worst case properties'
- Accounting results of -49% to +36% compared to reference measurement

- Multiple errors in energy calculation code of Slurm (now mostly fixed)
- Aliasing effects already in data coming from BMC
- High idle power variations – node measurement calibration highly inaccurate
- IPMI limitation of 8 bit per value: power measurement granularity of 2 to 7 W
HPC Power Measurements
Usage Models

Building Energy Efficient High Performance Computing:
4th Annual EE HPC WG Workshop

Daniel Hackenberg (daniel.hackenberg@tu-dresden.de)
Thomas Ilsche (thomas.ilsche@tu-dresden.de)
Robert Schöne (robert.schoene@tu-dresden.de)
**Power Measurement Data Management: Dataheap**

### Live viewers, e.g. Website

- [Chart showing live viewers data]

### Post-Mortem Analysis, e.g.
- Get energy of job
- Get energy for node
- Analyze long term energy savings since optimization
- Websites

---

**Flowchart:**

1. **Compute Node** → **Admin Node**
2. **Admin Node** → **Dataheap Server**
3. **Dataheap Server** → **Database**

- **Bulk IPMI** → **Admin Node**
- **Push Data** → **Dataheap Server**
Power Measurement Projects at TU Dresden

HAEC

1 Sa/s  >10k Sa/s

HDEEM Phase 2 High Res

COOL SiLiCON

1 node  100 nodes  1k nodes  full HPC system

HDEEM Energy Accounting

Daniel Hackenberg
The Classics: History View

- Power at $\geq$ 1 sample/s, temperature at $> 1$ sample/min
- All components: nodes, switches, storage etc.
- Full history, not data reduction/dumping
The Advanced: Performance + Efficiency Analysis at Scale

COSMO on 4 nodes of HRSK-II phase 1 (dual socket Sandy Bridge-EP)
Combining Performance Analysis and Energy Optimization

- Use performance counters to identify memory bound regions
- Next step: Reduce concurrency in these regions (DCT)

NAS Parallel Benchmark sp.C on Intel Xeon E5-2670
Combining Performance Analysis and Energy Optimization

- Concurrency reduced
- KMP_BLOCKTIME (default: 20 ms) prevents idle state
- Min. power 230W, avg. power 270W

NAS Parallel Benchmark sp.C on Intel Xeon E5-2670

Daniel Hackenberg
Reduced KMP_BLOCKTIME set to 0
Average power consumption 225 W
Minimal power consumption 170 W
Power Measurement Projects at TU Dresden

1 Sa/s  1k Sa/s  >10k Sa/s

HAEC

Phase 2 High Res

COOL SiLiCON

HDEEM Energy Accounting

1 node  100 nodes  1k nodes  full HPC system

Daniel Hackenberg
Combining Performance and Energy Efficiency Analysis

- SPEC OMP2001 (applu)
- How much power does my system consume at any given time during application execution?

![Graph showing power consumption over time with various measurements: maximum, average, and minimum.]

- Four different power-measurements
- application activity
- time →
Temporal granularity does matter when analyzing short code paths

- LMG450 (AC and DC): 20 Sa/s
- RAPL: 10 Sa/s (maximum 1 kSa/s*)
- Hall sensor / NI (DC): 200 kSa/s (with 4.5 kHz low-pass filter)

---

![Graph showing temporal analysis](image-url)
Temporal granularity does matter when analyzing short code paths

- LMG450 (AC and DC): 20 Sa/s
- RAPL: 10 Sa/s (maximum 1 kSa/s*)
- Hall sensor / NI: 200 kSa/s (with 4.5 kHz Filter)

tiny OpenMP regions affect power consumption
Thank you!
FIRESTARTER: A Processor Stress Test Utility

Intel Xeon X5670, Westmere-EP (2P), SSE routine

Intel Xeon E5-2670, Sandy Bridge-EP (2P), AVX routine

Intel Core i7 2600, Sandy Bridge-HE (1P), AVX routine

http://tu-dresden.de/zih/firestarter/
References

- http://tu-dresden.de/zih/forschung/projekte/coolscilicon
- http://tu-dresden.de/zih/forschung/projekte/hdeem
- http://tu-dresden.de/zih/forschung/projekte/haec


Michael Kluge, Daniel Hackenberg, Wolfgang E. Nagel: Collecting Distributed Performance Data with Dataheap: Generating and Exploiting a Holistic System View, Procedia Computer Science, Volume 9, 2012

Hackenberg, Daniel; Oldenburg, Roland; Molka, Daniel; Schone, Robert: Introducing FIRESTARTER: A processor stress test utility, International Green Computing Conference (IGCC), 2013
Acknowledgement: Cool Silicon Spitzencluster

Sustainability through energy efficiency. The region “Silicon Saxony” acted on this maxim and founded the Leading Edge Cluster “Cool Silicon - Energy Efficiency Innovations from Silicon Saxony”.

The Leading Edge Cluster Cool Silicon
- is aiming at providing a technological basis for increasing the energy efficiency in the information and communications technology sector.
- is assured Germany and Europe a leading role in the key technology of “energy efficiency in the information technology sector”.

Cool Silicon - Areas

Area 1
Micro- and Nanotechnologies
The core objective of the Area 1 project partners is the development of basis technologies, analysis and production methods for the production of energy efficient electronics and their application in order to decrease the energy consumption of computer systems.

Area 2
Communication Systems
In Area 2, the research and development projects are focussing on the improvement of energy efficiency in communications infrastructures and mobile devices.

Area 3
Network Sensors
The project CoolSensorNet is the Lead Project of Area 3. It conducts research on the whole electronic chain’s specific requirements, including sensors, analog electronics, A-D converters, processor systems and the telemetry unit.

Funded via BMBF 13N10186