Power Measurement Techniques for Energy-Efficient Computing: Reconciling Scalability, Resolution, and Accuracy

Thomas Ilsche, Robert Schöne, Joseph Schuchart, Daniel Hackenberg, Marc Simon, Yiannis Georgiou, Wolfgang E. Nagel:
Center for Information Services and High Performance Computing (ZIH), TU Dresden

EnaHPC 2017 – June 22, 2017 – Frankfurt

Why Measure Energy / Power?

- Energy monitoring, accounting
- Power capping
- Energy efficiency analysis/optimizations
- Imagine performance optimization with a clock that only updates once a second and has a 10% error

Key Criteria for Energy Measurements

1. Temporal granularity
   → Analyze short program phases
2. Spatial granularity
   → Distinguish individual components
3. Well-defined accuracy
   → Energy savings ≫ uncertainty
4. Scalability
   → Usage in HPC systems
5. Cost
   → Applicable to large production system or small experimental system

Related work

1. Temporal granularity
   → Common solutions between 1 s and 1 ms
2. Spatial granularity
   → DC power meters, e.g. PowerPack, PowerMon2, PowerInsight
3. Well-defined accuracy
   → Professional AC power meter
4. Scalability
   → Vendor power meter, integrated PDU measurement
5. Cost
   → Embedded CPU measurements, RAPL

Advancing Energy Measurement

HAEC

- Custom-built system
- Very high temporal resolution
- Good spatial resolution
- Well understood accuracy

HDEEM

- Vendor collaboration
- Highly scalable
- Verified accuracy
- Good temporal and spatial resolution
- Deployed in production HPC system with 1456 nodes

Systems Under Test

HAEC

- Single node workstation
- 2 × Intel Xeon E5 2690 v3
- 256 GiB memory
- 800 GB SSD
- Ubuntu 16.04 Server

HDEEM

- 1456 Bullx DLC nodes
- 2 × Intel Xeon E5-2680 v3
- 64 – 256 GiB memory
- 128 GB SSD
- Bull SCS4

Instrumentation points

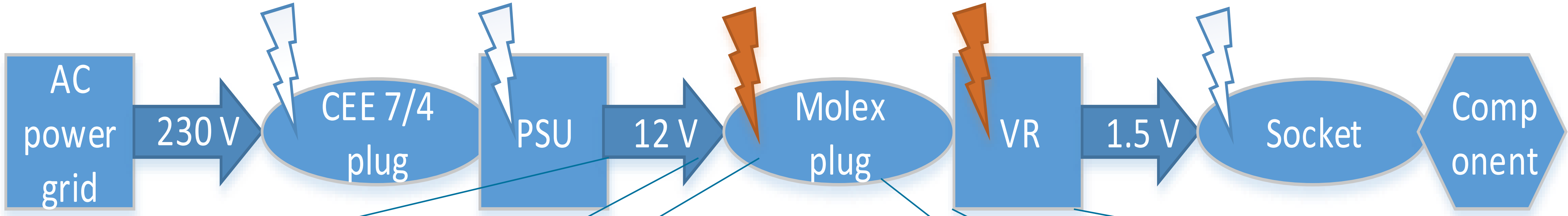
HAEC

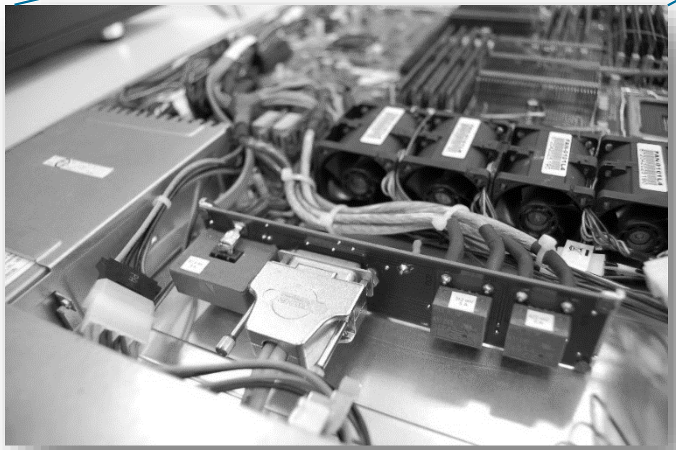
- 2 × sockets (CPU + Memory)
- Mainboard 12V, 5V, 3.3V ATX
- GPU via PCIe and 8+6 Molex
- SSD 12V, 5V
- Sum of all fans
- All DC power consumers

HDEEM

- Total DC node power
- 2 × CPUs
- 4 × DRAM DIMM groups

HAEC Instrumentation Points





Hall Effect Sensors

Shunts

Voltage Regulators

HAEC Measurement Chain

Sensor
- Shunts at all DC consumers
- Also capture all voltages

Analog processing
- Amplification to common voltage range
- Filtering

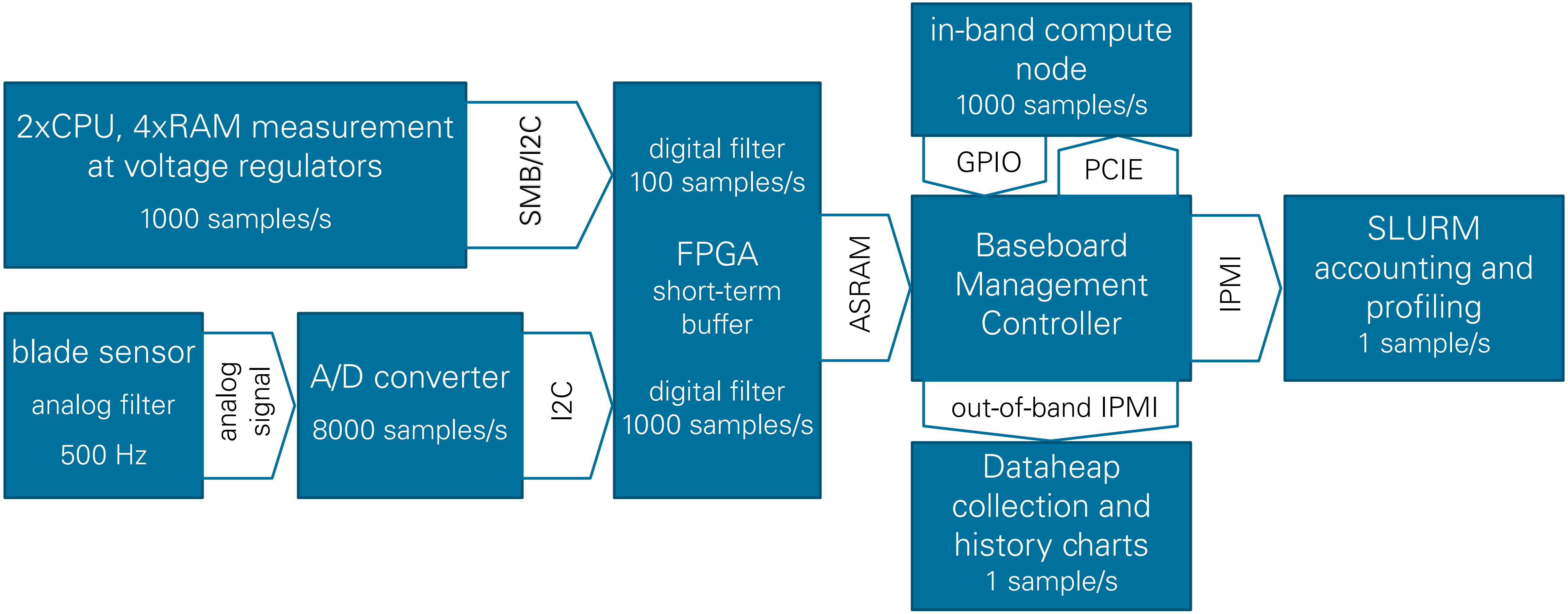
Data Acquisition
- Two NI DAQ cards
- Up to 500 kSa/s (for four sensors)
- All sensors 7 kSa/s

Digital processing
- Out-of-band measurement
- Analysis & integration

HDEEM Hardware

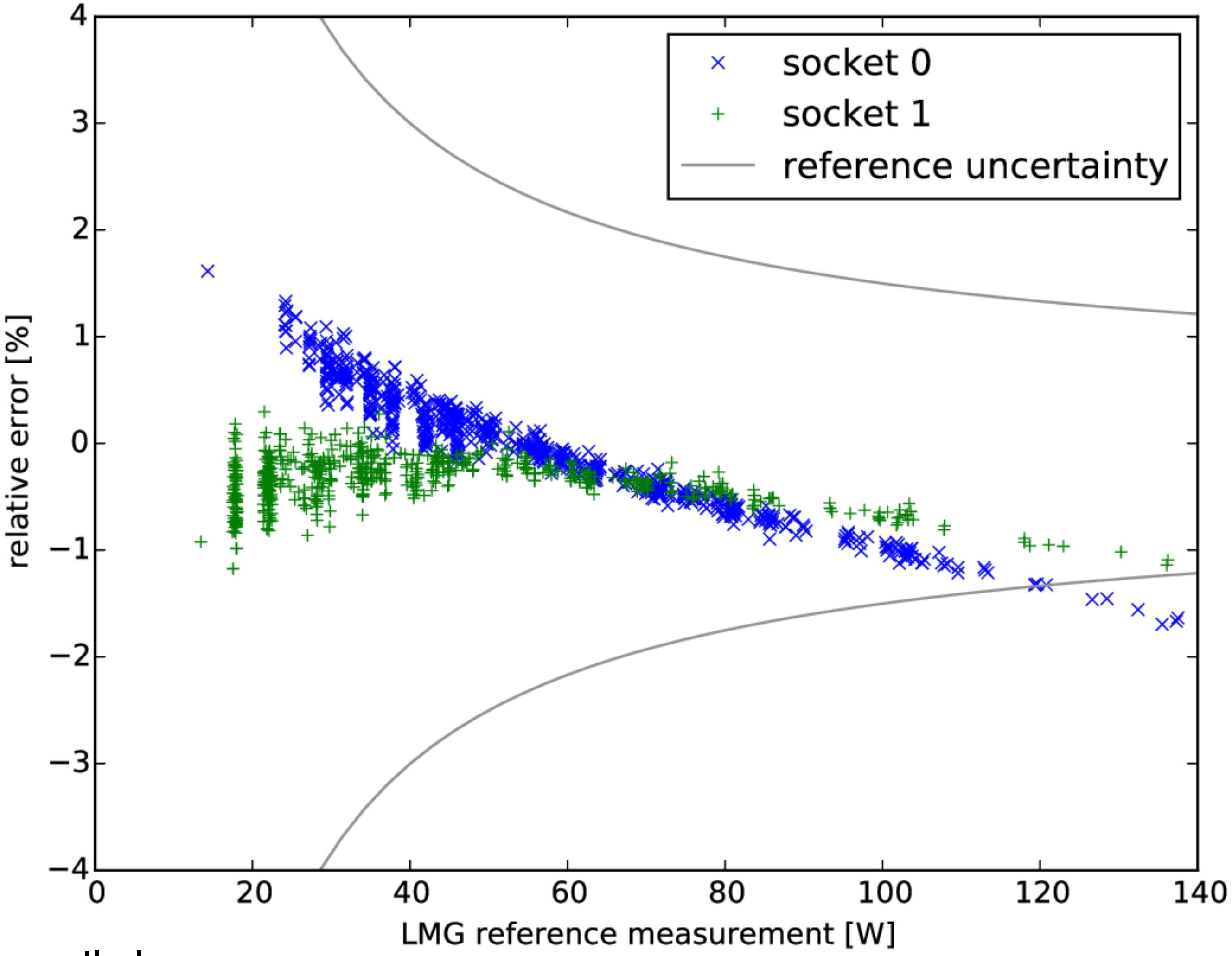
Thomas Ilsche

HDEEM Power Acquisition Scheme



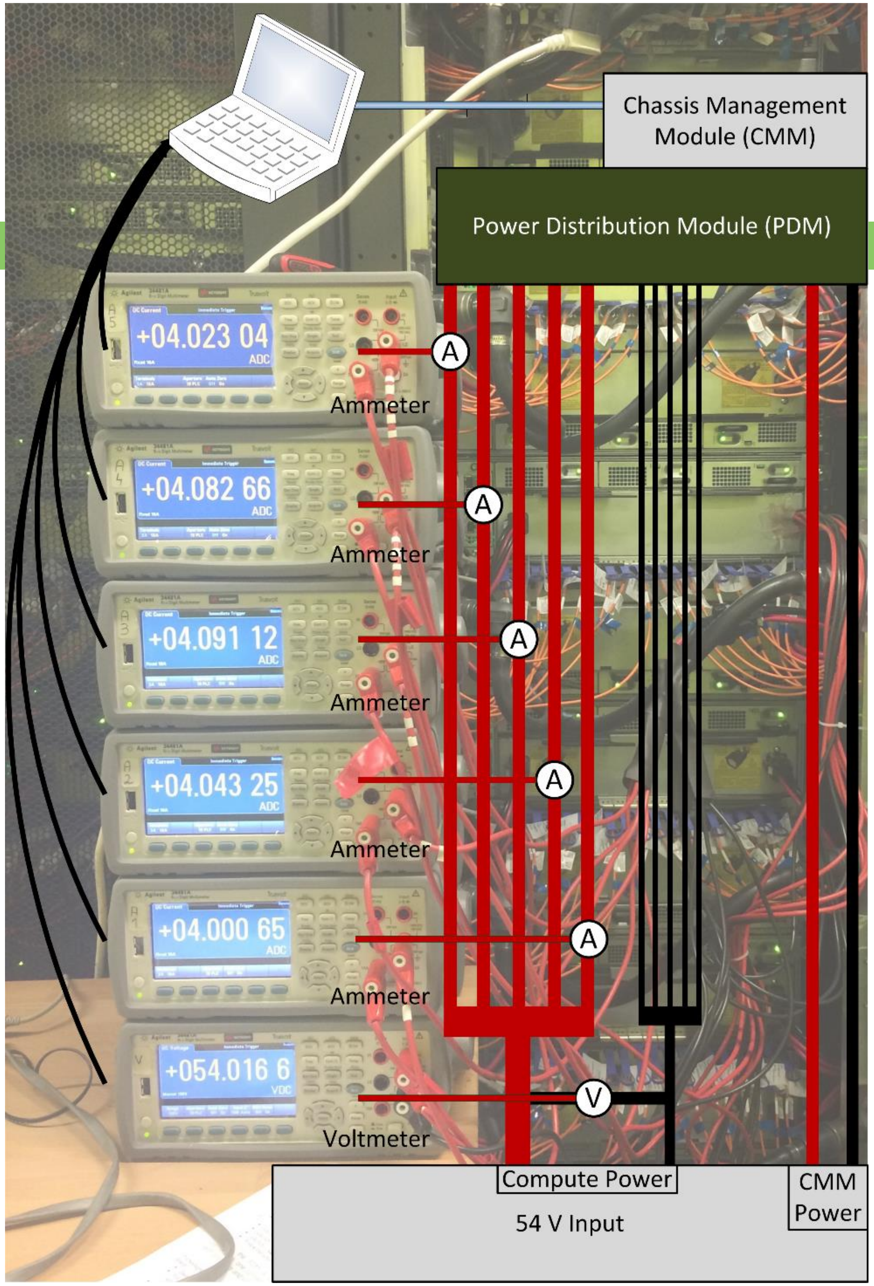
Calibration & Verification HAEC

- Measuring shunts in the setup using a variable load resistor
- Calibration factors in amplifiers
- Error < 2% compared to calibrated reference measurement
- Mostly within the uncertainty of the reference



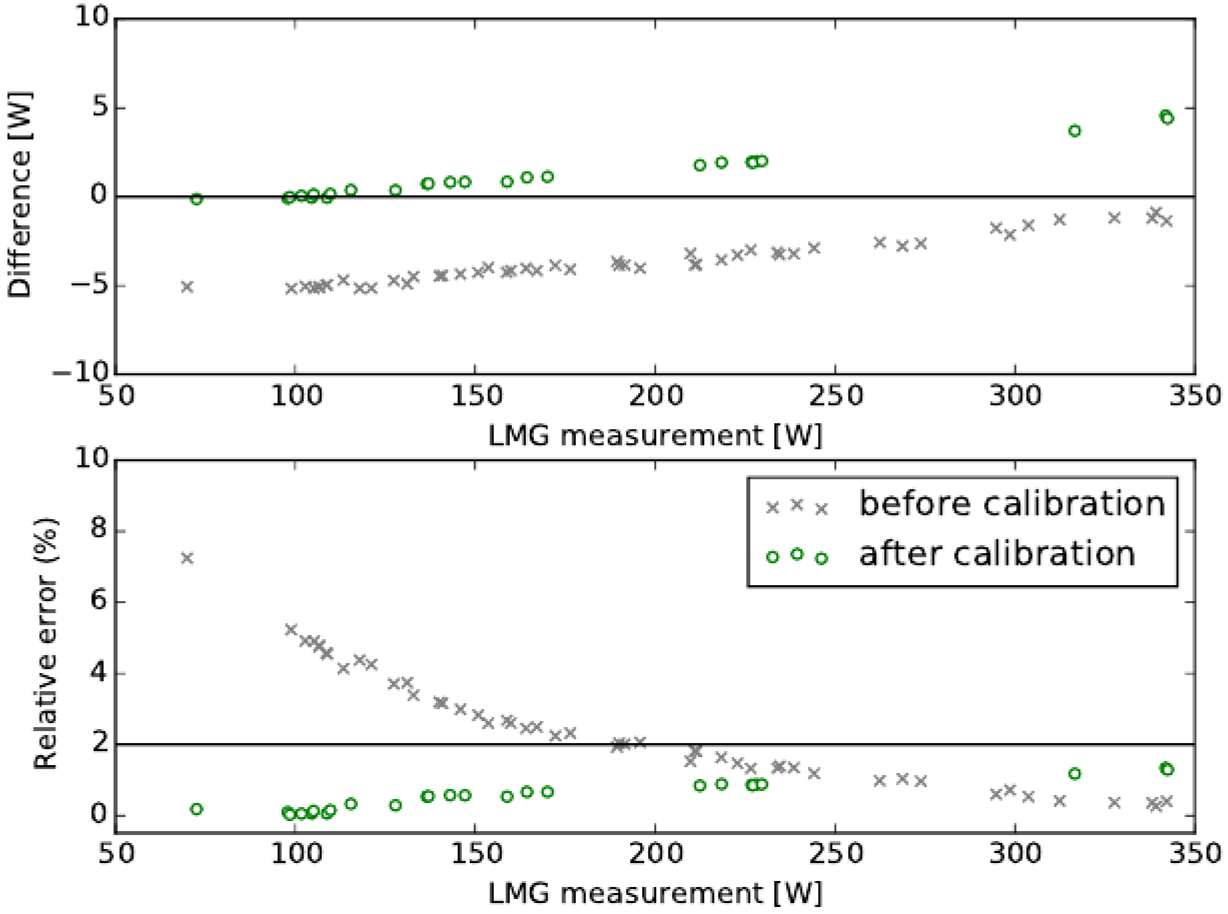
Calibration HDEEM

- Calibration of the deployed HPC system
- Determine calibration factor for each node
- Program correction factors for raw sensor values



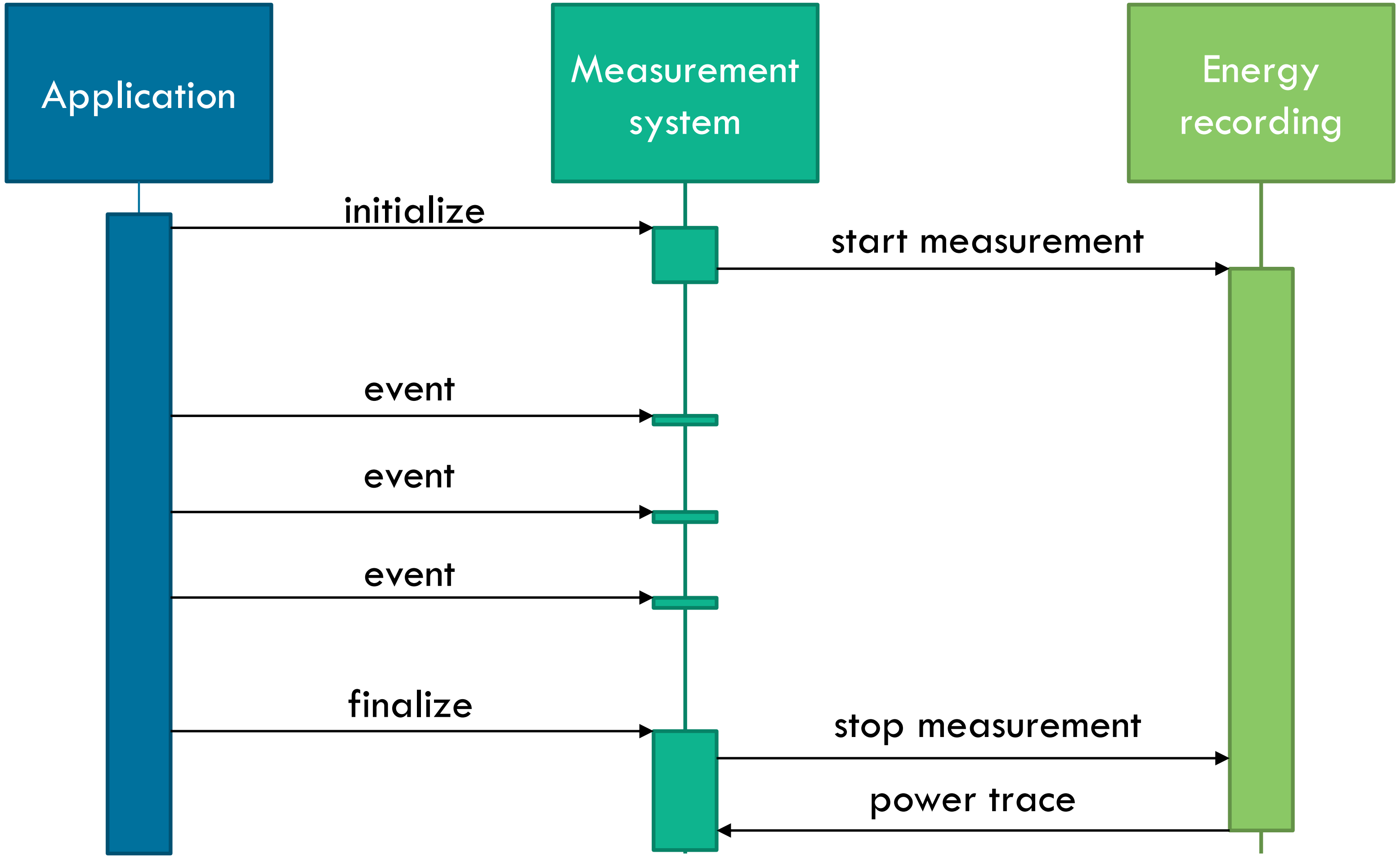
Verification HDEEM

- Independent measurement with calibrated power meter
- Challenging worst-case aliasing workloads
- Maximum error 3 %

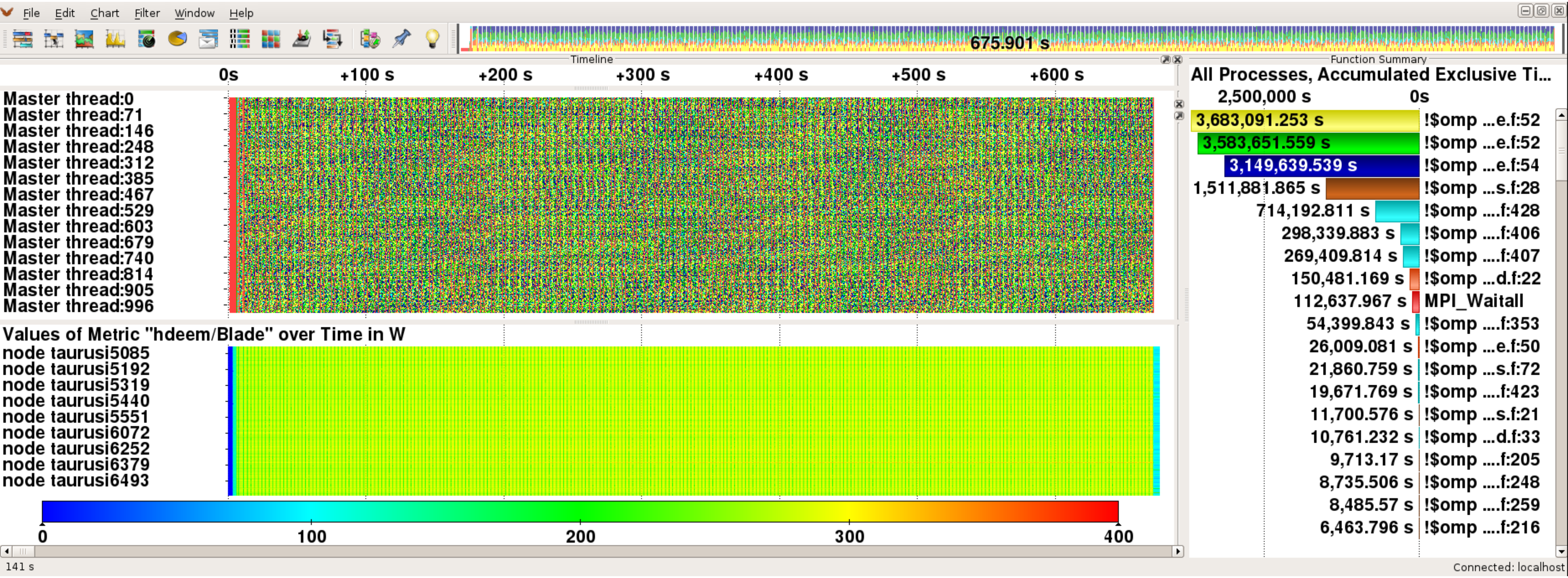


Application Energy Measurement

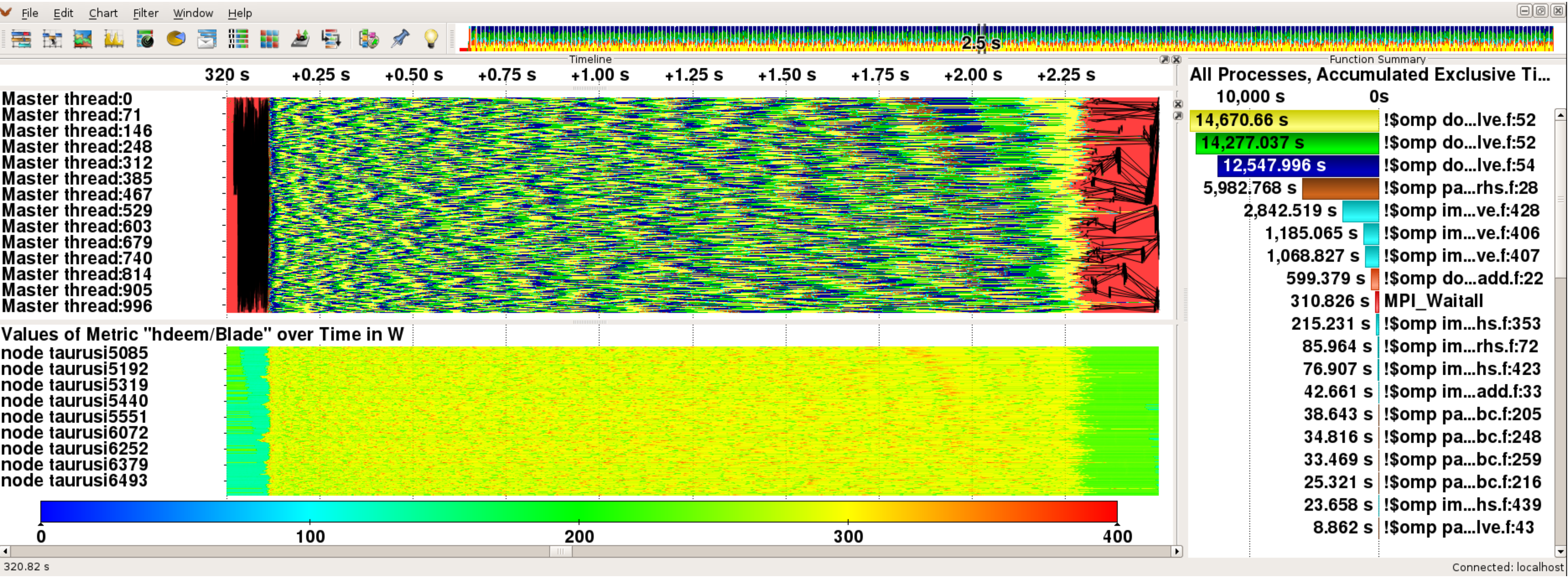
- Plugin metric for Score-P
- All measurement is asynchronous
  - No perturbation
- NAS parallel benchmarks
  - OpenMP & MPI
  - BT solver



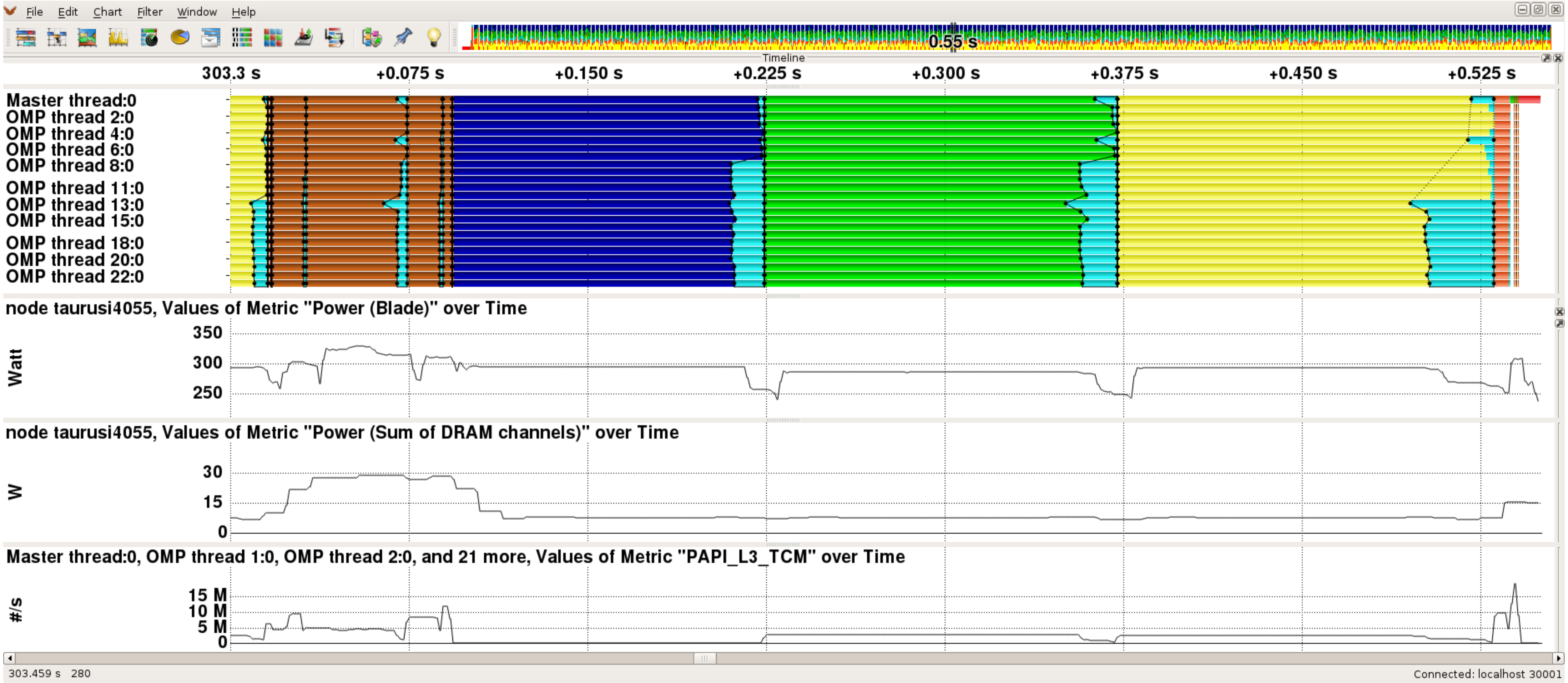
HDEEM Example (NPB on 1024 nodes)



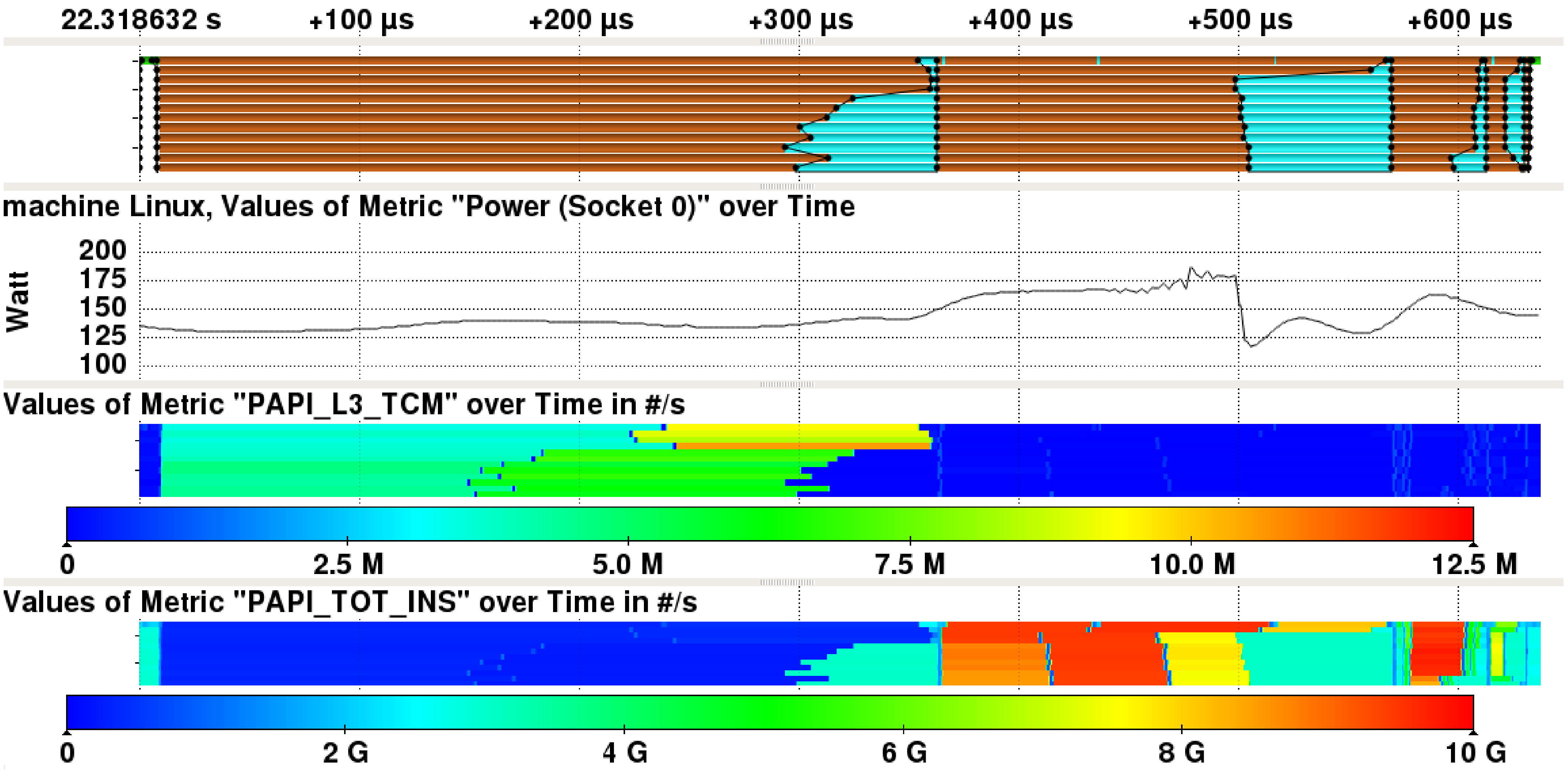
HDEEM Example (NPB on 1024 nodes)



HDEEM Example (NPB on 1024 nodes)



HAEC Example



Summary

- Application optimization needs good metrics
  - Energy efficiency optimizations brings together hardware and software
- Demonstrate two energy measurement infrastructures
  - Very high resolution at small scale
    - 500 kSa/s – observe applications regions of < 100 µs
    - Separate measurements per DC-component
  - High resolution measurement at HPC production scale (1456 nodes)
    - 1 kSa/s per node power
    - 100 Sa/s socket / DRAM measurement
  - Both measurements verified with high accuracy errors < 2 % / < 3 %

Questions