Data Management and Analysis for Energy Efficient HPC Centers

Presented by Ghaleb Abdulla, Anna Maria Bailey and John Weaver;
Lawrence Livermore National Laboratory

Data management and analysis for energy efficient HPC centers

Ghaleb Abdulla, Anna Maria Bailey and John Weaver

Lawrence Livermore National Laboratory

LLNL-PRES-XXXXXX

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Acknowledgement

- Philip Top (LLNL)
- Chuck Wells (OSIsoft)
- EEHPC D/R group
- EEHPC PUE/TUE group

Data collection helps assess the health and efficiency of our facilities

- Perform event analysis
  - Loss of power, voltage sag (dip), Voltage swell, etc.
- Understand where and how power is used
- Perform energy efficiency studies
- Plan for Exascale computing
  - Understand current usage patterns
- Relate component, system, and facility level data

Implement Centralized System

Data Sources
Rack, Equipment, Metering, Building Management, Utility

Interfaces
tens of thousands of Real Time Data Streams

Manage
Gather and Evaluate Large Amounts of Data

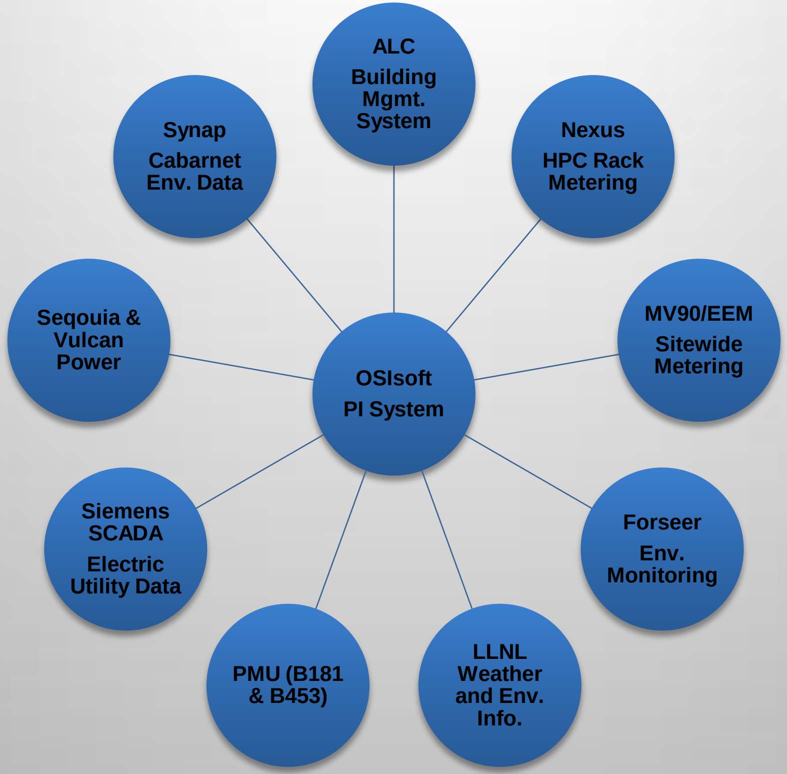
Analyze
Correlate Real Time Data

Notify
Centralized Event Notification

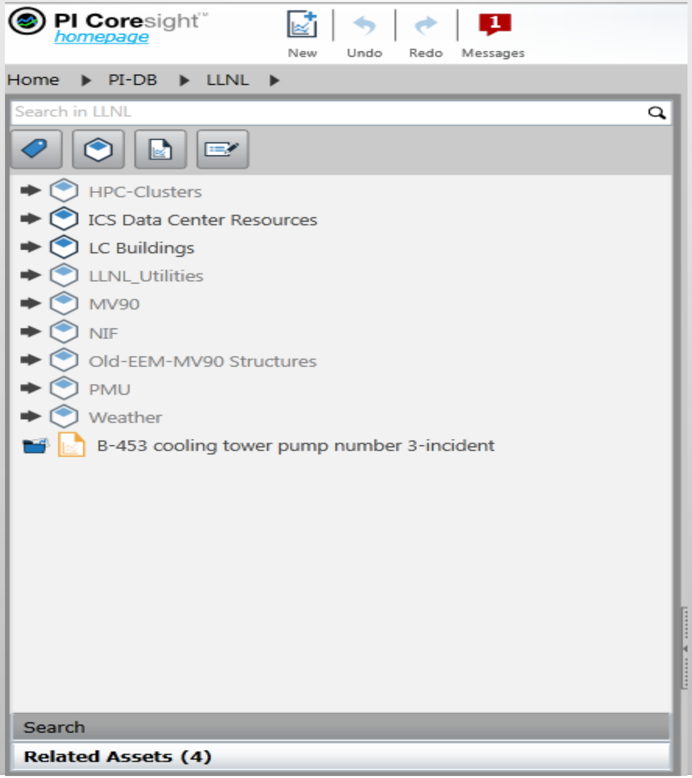
Visualize
View Data and Reports

Created an operational, event, and real-time data management infrastructure of all external and internal data sources

Current data sources spread across LLNL

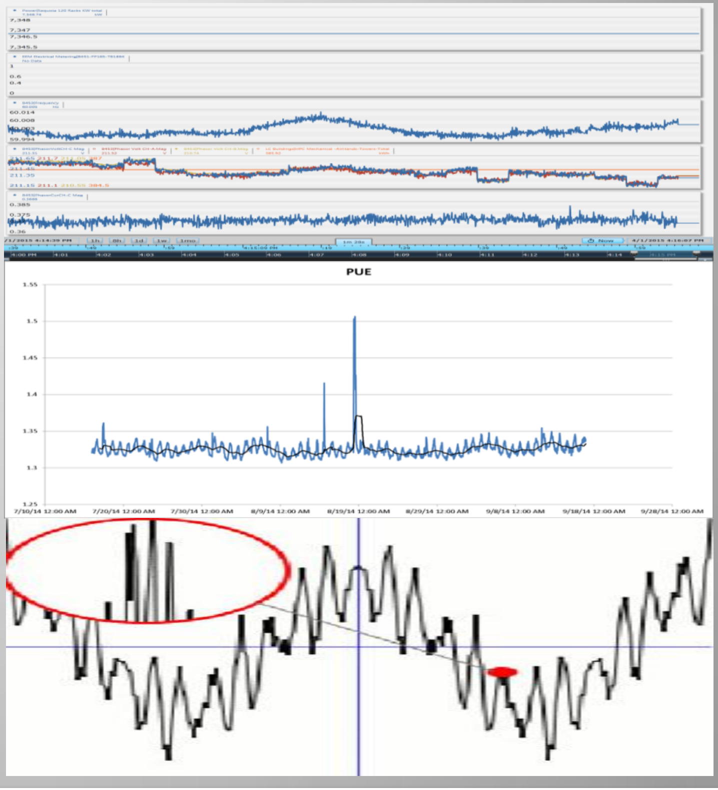


A hierarchical data structure for efficient data discovery



Data analysis

- Exploratory data analysis
  - Use coresight for quick data exploration tasks
- Use PI DataLink for reporting and some data analysis tasks
- Use R for more advanced or repetitive tasks
  - Example: Motif matching



Agenda for the rest of the talk

1. Demand response and Dynamic power management (DPM)
2. Energy efficiency metrics
3. Power quality and outage
4. Power usage characteristics of Sequoia
5. PI Data Server compression study

What is Dynamic Power management?

- Adjusting power parameters on-the-fly while ensuring deadlines of running software are met
- Several strategies, but we are looking into one fine-grained power management strategy:
  - Power capping, running the CPU with power bound

Cabernet (CAB) Overview

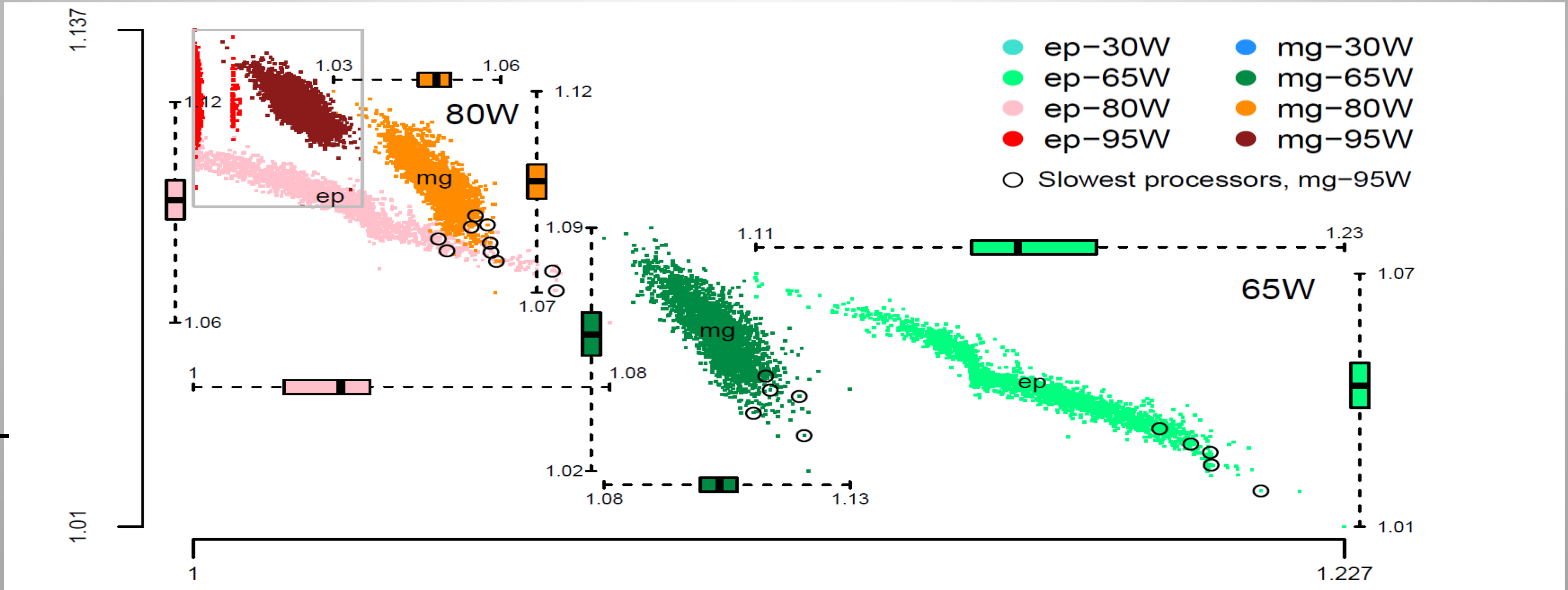
- Appro System
  - Intel Xeon ES-2670
- OS – TOSS
- Interconnect – IB QDR
- 426 TeraFLOP/s peak
- Memory 41,472 GB
- 1296 nodes, 16 cores/node
- Power – 564kW in 675 ft^2
- #94 on November, 2013 Top 500

Power bound experiment

- Embarrassingly parallel application and memory bound application, single socket runs
- Run the application with the same power bound across cluster processors
- Characterize processor variations across several power bounds
- Data shows that dynamic power management will be challenging

Processor performance with 80W and 65W PB

Effective clock frequency multiplier



Normalized slow-down

Energy efficiency metrics

- PUE/TUE
  - We found a meter not reporting correctly (data redundancy)

Energy efficiency metrics

$$PUE = \frac{mechanical + computing + other}{computing}$$

$$ITUE = \frac{total\ energy\ (that\ goes\ into\ the\ machine)}{energy\ into\ the\ computing\ nodes}$$

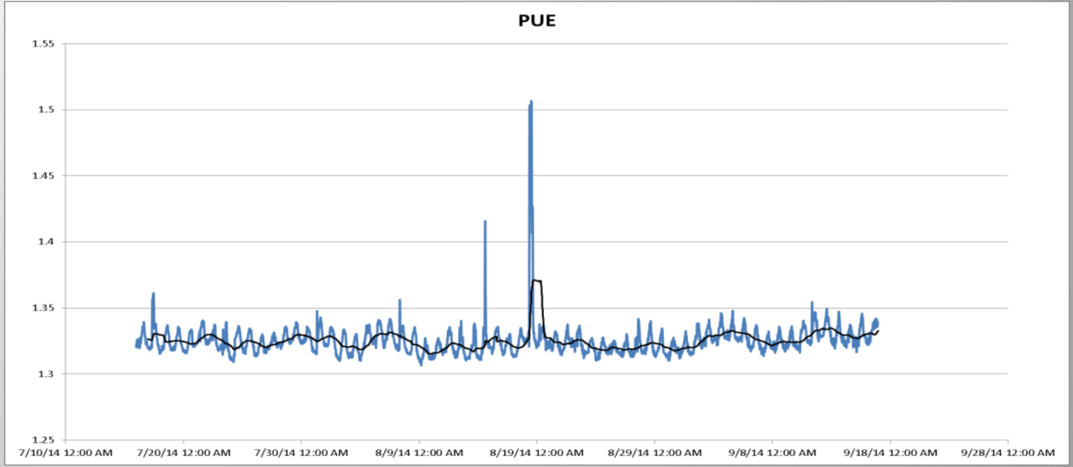
$$TUE = ITUE * PUE$$

Sequoia Parameters

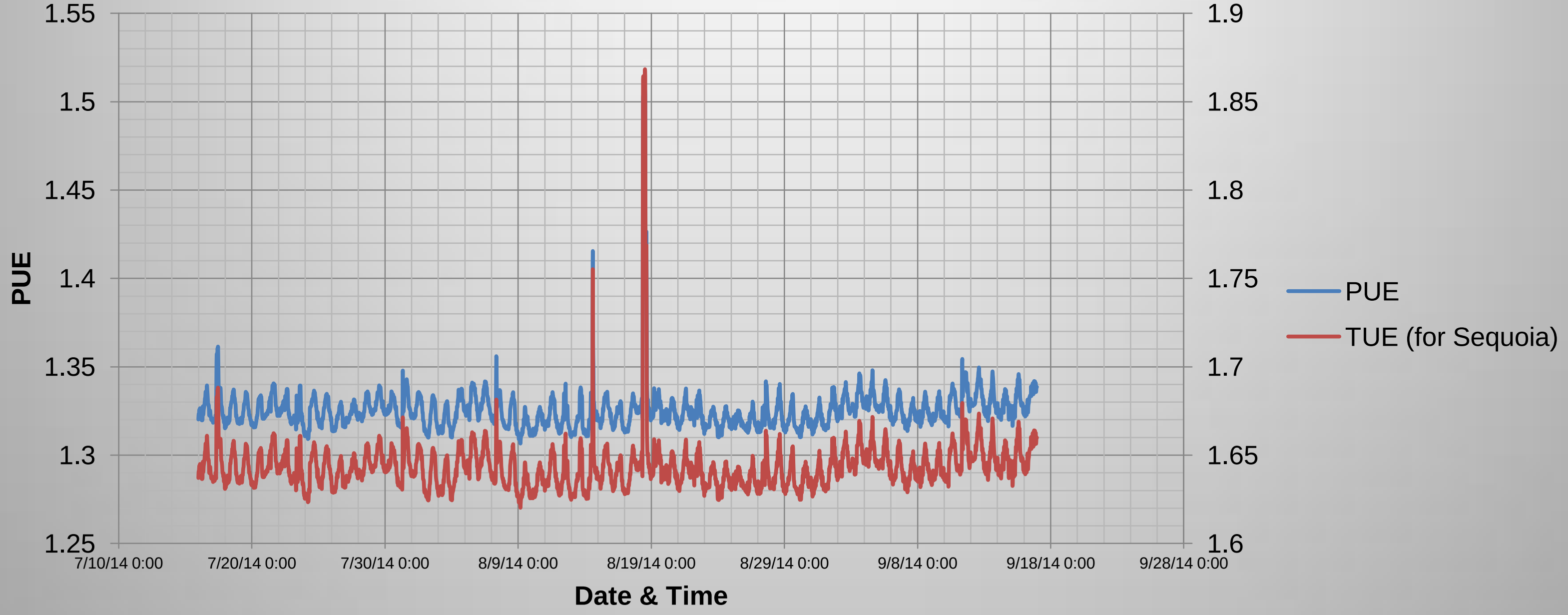
- IBM Blue Gene*/Q architecture
- 98,304 nodes
- 1,572,864 cores
- 20 PF, 3rd on Top 500 – June 2013
- 96 racks
- 91% liquid cooled
- 30 gpm/rack at 62 F
- 9% air cooled
- 1700 cfm/rack at 70 F
- 4800 square feet
- *Copyright 2013 by International Business Machine Corporation

PUE Dashboard

- PUE calculated using the metered data (not sequoia rack power)
  - PUE is now a tag in the DB
  - We found a meter not reporting correctly, data verification using different sensors and interfaces
- High spikes are when Sequoia is down for maintenance
  - Regular maintenance schedule with one major outage
- Daily and weekly cycles

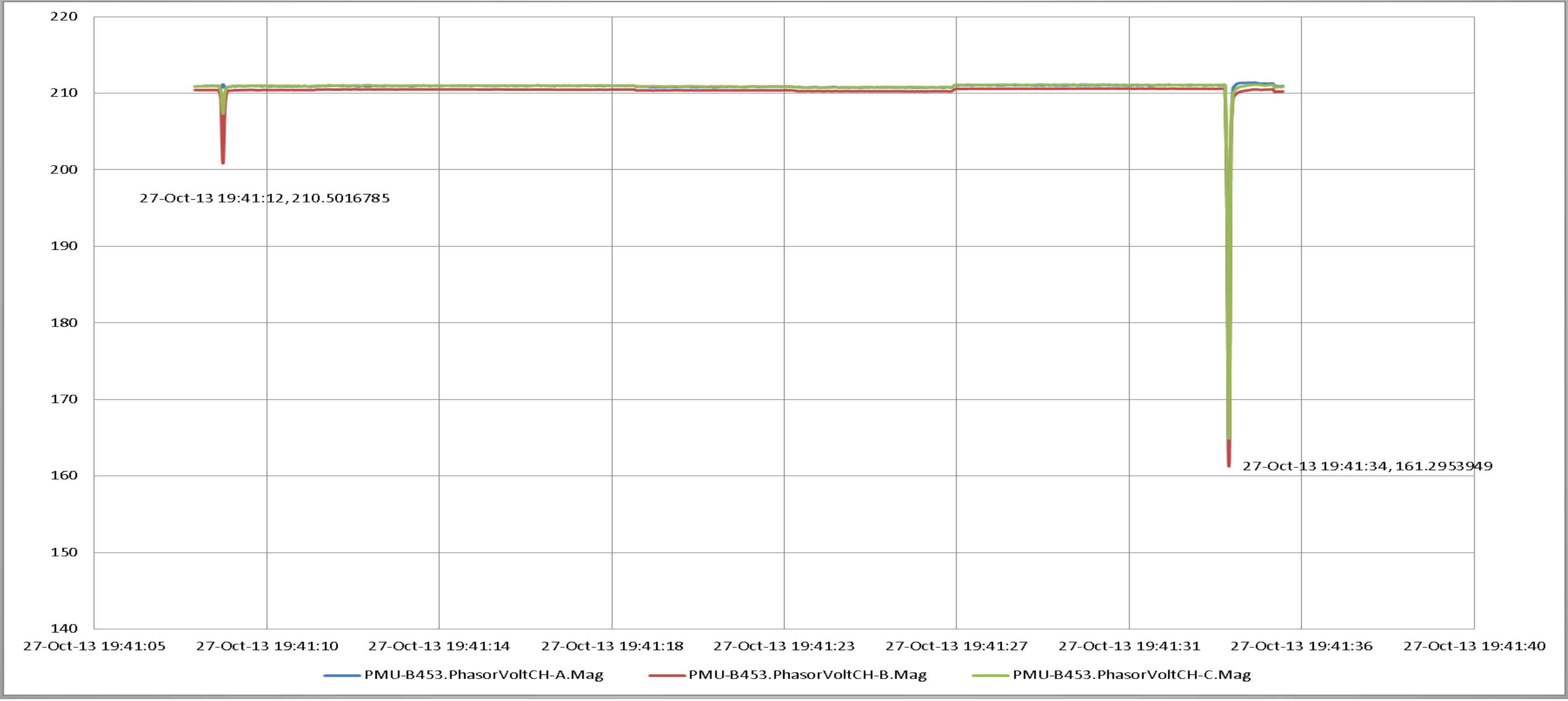


TUE graph for Sequoia

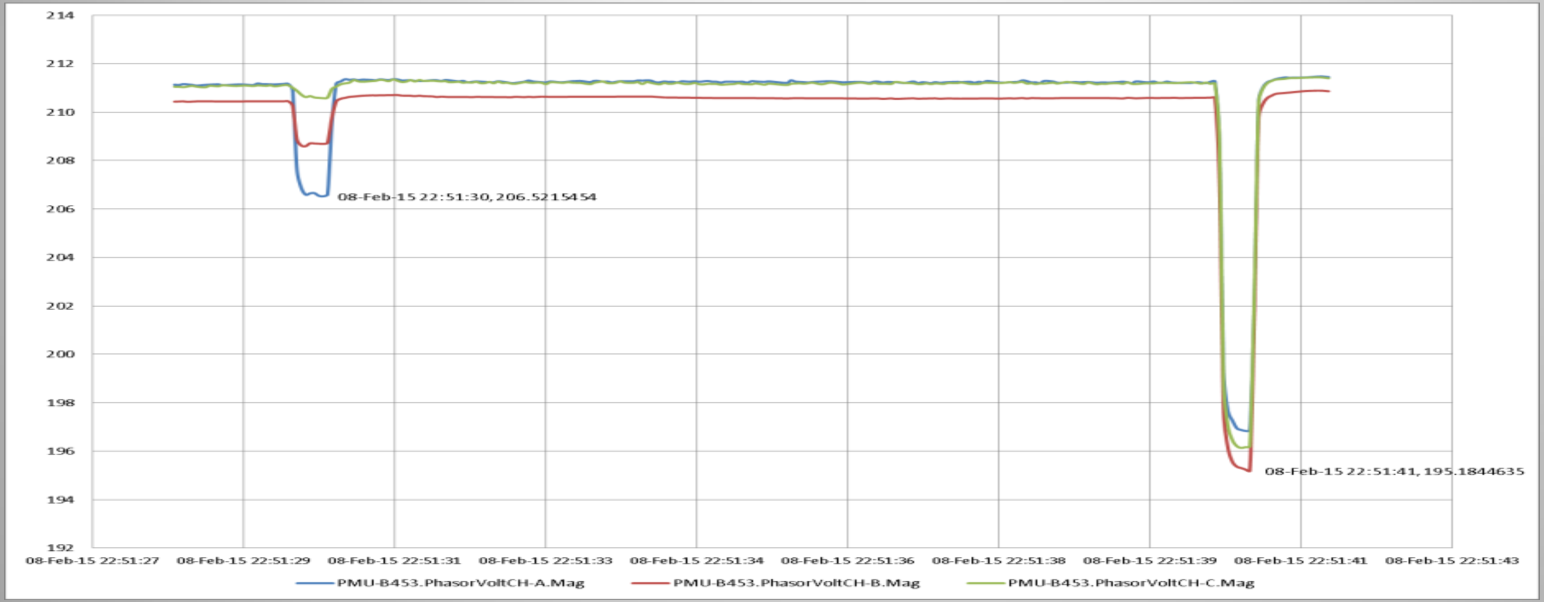


Power quality and outage

Event time and date7:41 pm on 10/27/2013



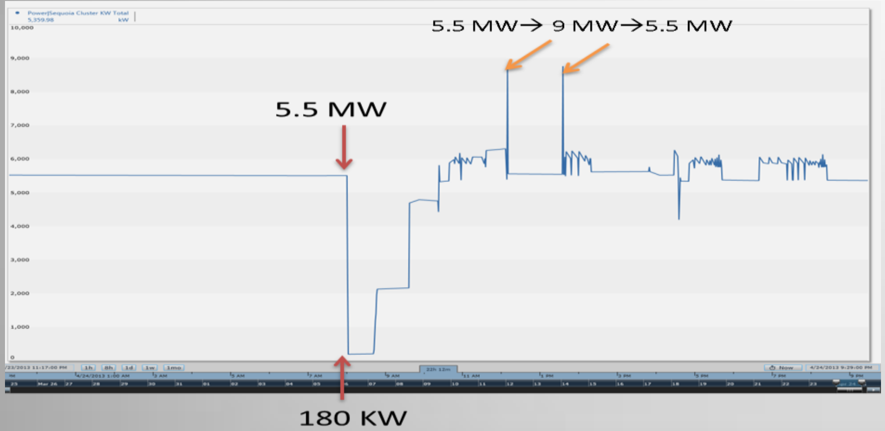
Event time and date: 10:51 pm on 02/08/2015



- Wind with gusts over 67 always has a SW and SSW direction
- Winds gusting below 67 come from other directions

Power usage characteristics of Sequoia

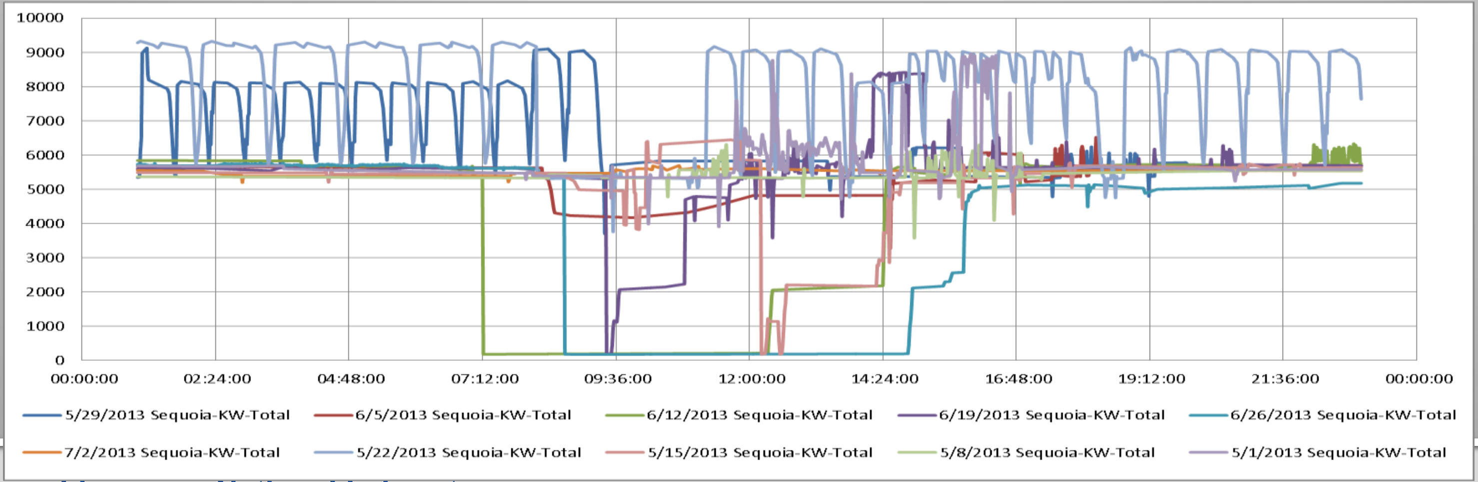
Bursty energy (power) usage



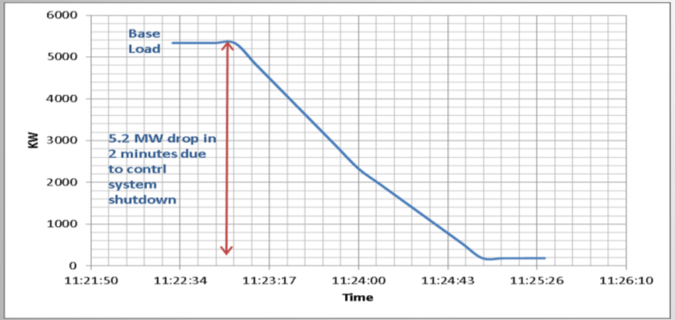
- Bringing the machine down for maintenance results in dumping over 5 MW of power back to the grid in a short period of time
- Bursty behavior of real workload, Power fluctuations are more abrupt

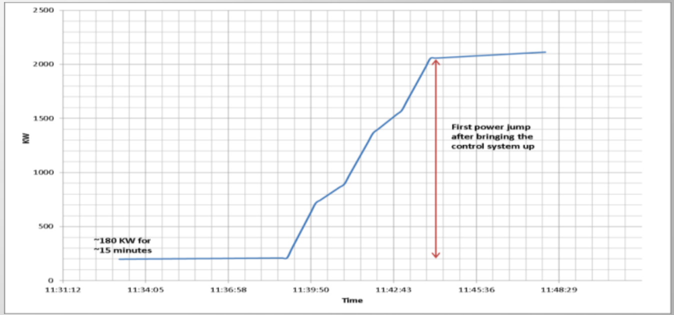
Characterizing the maintenance schedule

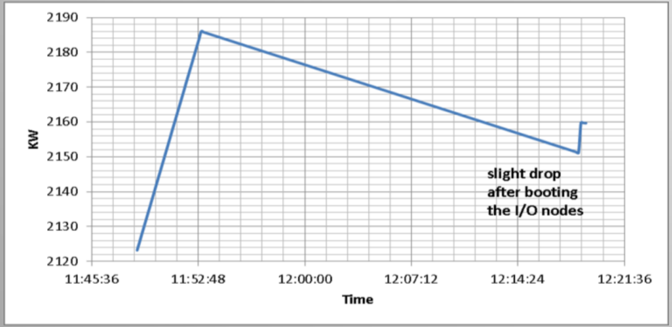
- Scheduled maintenance happens every Wednesday
  - Base load ~ 5MW
  - Can go down to 180KW or lower
  - Duration depends on what kind of maintenance will take place
  - ~ 50 seconds to go from 5.5 MW down to 100 KW.

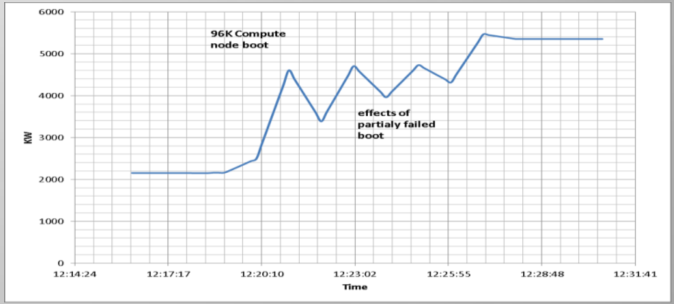


A closer look at the Sequoia shutdown events









PI Data Server Compression Study

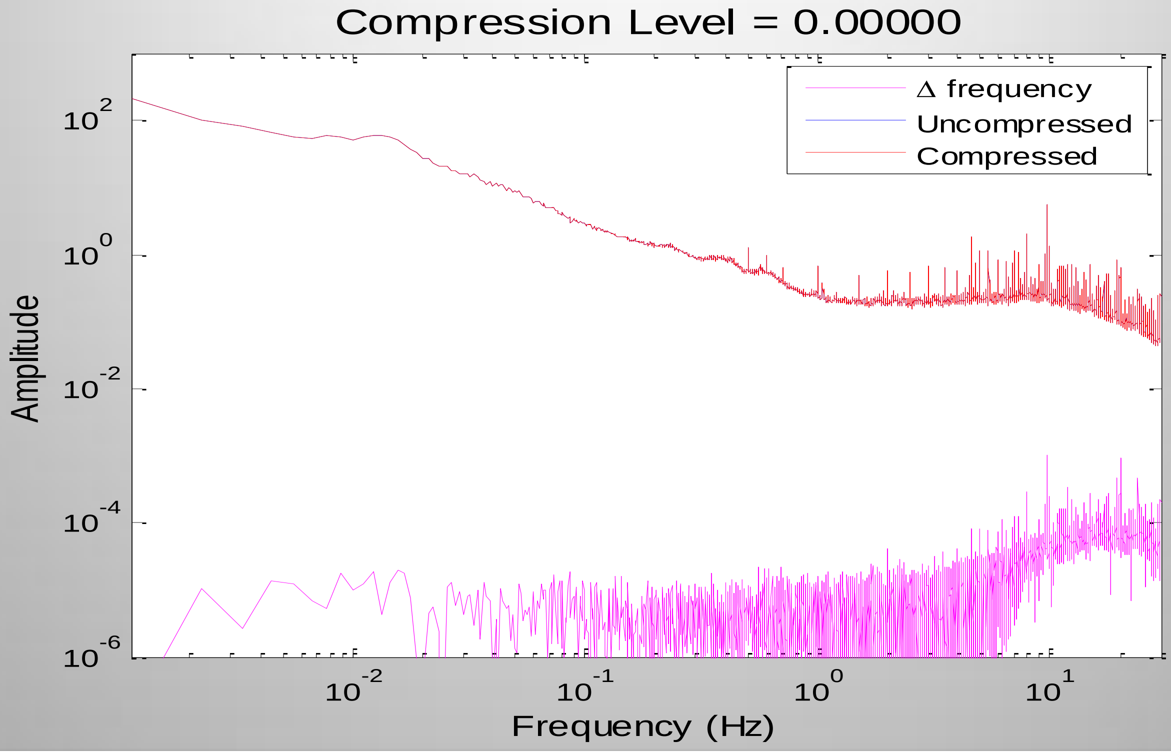
The question

- How does compression change the characteristics of the signal?
  - Frequency data from PMU device
    - — Arbiter Systems, model 1133A power Sentinel

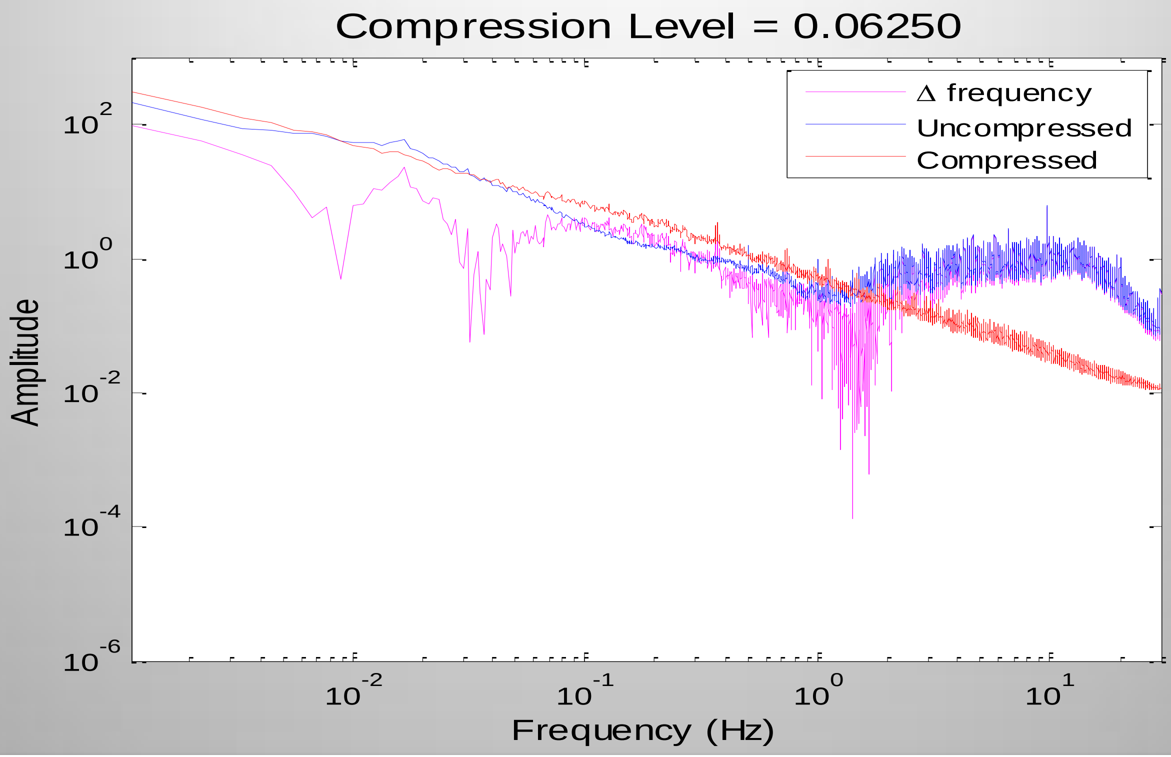
Experiment

- Collect raw data and store it in binary format on the file system
- Collect data into PI Data Server with different levels of compression
- Use FFT analysis to compare the signal with no compression and different levels of compression
  - Averaging 6 hours over 15 minute window

Distortion and compression level

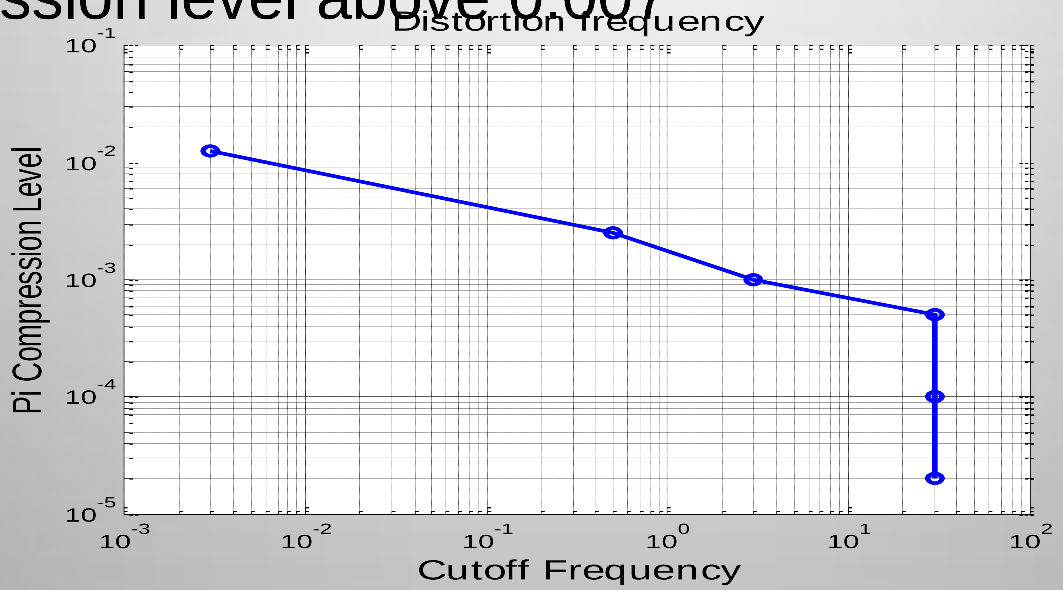


Distortion and compression level

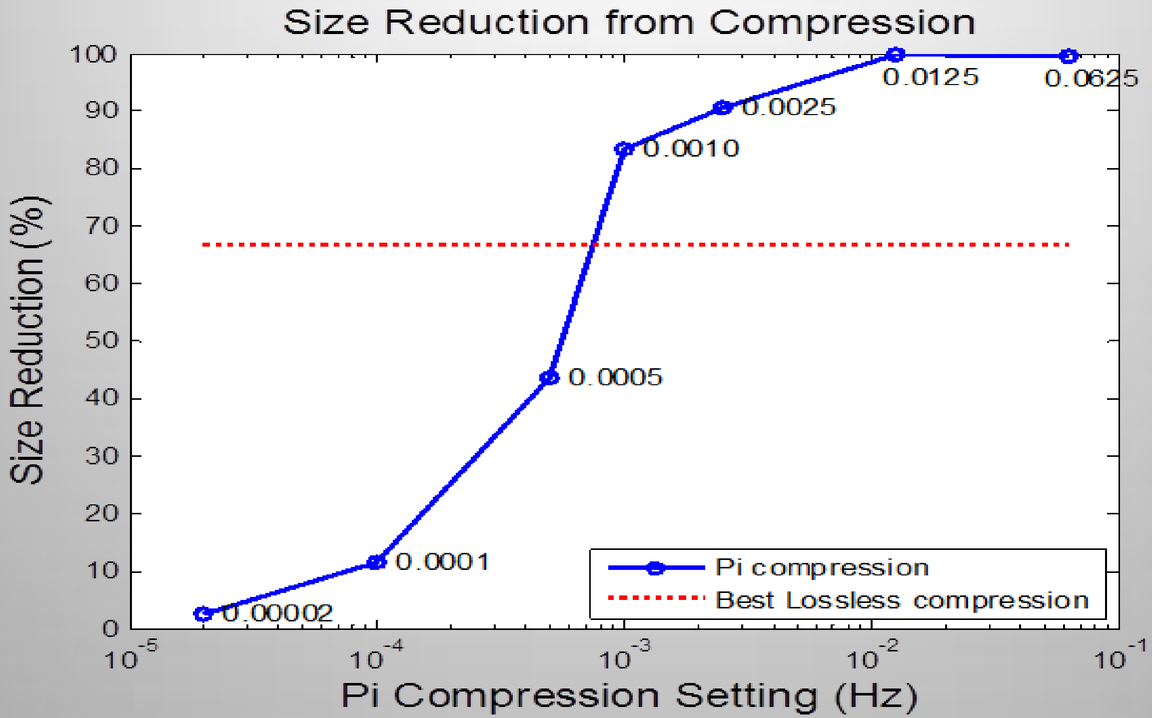


Distortion frequency

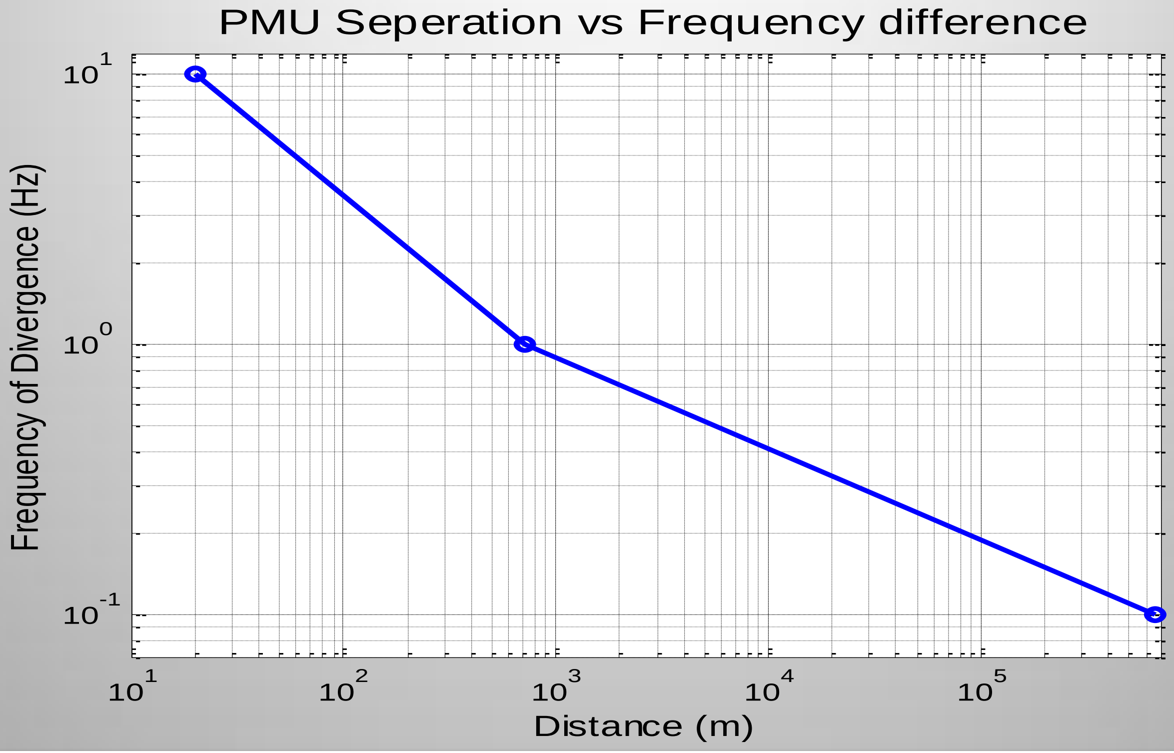
Frequencies higher than 10Hz will be distorted with compression level above 0.007



Compression Ratios



Distance & signal divergence



Ghaleb Abdulla

abdulla1@llnl.gov

Senior Computer Scientist

Director, Institute for Scientific Computing Research

Lawrence Livermore National Laboratory

Questions

Please wait for the microphone before asking your questions

State your name & company

THANK YOU