LLNL Sequoia-25 ISC13 Green500 Submission

Robin Goldstone, Anna Maria Bailey
Livermore Computing HPC Facility, LLNL
Sequoia Overview

- IBM BlueGene/Q System
  - 20 petaFLOP/s peak - 16.32 PF achieved
  - Memory 1.5 PB, 4 PB/s bandwidth
  - 98K nodes, 1.5M cores, 96 racks
  - 3 PB/s interconnect bandwidth
  - 0.5–1.0 TB/s Lustre bandwidth
  - 50 PB disk

- Power – 9.6 MW in 4,000 ft²

- #1 on June 2012 Top500 and Green500

  2,100.88 MFLOPS/W L1 submission (compute only, no network)
Sequoia Power Distribution

Utility Transformers  
Qty 6 @ 2000 KVA Transformers with Utility Metering.

Distribution Switchboards  
Qty 6 Distribution Switchboards, each feeding 16 underfloor PDUs.

Underfloor PDUs  
Qty 96 underfloor 480V PDUs, each feeding one Sequoia rack.
Sequoia Metering Components

Distribution Switchboard (x6)

Main Breaker Meter: Siemens 9510

16 Feeder Breaker Meters: Siemens MP636-EXTC
Data from Sequoia power meters, along with other LLNL HPC systems and facilities data, is collected and analyzed by the OSIsoft PI infrastructure management system.

The Siemens remote meter display unit (MDU) provides a local display of the energy readings. One MDU can support up to 256 meters. The MDU provides a quick way to view the energy information at an accessible location.
Sequoia-25

- Sequoia (20 PF peak) + Vulcan (5 PF peak)

- Systems were combined in early/mid 2013 to enable extreme-scale science calculations prior to transition of Sequoia to classified operations.

- Sequoia-25 achieved a Linpack result of 21.46 PF, however submission was rejected for inclusion in June 2013 Top500 list due to the fact that the system was subsequently de-coupled.
Despite not making the June 2013 Top500 list, Sequoia-25 still qualifies for Green500.

“In order to qualify for The Green500 List, a supercomputer must achieve performance at least as high as the 500th ranked system in the current Top500 List.”
Siemens meters report both instantaneous power (kW) and cumulative energy (kWh). When computing average power from the two data sets, we noted a discrepancy and subsequently discovered that 11 of our 120 meters were not correctly reporting their results to our data collection system.
## All L3 Requirements were otherwise met

<table>
<thead>
<tr>
<th>Aspect</th>
<th>L3 Requirement</th>
<th>Sequoia-25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect 1a: granularity of power measurements</td>
<td>Continuously integrated total energy</td>
<td>Met*</td>
</tr>
<tr>
<td>Aspect 1b: timespan of power measurements</td>
<td>A time series of equally spaced integrated total energy values</td>
<td>Met*</td>
</tr>
<tr>
<td>Aspect 1c: reported measurements</td>
<td>Core phase avg power, &gt;10 energy measurements within core phase, whole app avg power, idle power</td>
<td>Met</td>
</tr>
<tr>
<td>Aspect 2: machine fraction</td>
<td>Whole machine</td>
<td>Met</td>
</tr>
<tr>
<td>Aspect 3: subsystems included</td>
<td>All participating subsystems</td>
<td>Met</td>
</tr>
<tr>
<td>Aspect 4: power measurement point</td>
<td>Upstream of power conversion</td>
<td>Met</td>
</tr>
</tbody>
</table>

* Some meters failed to report due to misconfiguration.
Results

- **Sequoia-25 Linpack**: 21.41 PF
- **L2 Average Power**
  - Core phase: 11,501.98 MW
  - Full run: 11,470.29 MW
- **Green500 Result**: 1861.41 MFLOPS/W
  - ~12% less efficient than L1 measurement (with network excluded)
Takeaways

• In much the same way that preparing a Top500 submission can help shake out your HPC system hardware, preparing a Green500 submission can help shake out your facility metering infrastructure.

• … But it would have been better to have discovered misconfigured meters beforehand.

• LLNL’s multiple levels of metering (per transformer, per main breaker panel and per PDU) have proven to be very useful, allowing for cross-correlation of measurements as well as isolation of faulty meters and anomalous power conditions.