THE WORLD'S FIRST PETASCALE ARM SUPERCOMPUTER

Sandia National Laboratories

ASTRA

"Per aspera ad astra"

Green 500 BOF Presentation
Astra and Attaway Power Measurement

David Martinez
davmart@sandia.gov

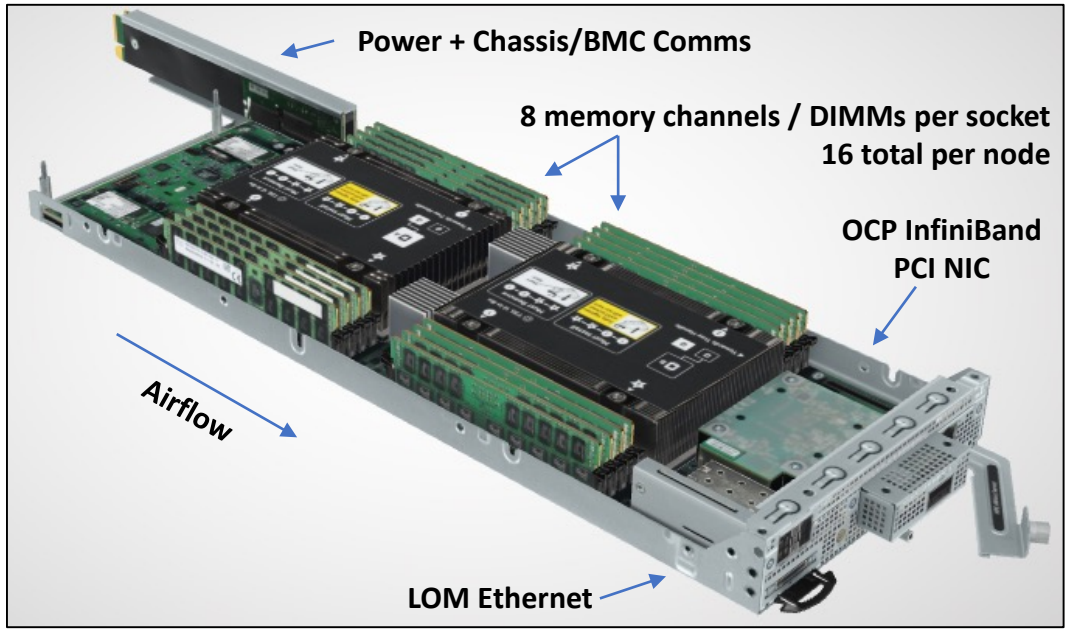
U.S. DEPARTMENT OF ENERGY NNSA

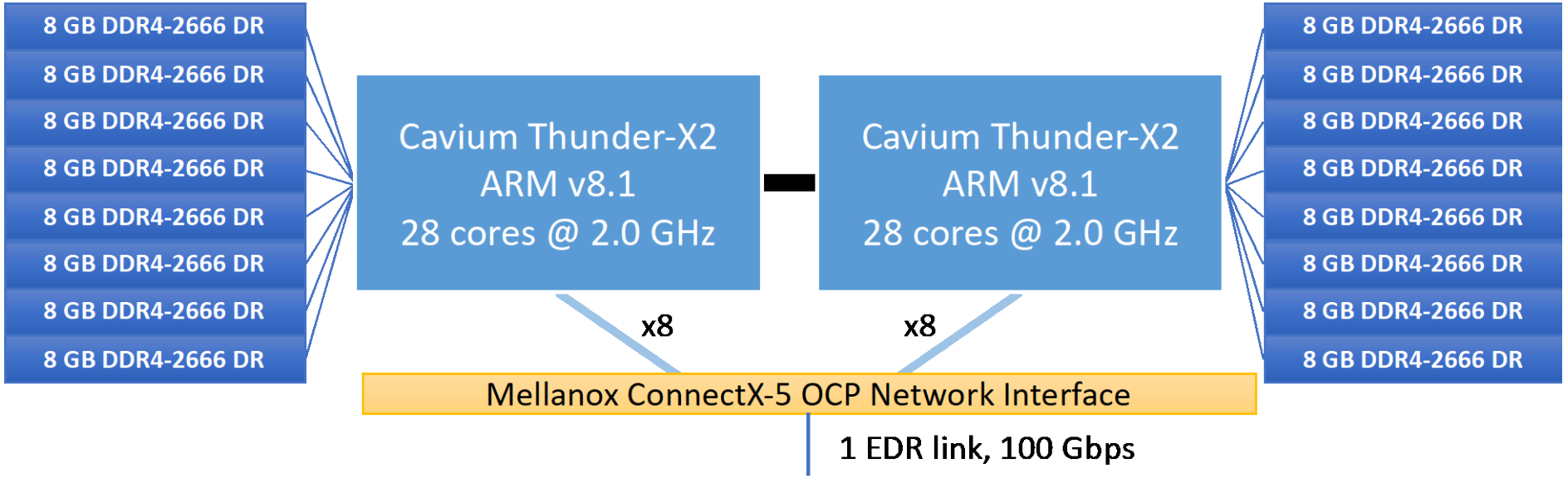
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525 Unclassified Unlimited Release .

ASTRA
VANGUARD

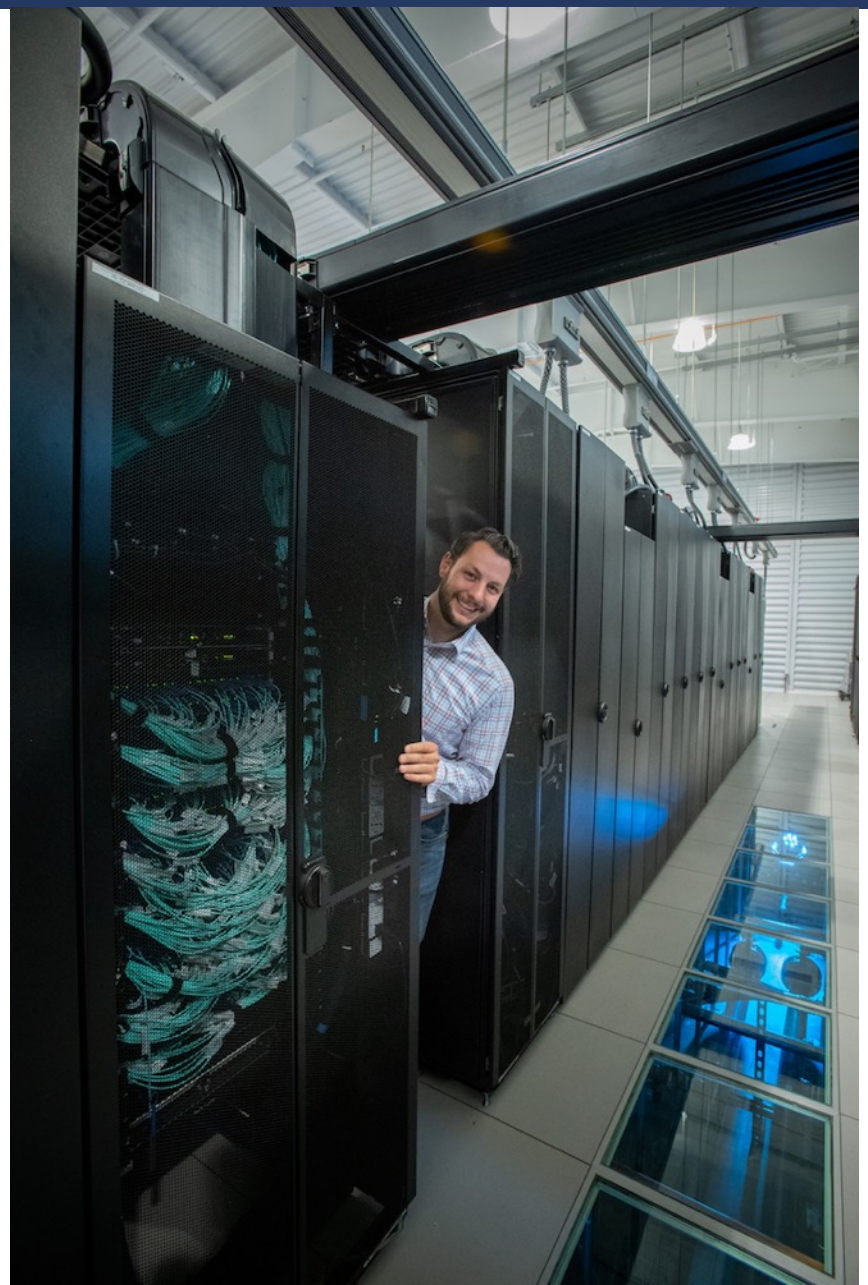
Astra Node Architecture

- **2,592** HPE Apollo 70 compute nodes
  - Cavium Thunder-X2 **Arm** SoC, 28 core, 2.0 GHz
  - 5,184 CPUs, 145,152 cores, 2.3 PFLOPs system peak
  - 128GB DDR Memory per node **(8 memory channels per socket)**
  - Aggregate capacity: 332 TB, Aggregate Bandwidth: 885 TB/s
- Mellanox IB EDR, ConnectX-5
- HPE Apollo 4520 All–flash storage, Lustre parallel file-system
  - Capacity: 990 TB (usable)
  - Bandwidth 244 GB/s





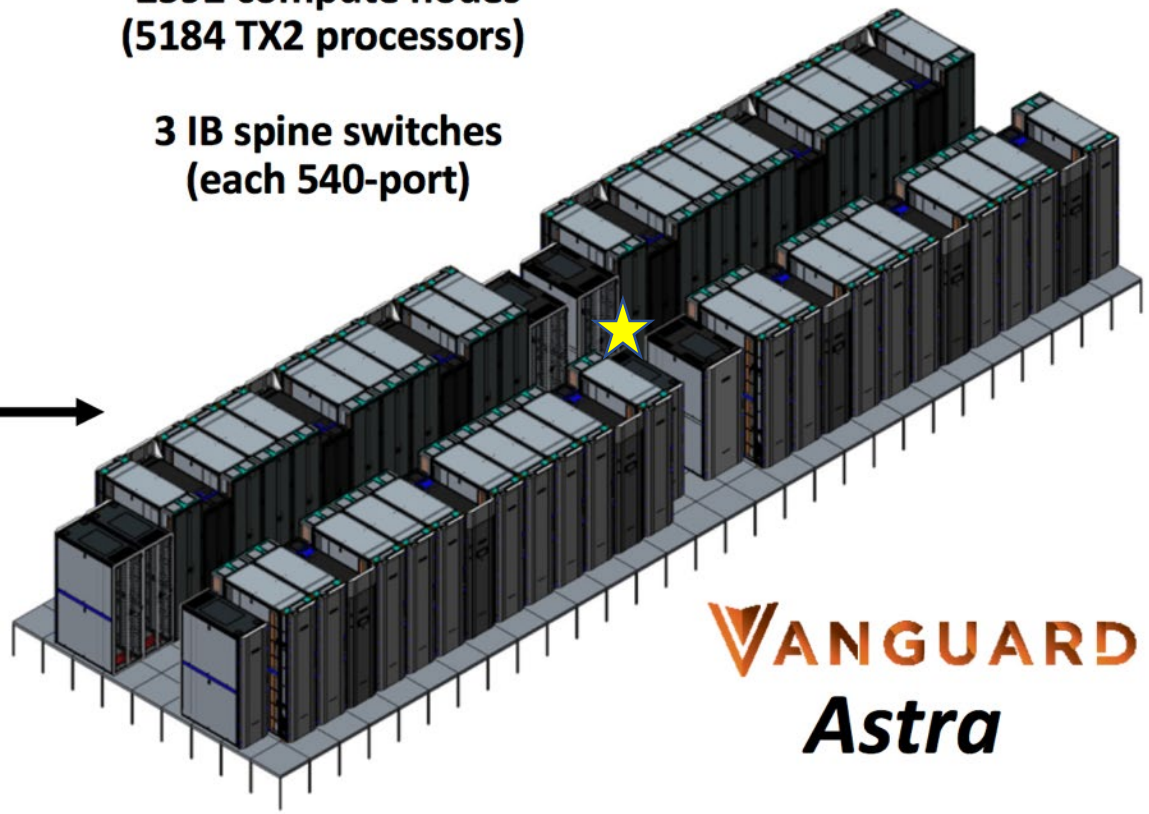
Astra System Architecture



36 compute racks
(9 scalable units, each 4 racks)

2592 compute nodes
(5184 TX2 processors)

3 IB spine switches
(each 540-port)



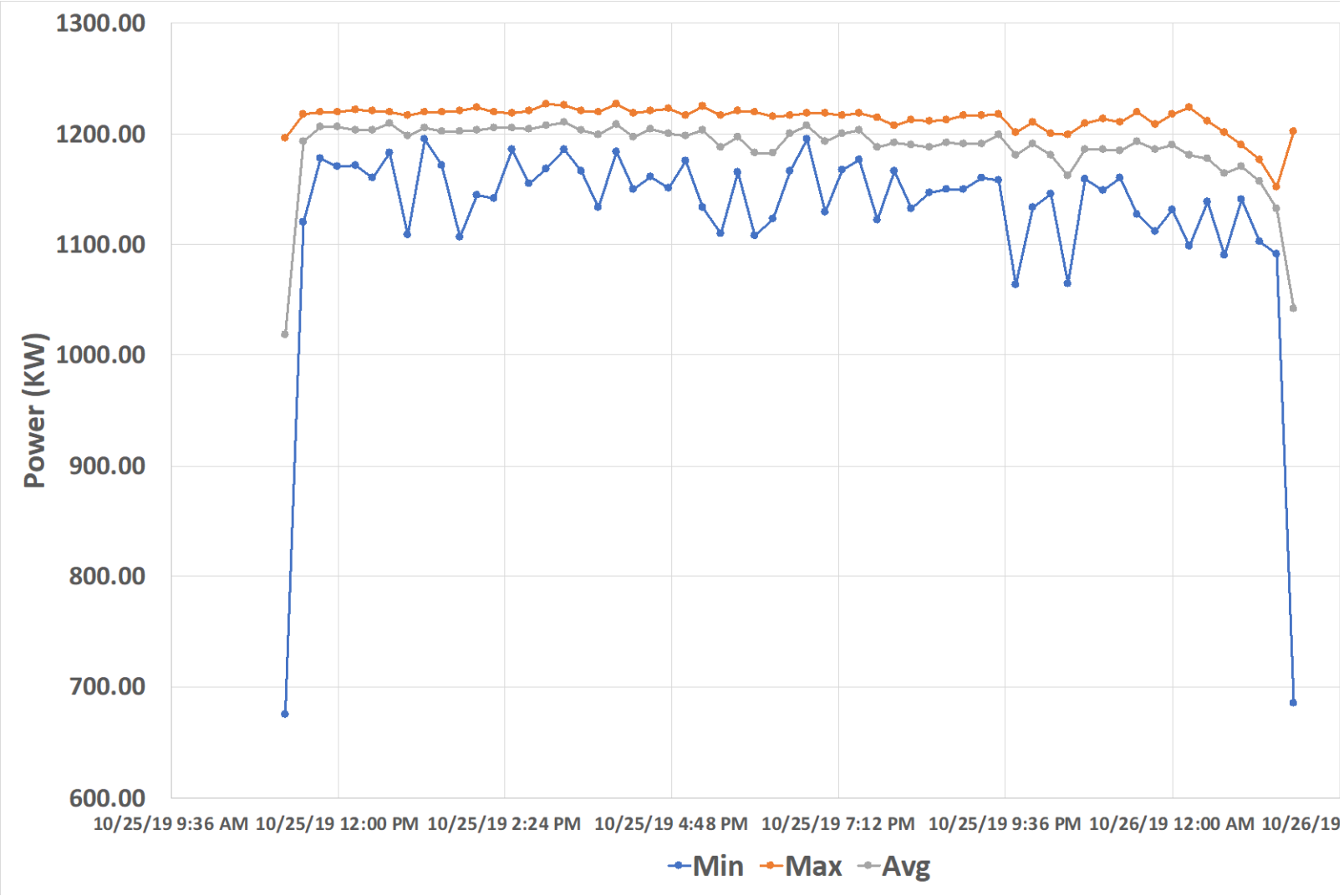
VANGUARD
Astra

Astra Power Measurement Capabilities

- Node-level
  - In-band TX2MON Linux driver and tool cpu/mem/uncore power usage
  - Out-of-band BMC power measurement for whole node
- Rack-level
  - PDU power + energy measurement of full rack and each breaker
- System-level
  - Starline Busway T5 1200 A for 480 V (compute nodes, 1 bus bar per row)
  - Liebert PDUs for 208 V (Cooling, I/O, InfiniBand, and service nodes)

System-level Used for Level 3 Astra HPL Power Measurement

Astra HPL Power Nov Top500 Submission



- HPL Performance:
  1.833 PF on 2565 nodes
  (143,640 cores)
- Core HPL Region Average Power:
  1192.54 KW
- Power Efficiency:
  1.833 / 1192.54 = 1.537 GFlops/Watt
- Peak Power During HPL Run:
  1227 KW
- Idle Power:
  662 KW

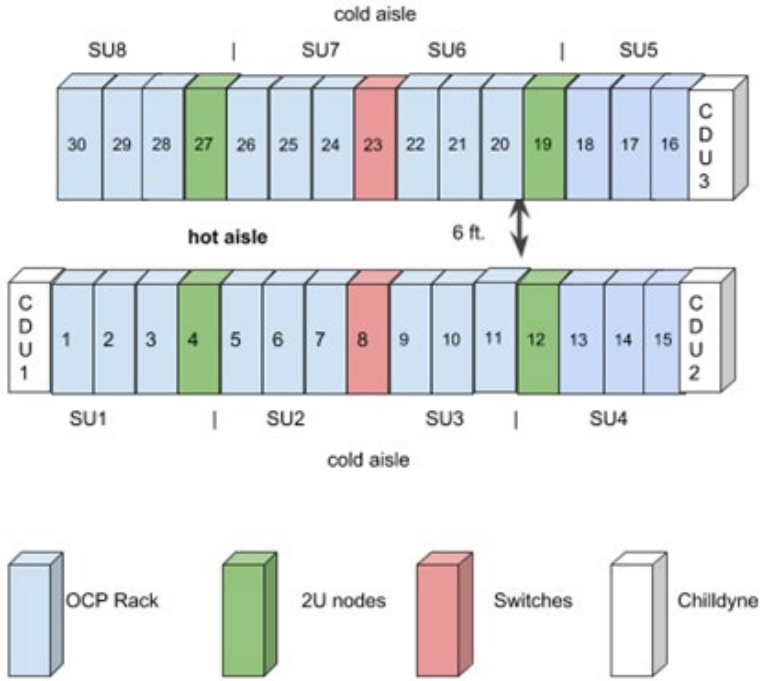
“Attaway”

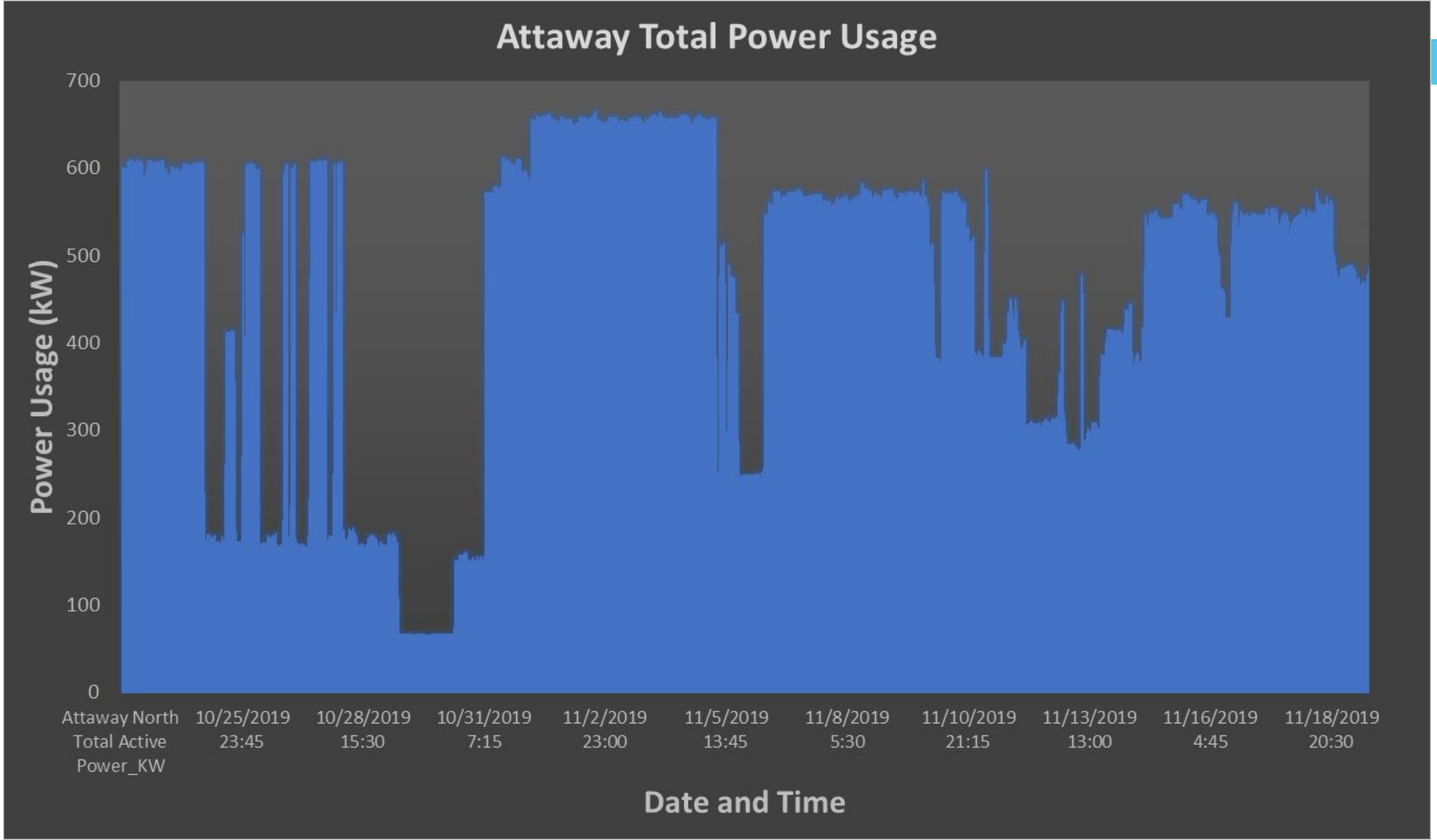
Hardware facts:

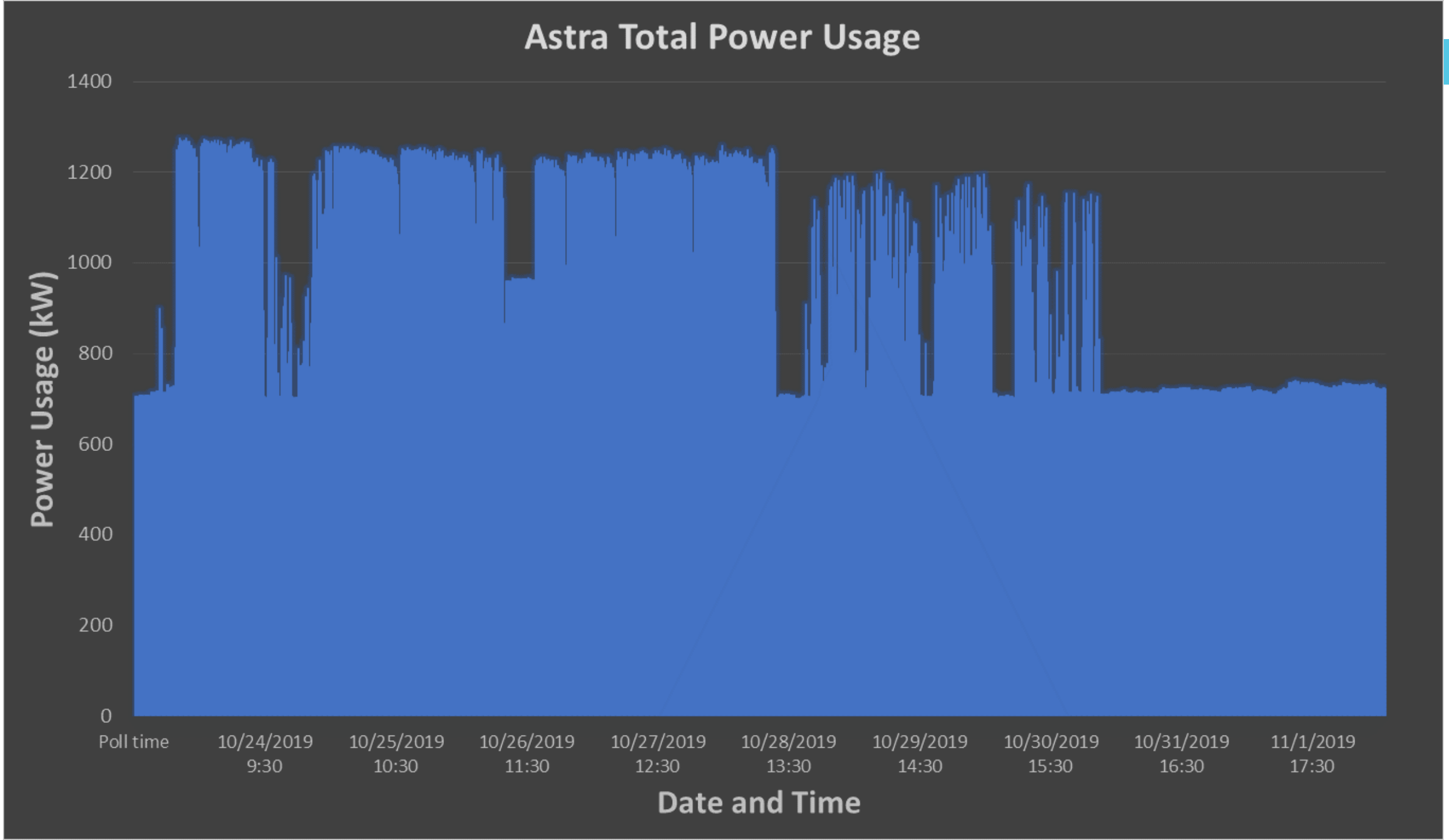
- **1,488 compute nodes (8 scalable units)**
  - **2.7 PFLOPs measured**
  - **Fastest production cluster at Sandia**
- 2.3 GHz processors
  - Dual sockets with 18 cores each
  - Intel Xeon Gold 6140 “Skylake”
- 192 GB RAM per node (5.3 GB per core)
- Intel Omni-Path high speed interconnect
  - 2:1 oversubscription
  - Mellanox ConnectX5 HCA’s in IO nodes
- Penguin Tundra™ hardware platform
  - Open Compute (OC) 21’’ racks
  - 3 compute nodes slotted in 1 Open Unit (OU)
- 480 VAC direct to rack power shelf
  - 12 Volt DC power distribution within the rack
  - In rack monitoring is provided by Penguin's in-rack P1000 Rack Controllers (small controller using a Compute RPI running Linux) - provides power monitoring, supply/return temperature, leak detection and remote rack EPO
- Next generation water cooling:
  - Processors are directly cooled by cold plate
    - Fins provide additional air cooling/heat exchange
  - Floor based CDU’s provide negative pressure to prevent leaks and to improve system resilience.
  - ~60-70% of node heat is removed via water
  - System can run with 2 of 3 CDU's using Chilldyne's automatic failover valve system

Software facts:

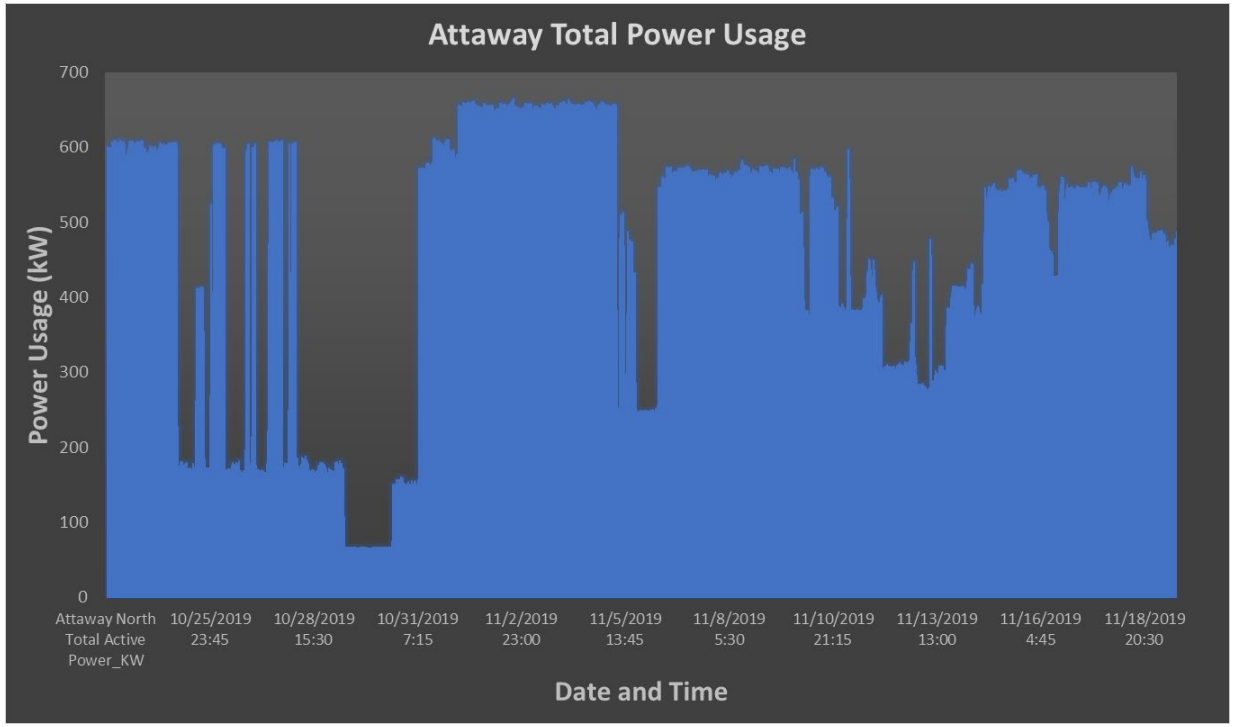
- TOSS 3.x.x (Red Hat EL 7.x)
- Lustre 2.x client
- Intel Fabric Manager
- Diskless boot over gigE private LAN
- Slurm for user job scheduling

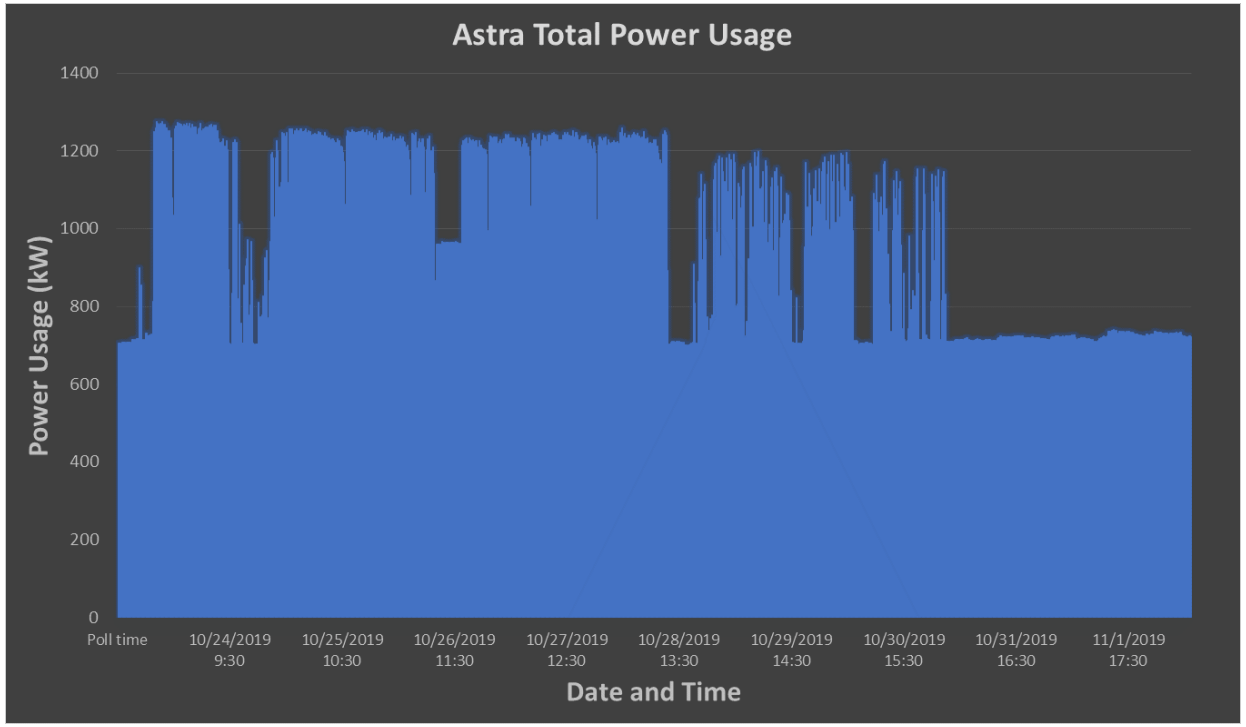



Attaway Total Power Usage
Power Usage (kW)
700
600
500
400
300
200
100
0
Attaway North Total Active Power_KW
10/25/2019 23:45
10/28/2019 15:30
10/31/2019 7:15
11/2/2019 23:00
11/5/2019 13:45
11/8/2019 5:30
11/10/2019 21:15
11/13/2019 13:00
11/16/2019 4:45
11/18/2019 20:30
Date and Time


Astra Total Power Usage
Power Usage (kW)
1400
1200
1000
800
600
400
200
0
Poll time
10/24/2019 9:30
10/25/2019 10:30
10/26/2019 11:30
10/27/2019 12:30
10/28/2019 13:30
10/29/2019 14:30
10/30/2019 15:30
10/31/2019 16:30
11/1/2019 17:30
Date and Time

Attaway vs. Astra Power Usage





Questions?

Sandia National Laboratories

Exceptional Service in the National Interest