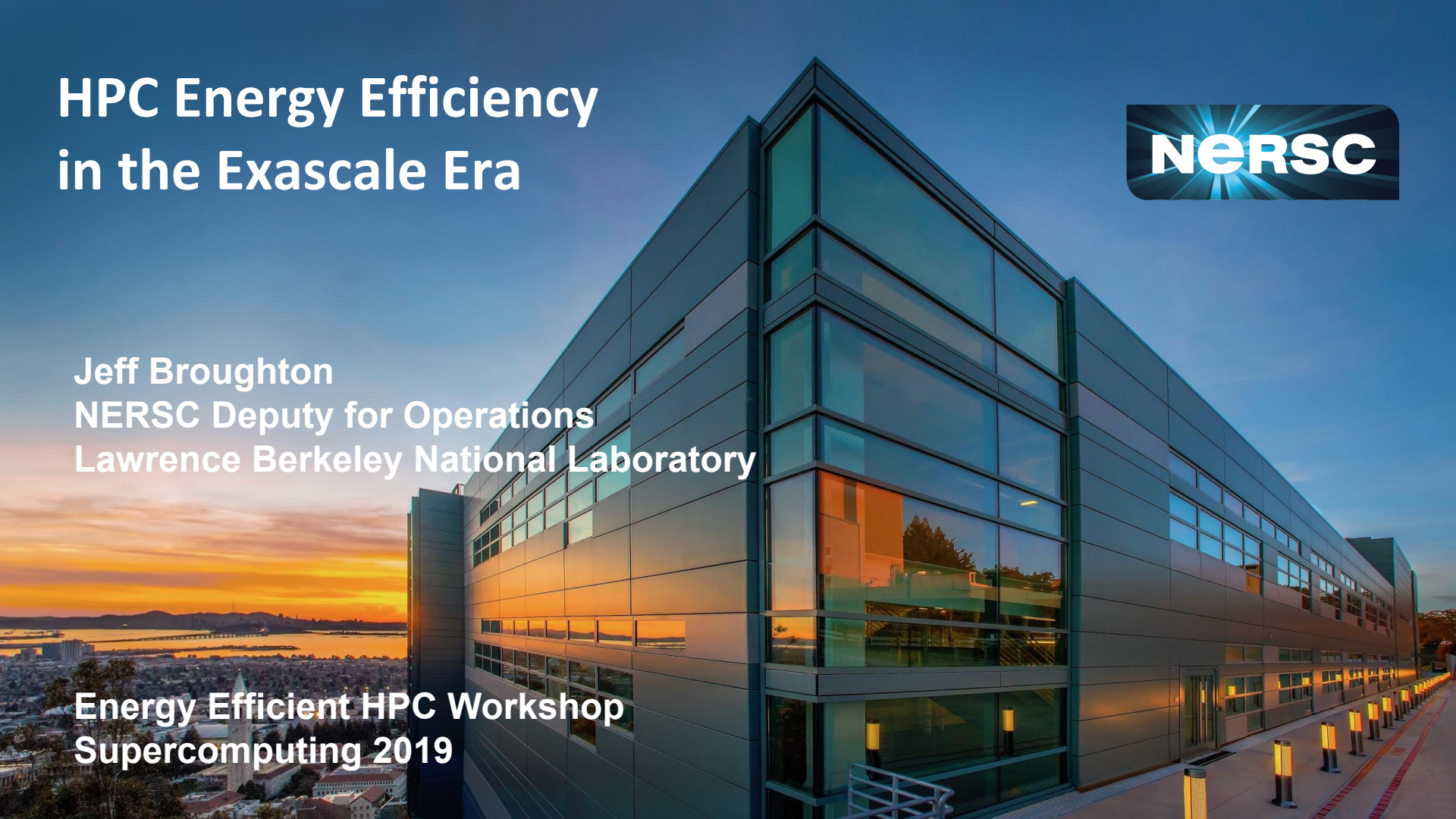
HPC Energy Efficiency in the Exascale Era



Jeff Broughton
NERSC Deputy for Operations
Lawrence Berkeley National Laboratory

Energy Efficient HPC Workshop
Supercomputing 2019

Exascale by the Numbers

1,000,000,000,000,000,000 flops / sec

500 x U.S. national debt in pennies

100 x number of atoms in a human cell

2 x number of seconds since the Big Bang

1 x number of insects living on earth

Exascale by the Numbers

> 1,000,000,000,000,000,000 flops / sec

> $600,000,000 life cycle cost *

~ 200,000,000 kWhr / year

~ 150,000 $MTCO_2e$ / year

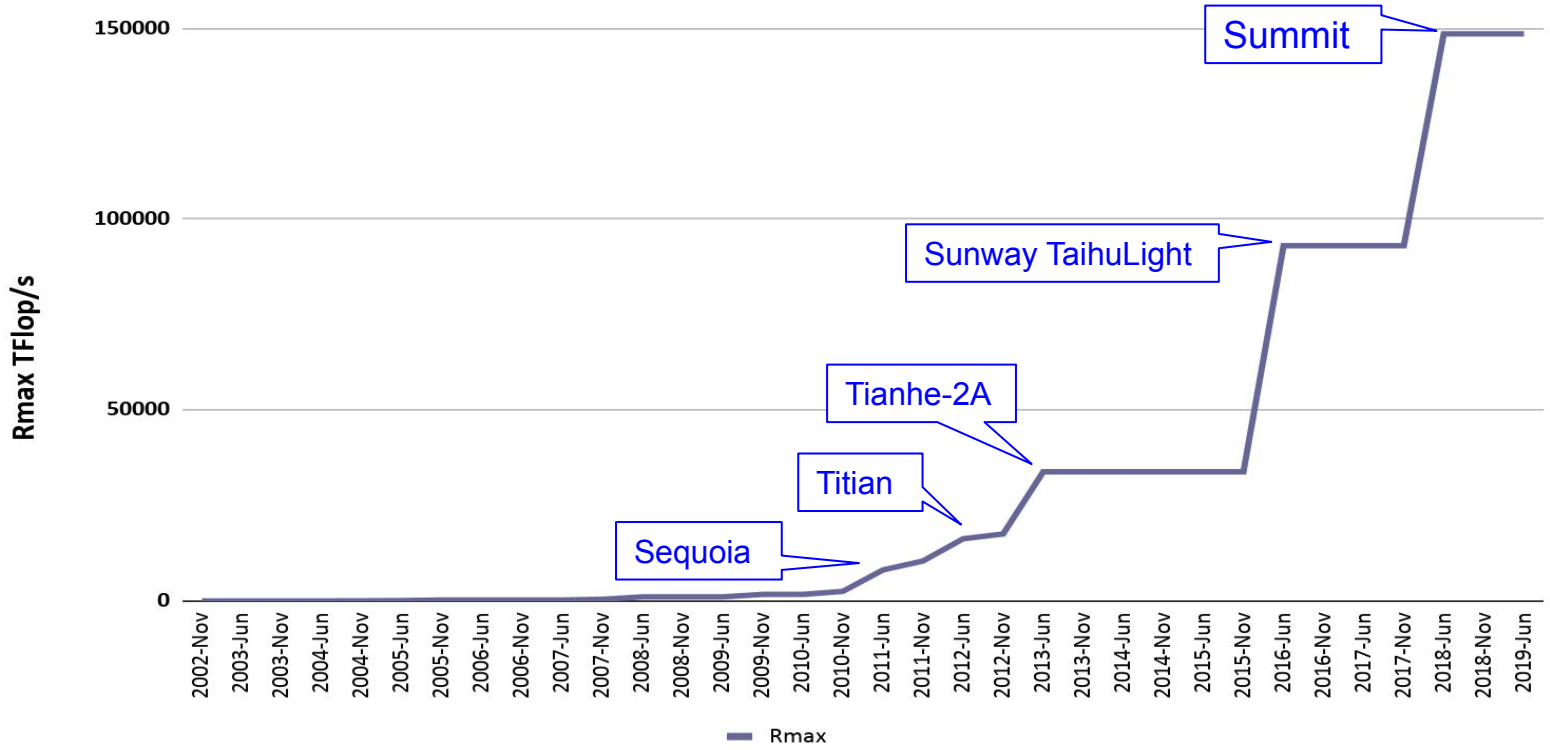
< 100 apps ready

(*) Not including NRE, construction or staff time

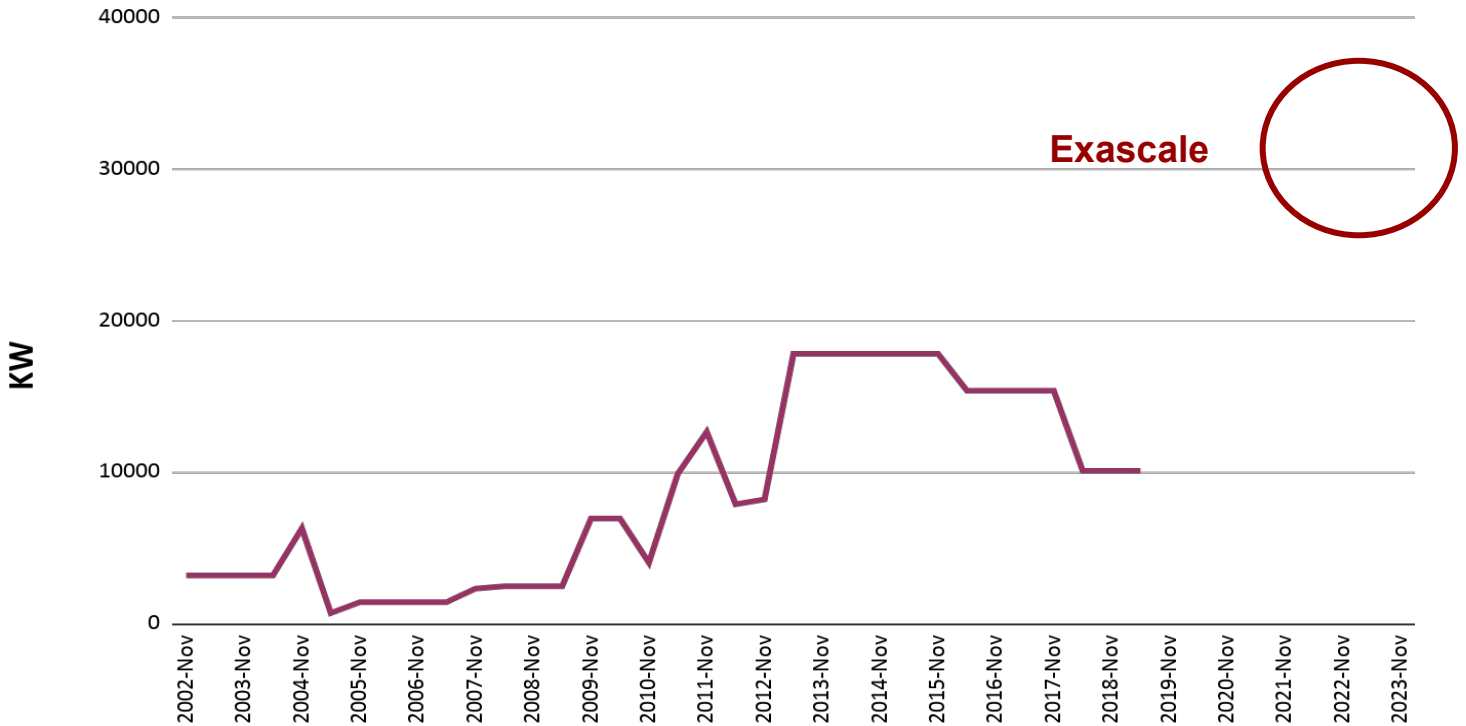
NERSC

Four facets of energy efficiency

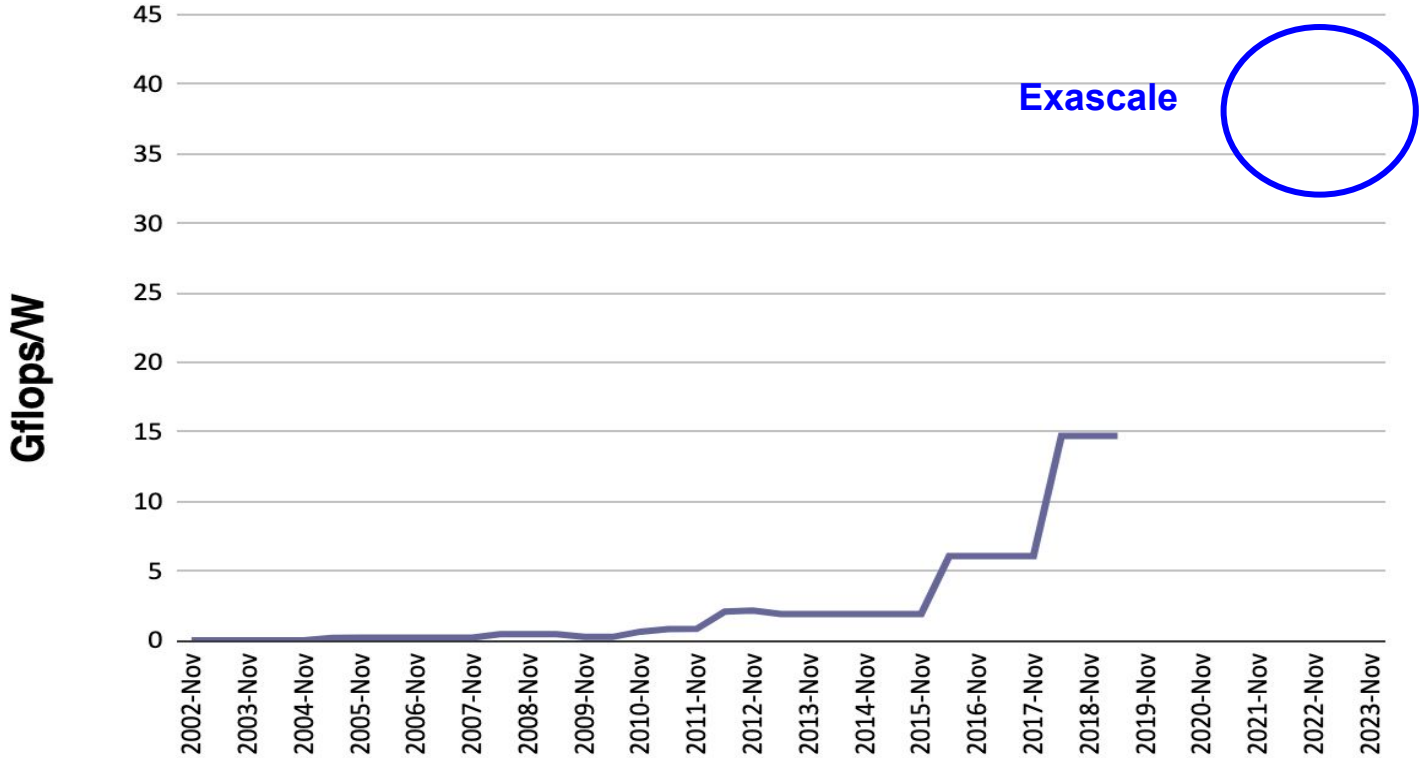
Performance of Top 1 System (Rmax)



Power consumption of Top 1 system



Power efficiency of Top 1 system



Technology evolution has delivered substantial energy efficiency improvements over time

Successes

- Moore's law
- Volume Server Technology
- GPUs!
- Incremental Improvements
- Bigger Budgets!!

Noble Attempts

- BlueGene
- Knights Family
- Hero systems

Near Future Wins?

- Flash and storage class memory
- Specialized accelerators

Four facets of energy efficiency

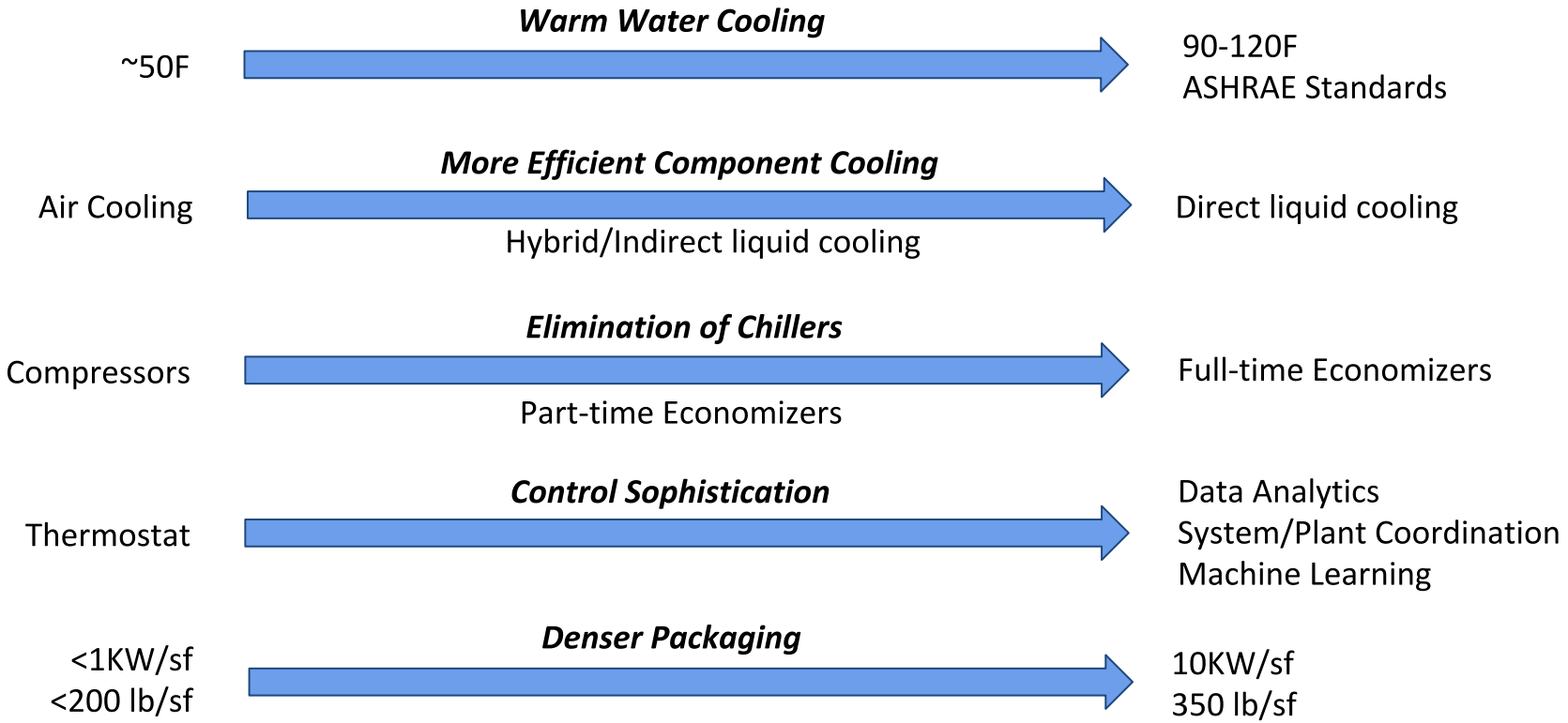
How much power does your system use?

2 year history of Cori power consumption

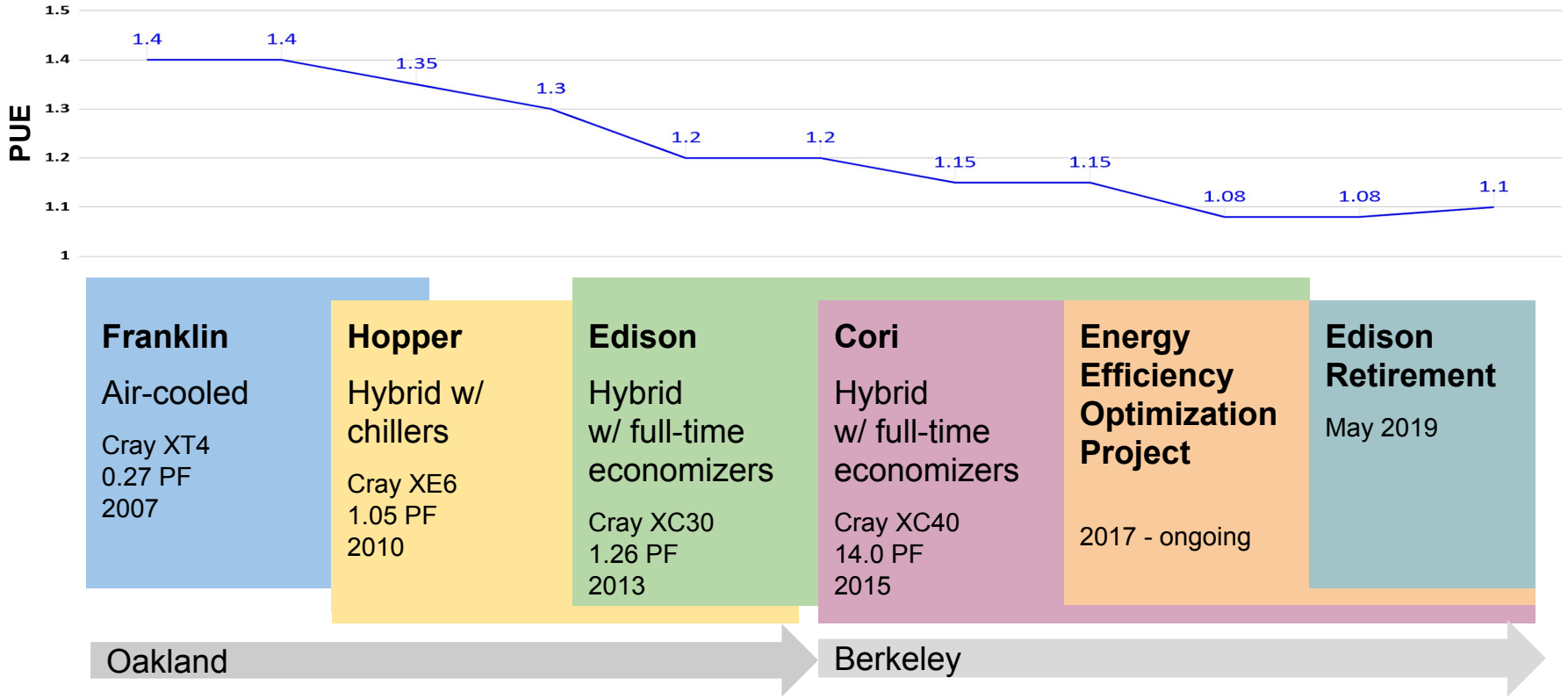
Cori rated 5.7 MW, observed peak 4.6 MW, typical 3.9 MW, idle 2 MW

Underused capacity is an economic and efficiency problem

Big Trends in Packaging and Energy Efficiency

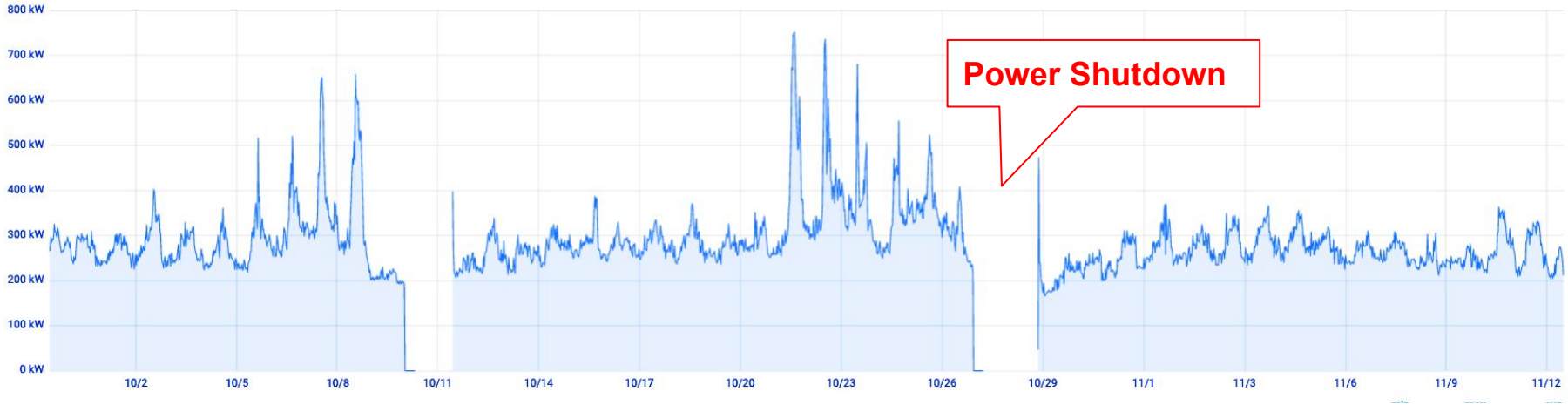


Incremental improvements to energy efficiency over the years



New cooling methods require sophisticated controls

45-day history of Cori physical plant power consumption



Responds to both system load and to weather conditions

Varies CW temp in response to wetbulb temp to optimize fan and pump energy

Varies CW pump speed in response to system load

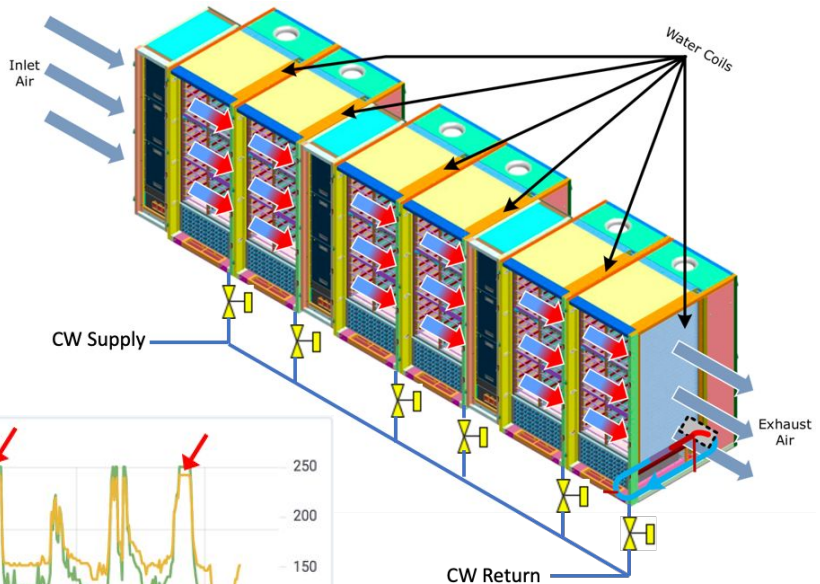
System and infrastructure must communicate

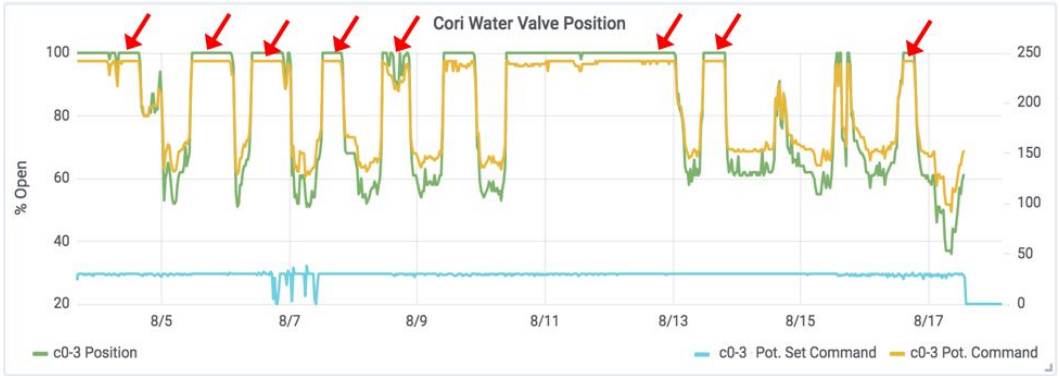
Out of the Box Operation: ***Static cabinet air temp setpoint***

- Can't set for all weather, therefore CW pumping and fan energy waste

Interactive Operation: ***Dynamic cabinet air temp setpoints***

- Cooler CW temps - Cooler cabinet air therefore fan speed turndown
- Warmer CW temps - Fan speeds compensate. Dynamic setpoint reduces excessive CW demand valve positions





Gather the data and the Q's will come

OMNI Data Collection System

Current Data Sources

- Substations, panels, PDUs, UPS
- Cray XC internal SEDC data
- Onewire Temp & RH sensors
- BMS through BACNET
- Indoor/Outdoor Particle counters
- Weather station

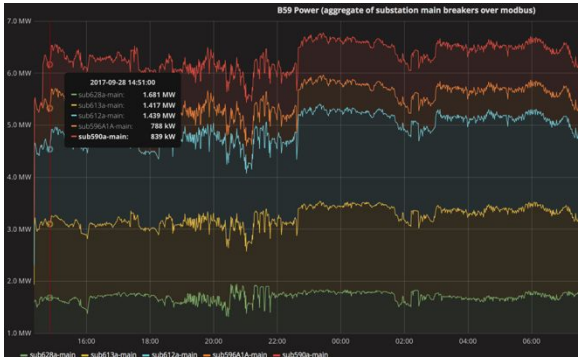
System Data Sources (WIP)

- Syslog
- Job Data
- Lustre + GFPS statistics
- LDMS

Rabbit MQ, Elastic, Linux

- Collects ~20K data items per second
- Over 100TB data online
- 45 days of SEDC (versus 3 hours on SMW)
- 180 days of BMS data (6X more than BMS)
- > 2 years of power data

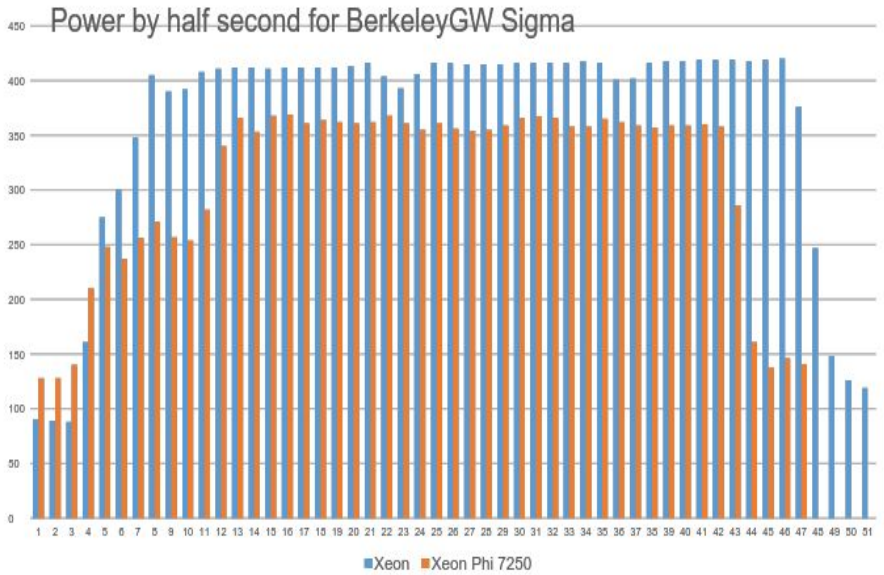
Kibana, Grafana, Skyspark



Four facets of energy efficiency

Reducing time to solution is the first order approach to use energy efficiently

- Application Energy Efficiency ~ 1 / Wall-Time
- Potential application power savings << typical power consumption
- Goal of capability/capacity systems is more science for ~ same power



Software optimizations can improve performance and reduce energy

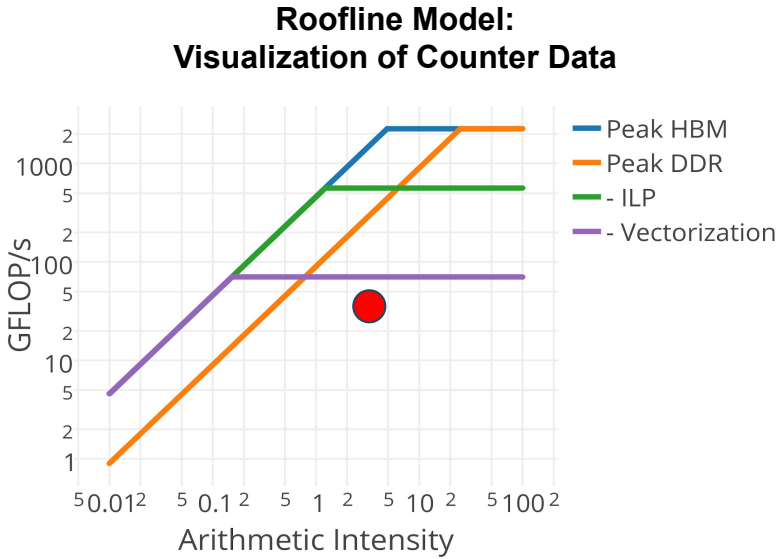
Energy-efficient processors have multiple HW features to optimize for both performance and energy efficiency

- Many (Heterogeneous) Cores
- Bigger Vectors
- New ISA
- Multiple Memory Tiers

Opportunities for energy reduction:

- Reduce CPU power consumption in Memory or IO intensive code-regions (and vice versa)
- Avoid highly masked or inefficient vector instructions
- Change algorithm to minimize data motion

Large gains possible by connecting users to actionable performance data



Major efforts underway to optimize applications for exascale-class systems

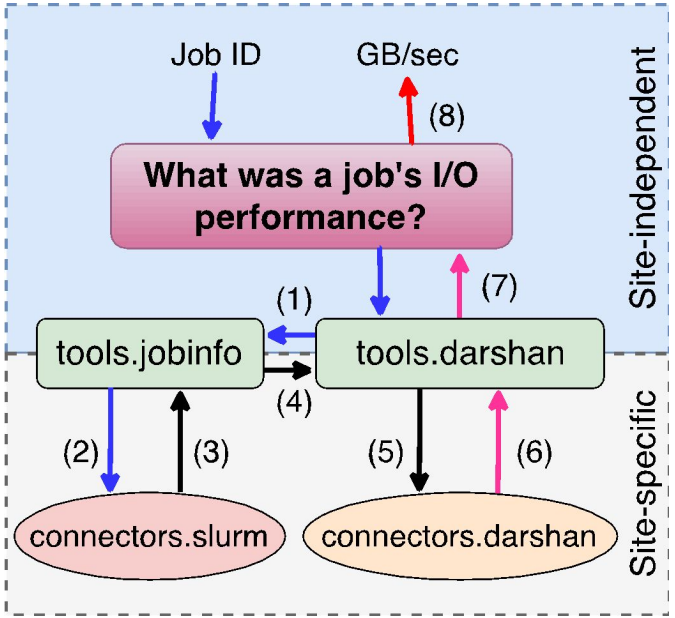
- Multi-year, cooperative efforts
  - DOE National Labs
  - Vendors
  - Application teams
- Approach
  - Domain and computer scientists
  - Focused optimization sessions
  - Dedicated app team support
  - Education, training and outreach
- Parallel effort on performance portability

GPU For Science Days


GPU Community Hack-a-thon

System software opportunities to improve operational efficiency

- Power saving
  - Idle time reduction
  - Detection of failing / aggressor jobs
- Power management
  - Power capping / band limiting
  - Automatic demand-response
- Data Collection
  - Job, application and I/O statistics

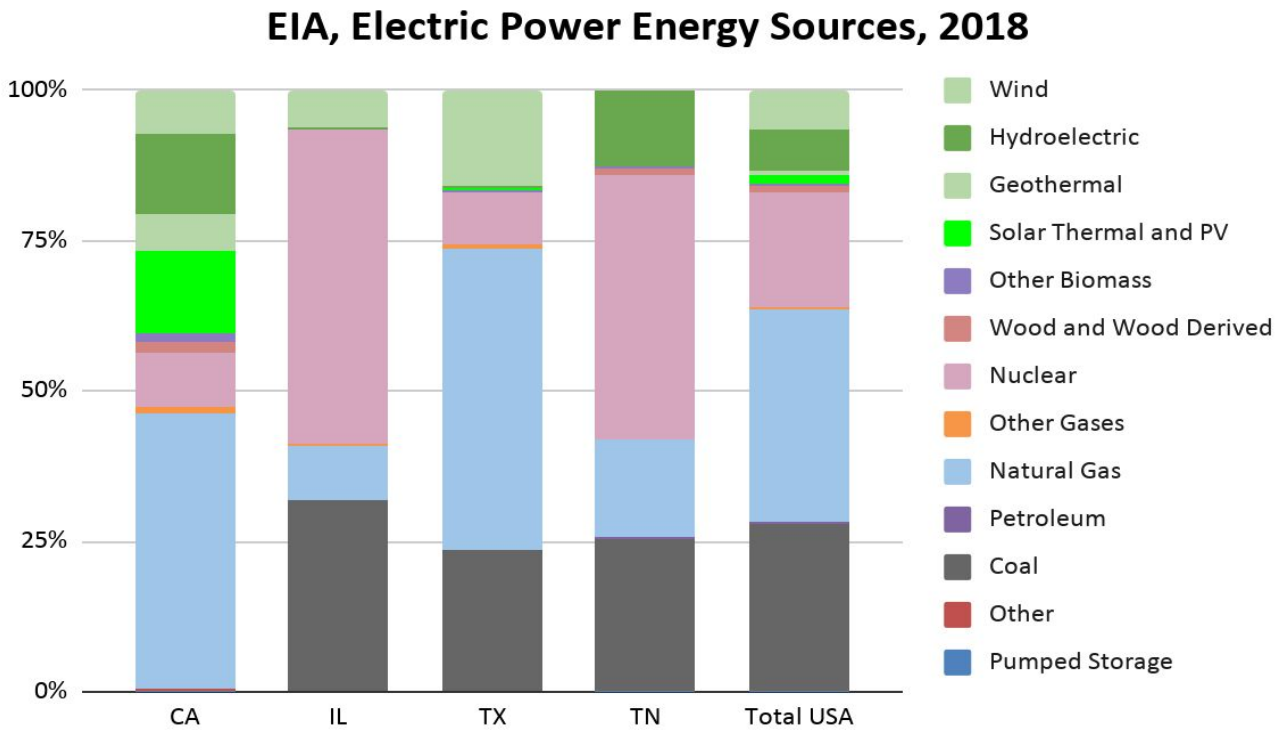


G. K. Lockwood et al, "TOKIO on ClusterStor: Connecting Standard Tools to Enable Holistic I/O Performance Analysis," in Proceedings of the 2018 Cray User Group. 2018.

Four facets of energy efficiency

NERSC

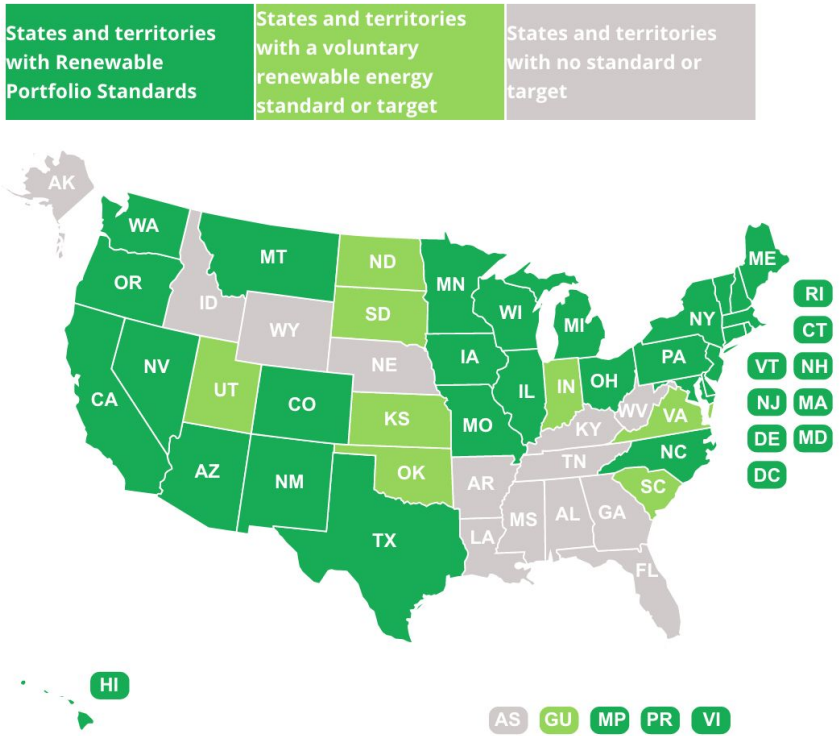
HPC needs to use greener energy



Greenhouse gas mitigation requirements are increasing

Renewable Portfolio Standards (RPS)

- California + 28 states + DC have mandatory programs
- 8 more have voluntary programs.
- ~ one half of renewable energy growth is due to these standards



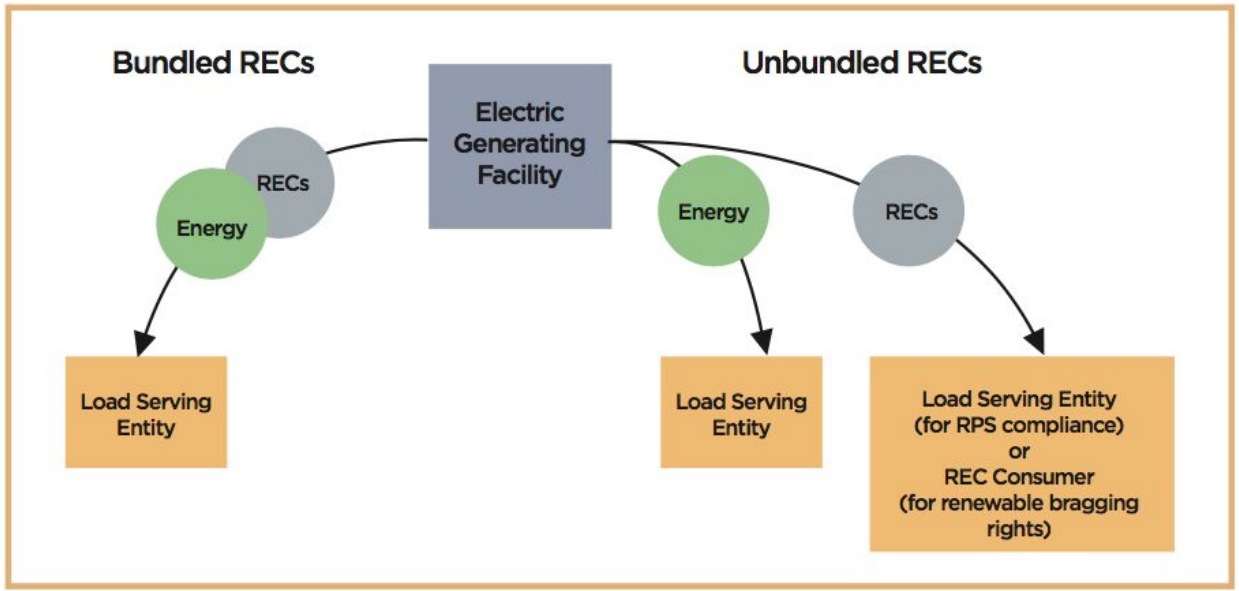
Source: National Conference of State Legislatures
http://www.ncsl.org/research/energy/renewable-portfolio-standards.aspx

Greenhouse gas mitigation requirements are increasing

California Environmental Quality Act (CEQA) creates a legal obligation to meet GHG mitigation goals

- Must buy Renewable Energy Credits (RECs) to mitigate nearly all GHG for NERSC-9
- University of California developing renewable energy projects to achieve net zero GHG by 2025

RECs come in many forms and prices



Pinkel and Weinrub, 2014, What the Heck is REC

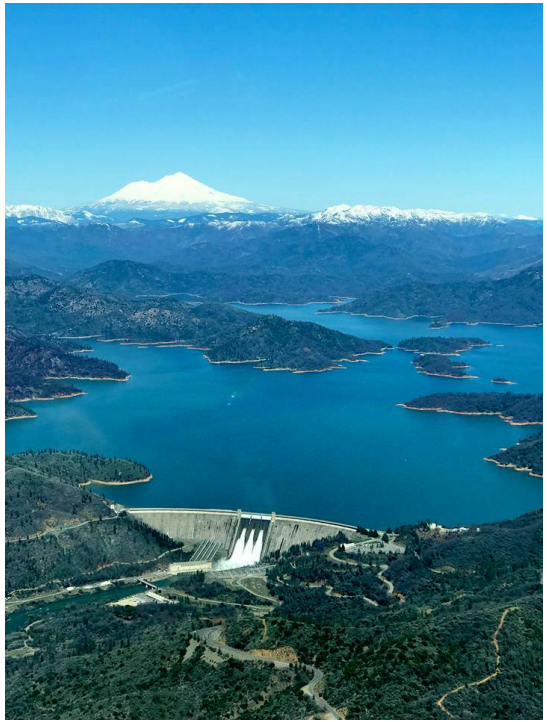
- Electricity and RECs are the same product, on the same grid
- Recent historical typical price: 1.5¢ to 4¢ per kWh

- Environmental attribute is separate from the electricity
- Recent historical typical price: 0.1¢ to 2¢ per kWh

Renewables are becoming less expensive than conventional power.

Can we save water too?

- Cooling towers consume up to 3.5M gallons per year per MW
- ~14 $MTCO_2e$ per MW from energy to deliver the water
- Alternatives to cooling towers
  - Dry coolers
  - Thermosyphons
  - Geothermal cooling
  - Deep water cooling



Final thoughts

Energy efficient improvements for HPC require:

- A holistic approach
- Some breakthroughs... but mostly incremental improvements
- Data analytics to drive and measure progress

Thank you!