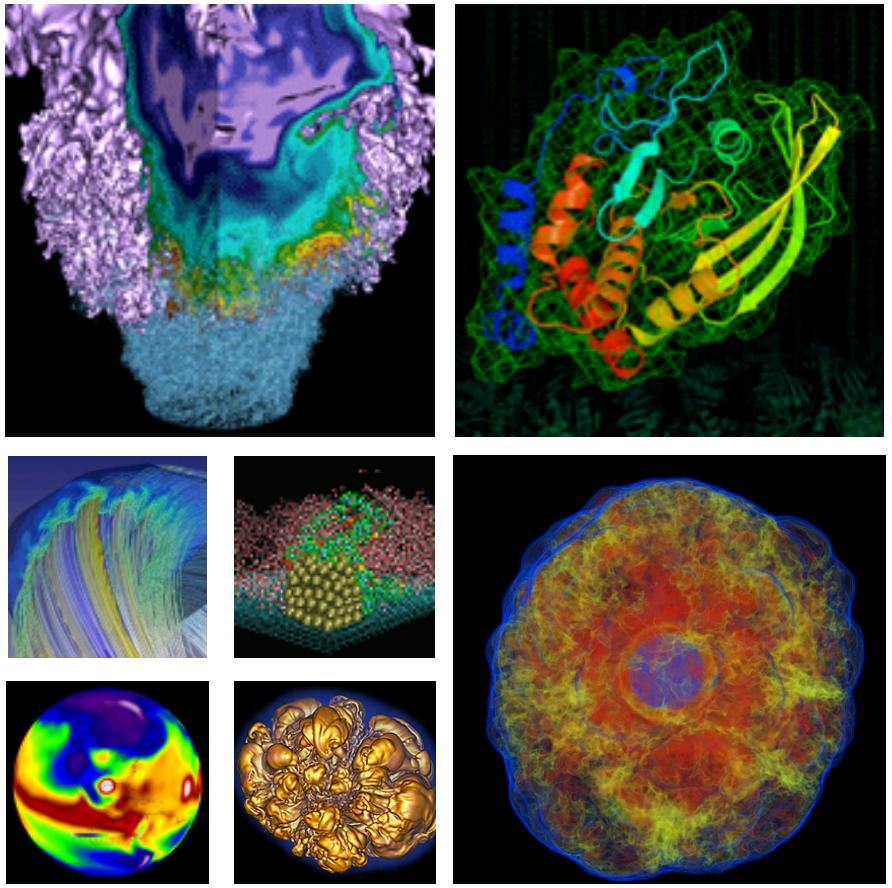
Edison Update



Richard Gerber, NERSC 7 Deputy
Tina Declerck, NERSC 7 Systems Lead
Zhengji Zhao, NERSC 7 Services Lead

February 12, 2013

Edison - Cray XC30 Phase 1 [Phase 2]

- 10K [104K] compute cores
- Cray ”Aires” interconnect
- Two 8-core Intel ’Sandy Bridge' 2.6 GHz processors per node [Phase 2 TBA]
- 16 [TBA] processor cores per node
- 64 GB of memory per node
- 42 [333] TB of aggregate memory
- Edison Phase 1 access Feb. 2013 [Phase 2 in Summer or Fall 2013]

- 4 [TBA] GB memory per core for applications
- 1.6 / 6.4 PB of scratch disk
- CCM compatibility mode available
- Intel, Cray, GNU compilers

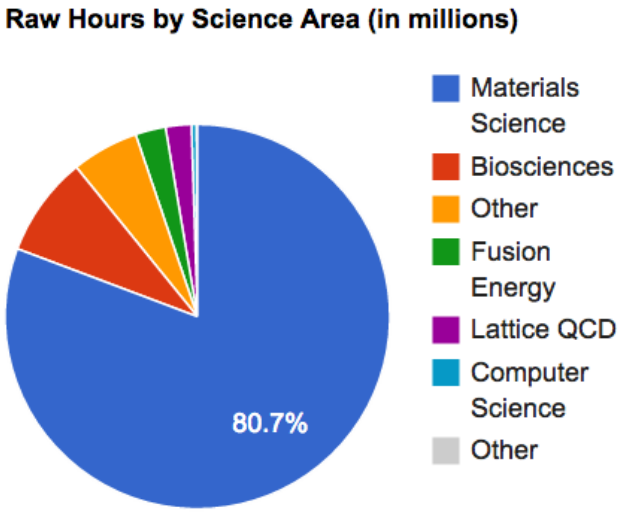
Edison Phase I is Here!

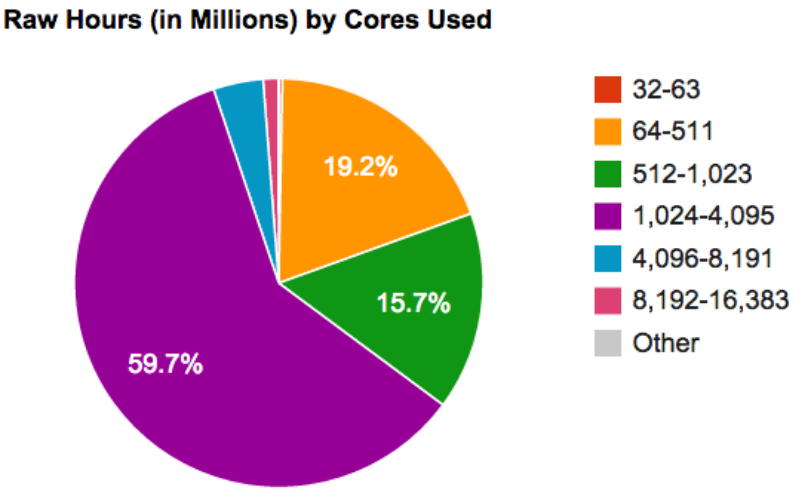
Quick Status

- **Edison Phase I is up and running**
- **Added first early users on Feb. 4**
  - We'll add you now if you're not on yet. Just ask.
- **System has been extremely stable**
  - Batch system software issues persist (not Cray software)
- **Early tests: about 2X Hopper performance per core**
- **Compute**
  - Meeting or exceeding expectations
- **I/O**
  - Not completely there yet (mostly reads and metadata)

Users Are Up and Running Quickly

- **2.5 million hours in a week**
  - 5 M Hopper-equivalent (?)
- **But Intel environment (the default) is still not complete**





Get on Edison and try it out!

Then give us your feedback.

Installation Timeline

- **Phase 0**
  - 6 login nodes ✓
  - 1 file system (35GB/s, 1.6PB) ✓
- **Phase 1 (Now)**
  - ~~4~~ 2 cabinets with ~10K Sandy Bridge processors ✓
  - Start Phase 1 acceptance test in February ✓
  - Start DARPA Mission Partner Access ✓
  - Add ~~2~~ 3 file systems (105GB/s, 4.8PB)
- **Phase 2 (early July)**
  - Add 24 cabinets with Ivy Bridge processors
  - Upgrade first 4 cabinets with Ivy Bridge
  - About 100K cores total
  - Will require downtime of a few weeks

Edison Scratch File System (March)

3 Scratch File Systems

2 - 1.6 PB @ 35 GB/sec

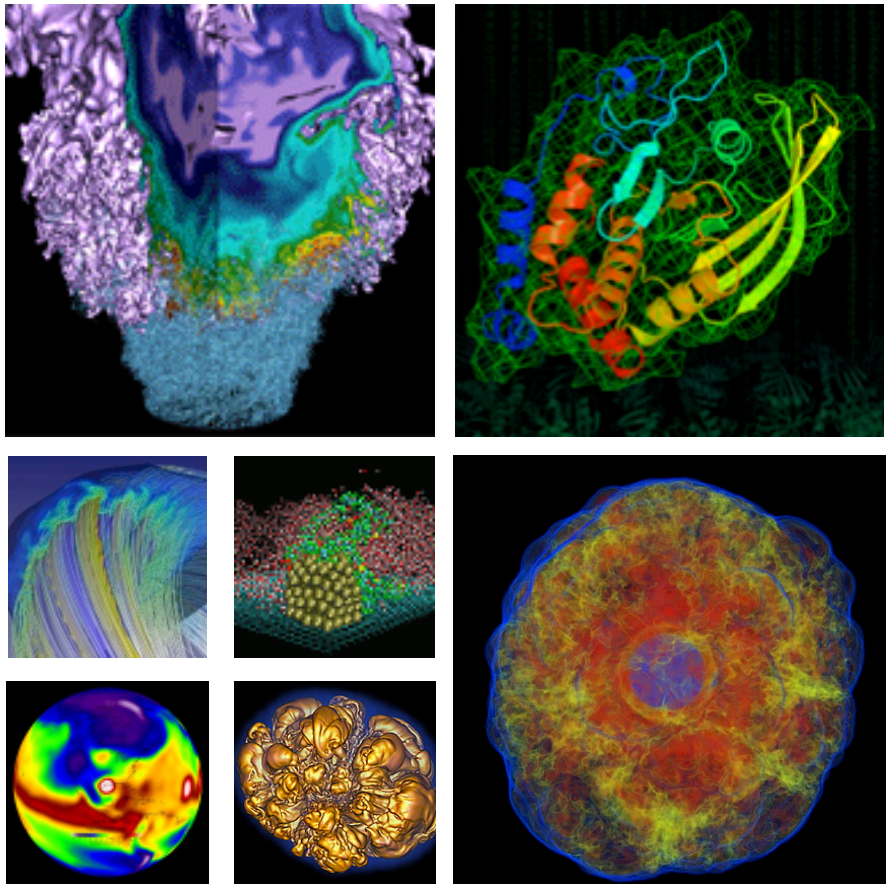
1 - 3.2 PB @ 70 GB/sec

- Multiple file systems to reduce contention, improve metadata performance
- One 2X Hopper's size & performance for I/O intensive applications
- Help us decide allocation policies

Next Presentations

- **Installation and Configuration Update**
  - Tina Declerck, NERSC 7 Systems Lead
- **Software and Programming Environment Update**
  - Zhengji Zhao, NERSC 7 Services Lead

Edison Systems Update



NERSC

Tina Declerck

12 February, 2013

U.S. DEPARTMENT OF ENERGY | Office of Science

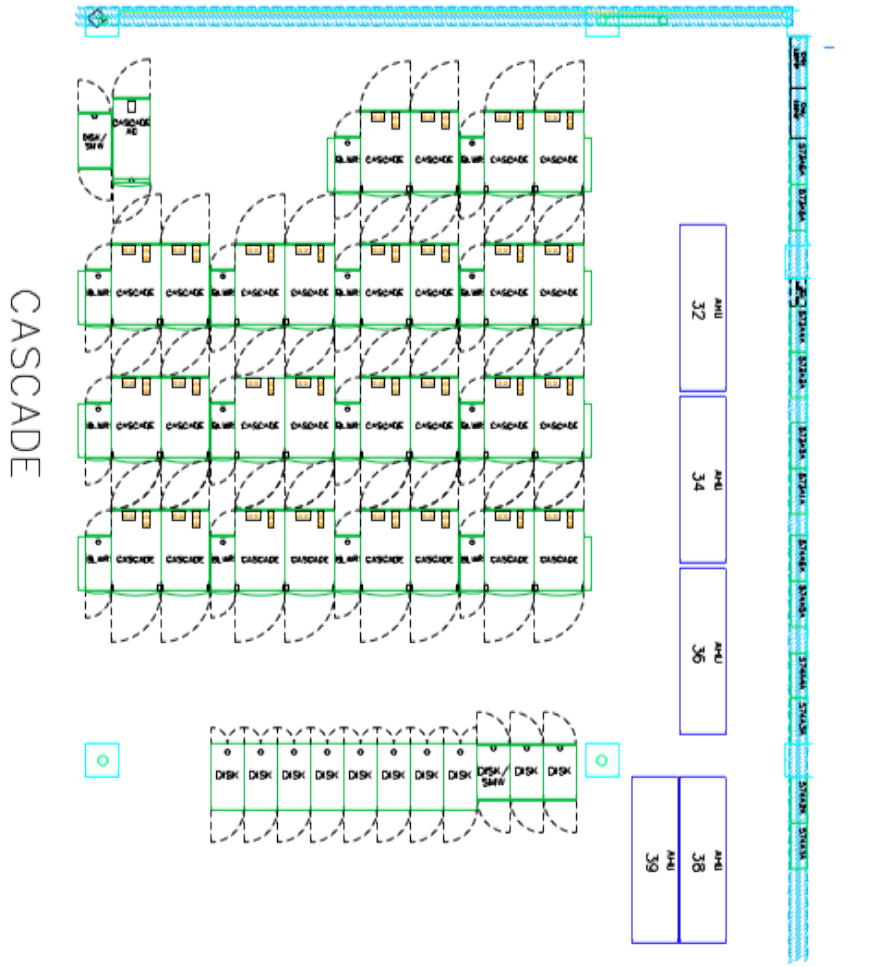
BERKELEY LAB
Lawrence Berkeley National Laboratory

Overview

- **Installation – prepping for Edison**
  - Power
  - Piping
  - IsoBases
- **Free Cooling**
- **Scheduling**
- **NERSC Prep**
- **Power monitoring**

Location, Location, Location

- **Where can the system go?**
  - Power considerations
  - Cooling – piping/ AHUs
  - System space requirements



Installation – Prepping for Edison

- **Prep for construction**
  - Constructing a wall
    - Between construction and the rest of the center
    - Purpose: Debris protection
    - Negative pressure

- **Equipment Removal**
  - 9 AHUs
  - Chillers
  - Power whips

Power / Cooling

- **Power work on power substations**
  - Outage on Hopper
- **Different electrical cables**
- **Drain existing pipes**
- **Pipe removal for rework**
- **Layout**
  - Requirements for hoses
  - Hose length measurements
- **Heat exchangers**

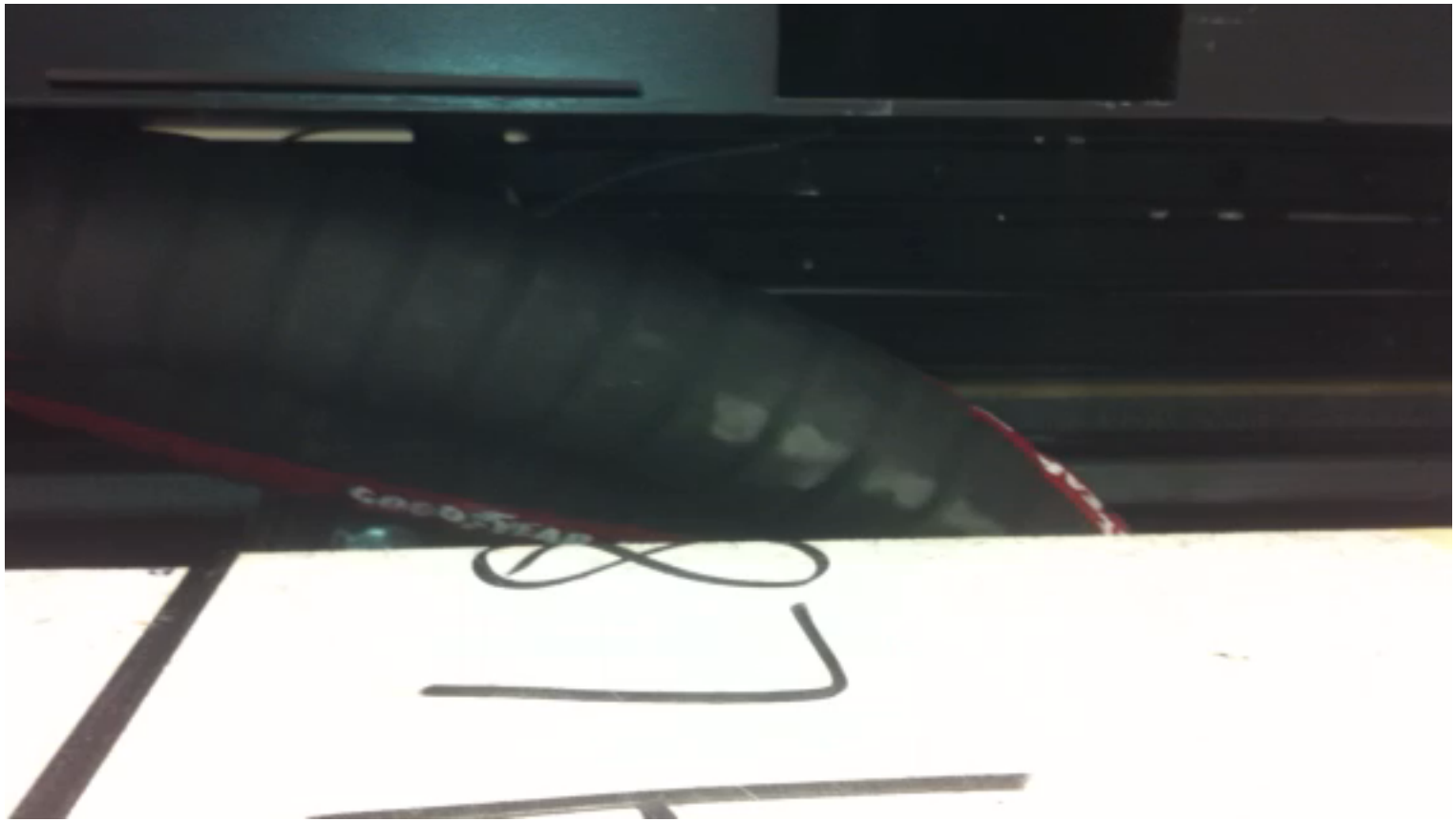
IsoBases

Can you make hoses & cables stretch?

- Description of how we designed the isobases and the power cable and water hose layout
- Great help from Cray

Power Cables

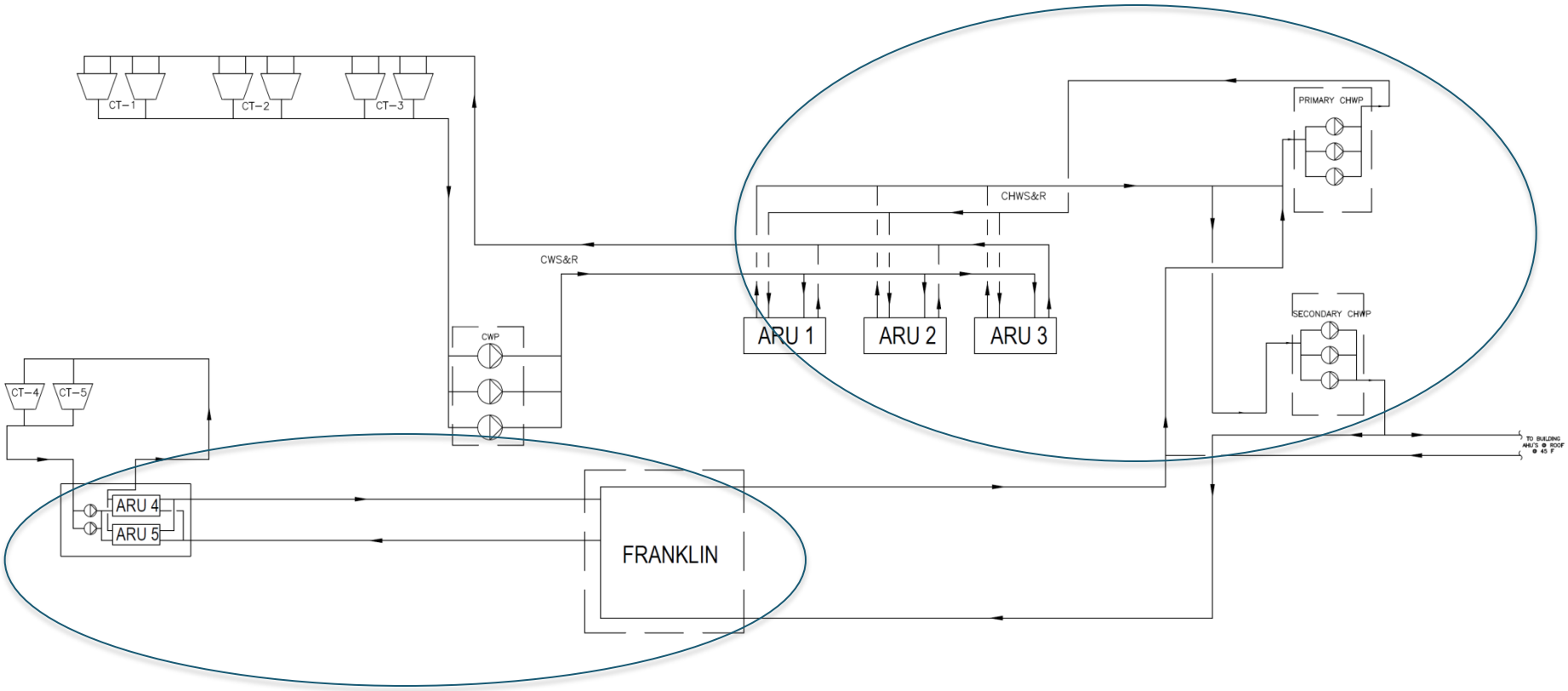
Hoses



Free Cooling

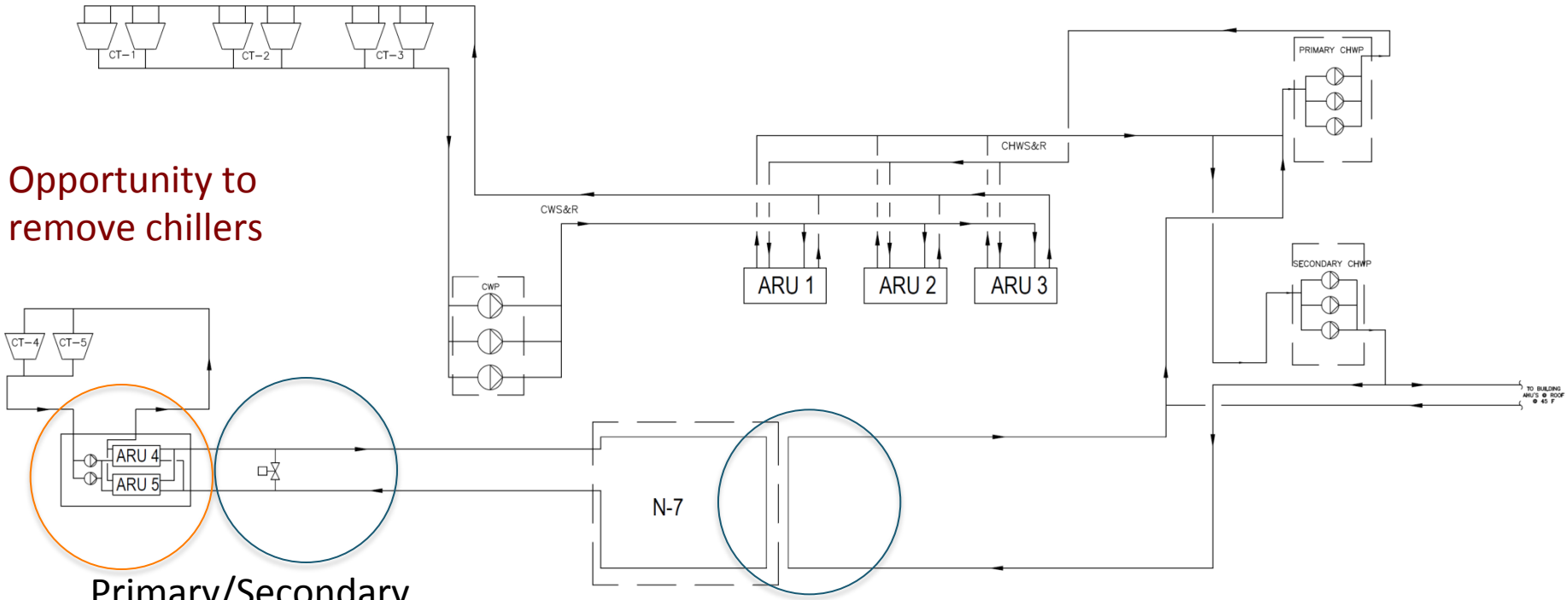
- **First DOE petaflop system with year-round free-cooling**
  - Order of magnitude improvement in energy efficiency (TF/W) over Franklin
  - PUE for N7 < 1.1
  - Exercise free-cooling capability before move to CRTF

OSF Mechanical One-line



Primary Injection Only = Δ-T 10F

N7 Mechanical Options

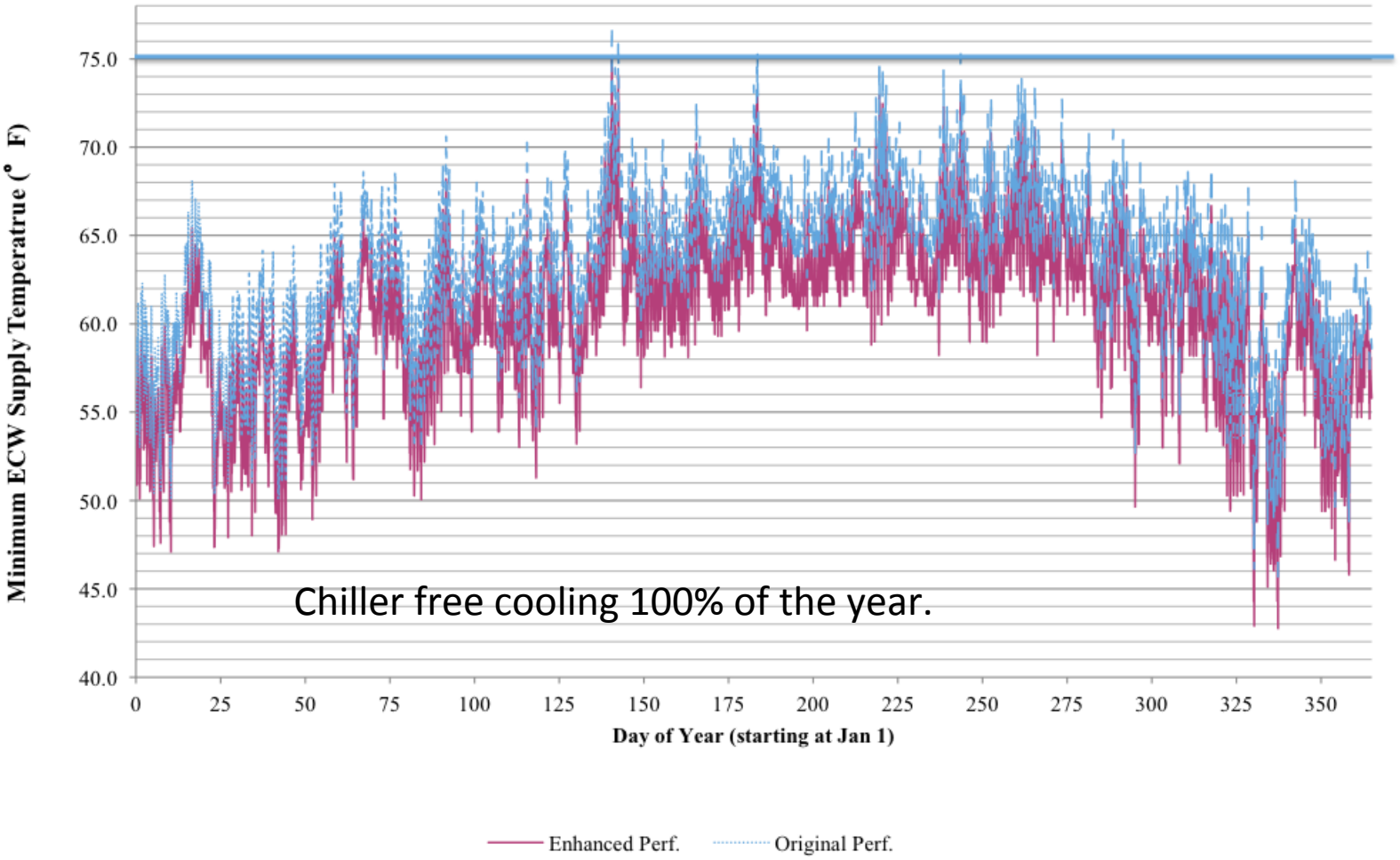


Opportunity to remove chillers

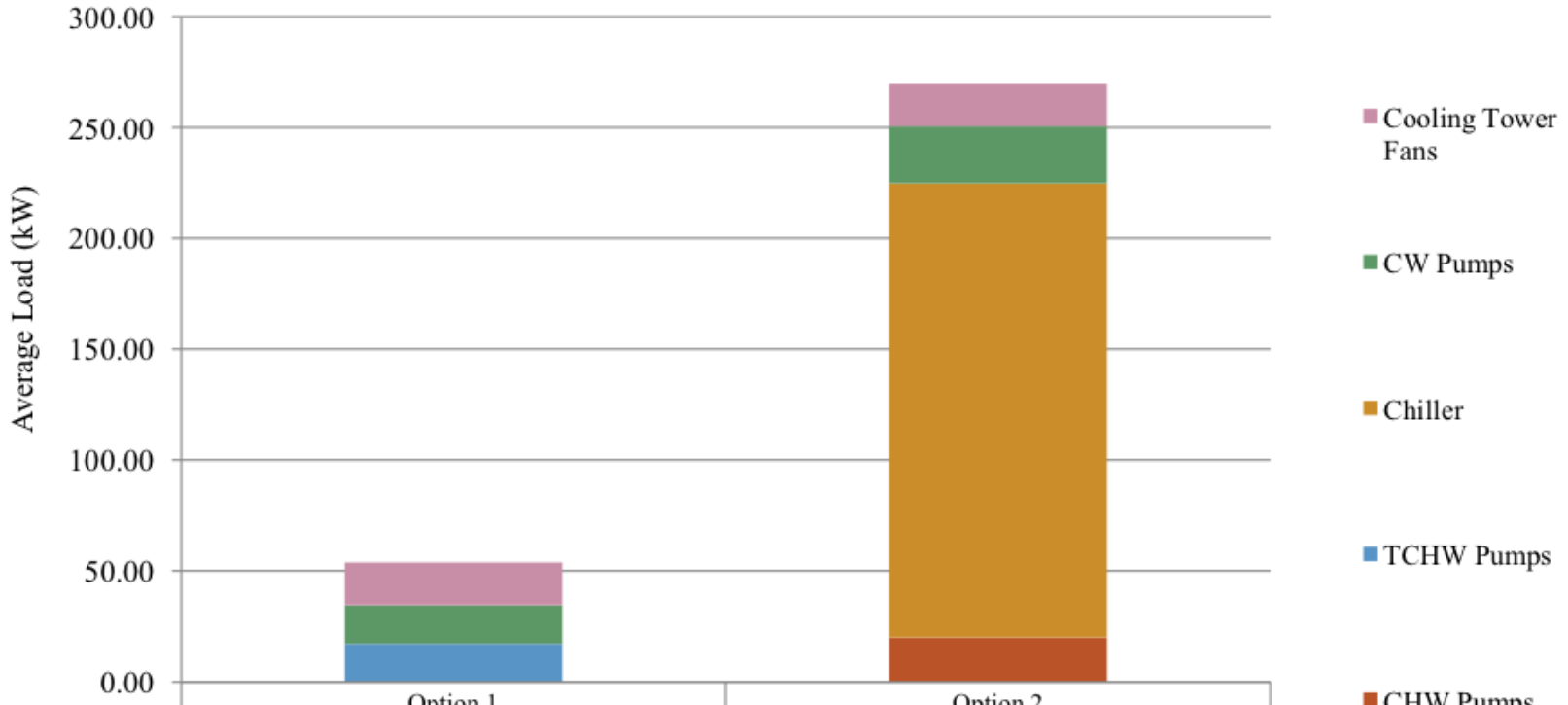
Primary/Secondary Pseudo loop

Separation valves

Oakland Climate Permits Use of Chiller-less Free Cooling

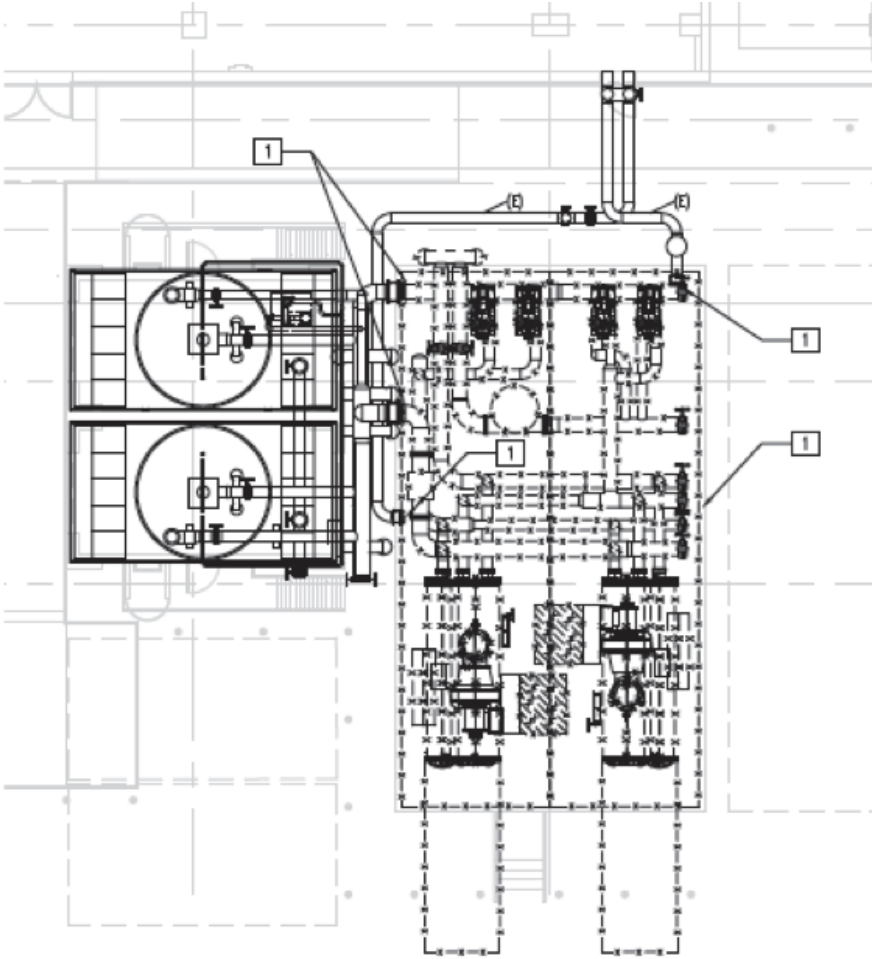


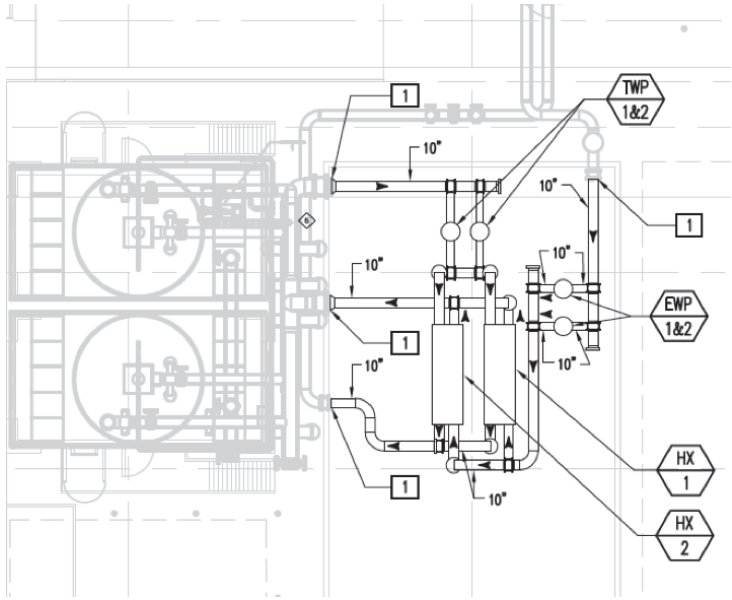
Power for Cooling with and without chillers



	Option 1	Option 2
Cooling Tower Fans	19.32	19.32
CW Pumps	17.38	25.71
Chiller	0.00	204.98
TCHW Pumps	17.17	0.00
CHW Pumps	0.00	19.97

Change to chiller free cooling





- Simplified mechanical plant
- Greater reliability
- Reduced maintenance cost

NERSC Site Prep

- **Network**
  - Fully redundant all the way to the network core
- **NGF**
  - Move from FC/IB to full IB
- **Monitoring**
  - Physical monitoring
  - System monitoring
- **NERSC Infrastructure**
  - NIM
  - LDAP
  - etc

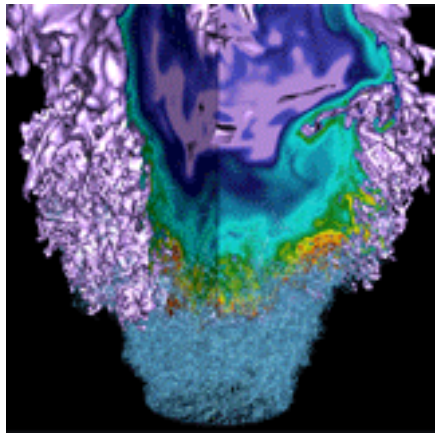
Scheduling

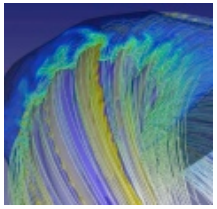
- **Same Apps as Hopper**
  - Torque / Moab
- **Using Fair Share features**
  - Ensure fair sharing for DARPA access
  - May require some experimentation
  - When jobs will run may differ from current configuration
  - Interested in user feedback

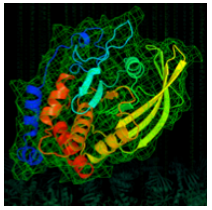
Power Monitoring

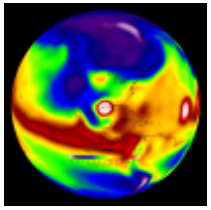
- **New Feature on Edison**
- **Currently at the rack level**
- **Allows power capping**
  - Only at the system level initially
- **Will be able to track at job level**
- **Users will be able to set**

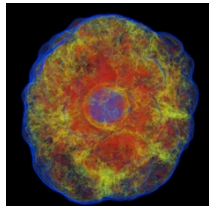
Software and Development Environment

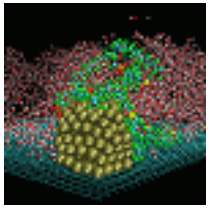












Application and Development Environment

- **Edison programming environment is remarkably similar to that of Hopper**
  - Easy to port and run production codes
- **Supports production software applications, libraries, and tools needed by the entire NERSC workload**
  - A robust set of programming languages, models
  - A rich set of highly optimized libraries, tools and applications
  - Community and pre-packaged applications
  - Shared-object libraries and socket communication
- **Enables effective application performance at scale, single node (high-throughput computing), and everything in between**

Supported Programming Languages, Programming models, and compilers

Supported Programming Languages	Supported Programming Models	Supported Compilers
Fortran	MPI	Intel*
C, C++	OpenMP	Cray
UPC	Cray Shmem	GNU
Python, Perl, Shells	POSIX Threads	
Java	POSIX Shared Memory	
Chapel	UPC	
	Coarray Fortran	
	Chapel	

***) Intel compiler is the default compiler on Edison.**
PGI compiler is not supported on Edison.

Cray Scientific and Math Libraries

Libraries	Intel	Cray	GNU
LibSci: LAPACK, ScaLAPACK, BLACS, PBLAS		✔	✔
Third Party Scientific Libraries: MUMPS, SuperLU, ParMETIS, HYPRE, Scotch		✔	✔
Trilinos		✔	✔
FFTW	✔	✔	✔
PETSc		✔	✔
DMAPP API for Aries	✔	✔	✔
MPI-IO Library	✔	✔	✔
IO Libraries: HDF5, NetCDF, Parallel-netcdf	✔	✔	✔
Intel MKL	✔	✔	✔

More libraries, Tools, and Applications

Scalable Debuggers	Performance Tools	Libraries and Tools	Applications	Visualization tools	ACTS Libraries and tools
DDT	Craypat/ Apprentice2	Python	VASP	VisIt	
TotalView	PAPI	GSL	QUANTUM ESPRESSO		
ATP	IPM	NCAR	LAMMPS		
		ADIOS	NAMD		
			BerkelyGW		
			MOLPRO		
			G09		

Shared libraries are supported on compute nodes

- Applications may be launched in a full Linux environment by choice
- Supports both dynamic and static libraries
- Accommodates shared library applications, which is one of our important workloads, to run on compute nodes, eg., python applications
- Cray is committed to improve the performance of the shared library applications at scale

Cluster Compatibility Mode is supported (CCM)

- CCM is a software solution to run applications which require the standard set of Linux Services, eg., ssh, nscd, DNS, rsh, and LDAP
- Applications run "out of the box" or can be compiled under CCM with third party MPI libraries, making use of the ISV Application Acceleration layer that translates InfiniBand verbs into Cray HSN API calls.
- It dynamically allocates and configures compute nodes at a job start, and releases nodes to the main computing pool at the job exit (ccm_queue).
- It accommodates some of our important TCP/IP workload, eg., G09, WIEN2k.

New features and technologies available on Edison Compared to Hopper

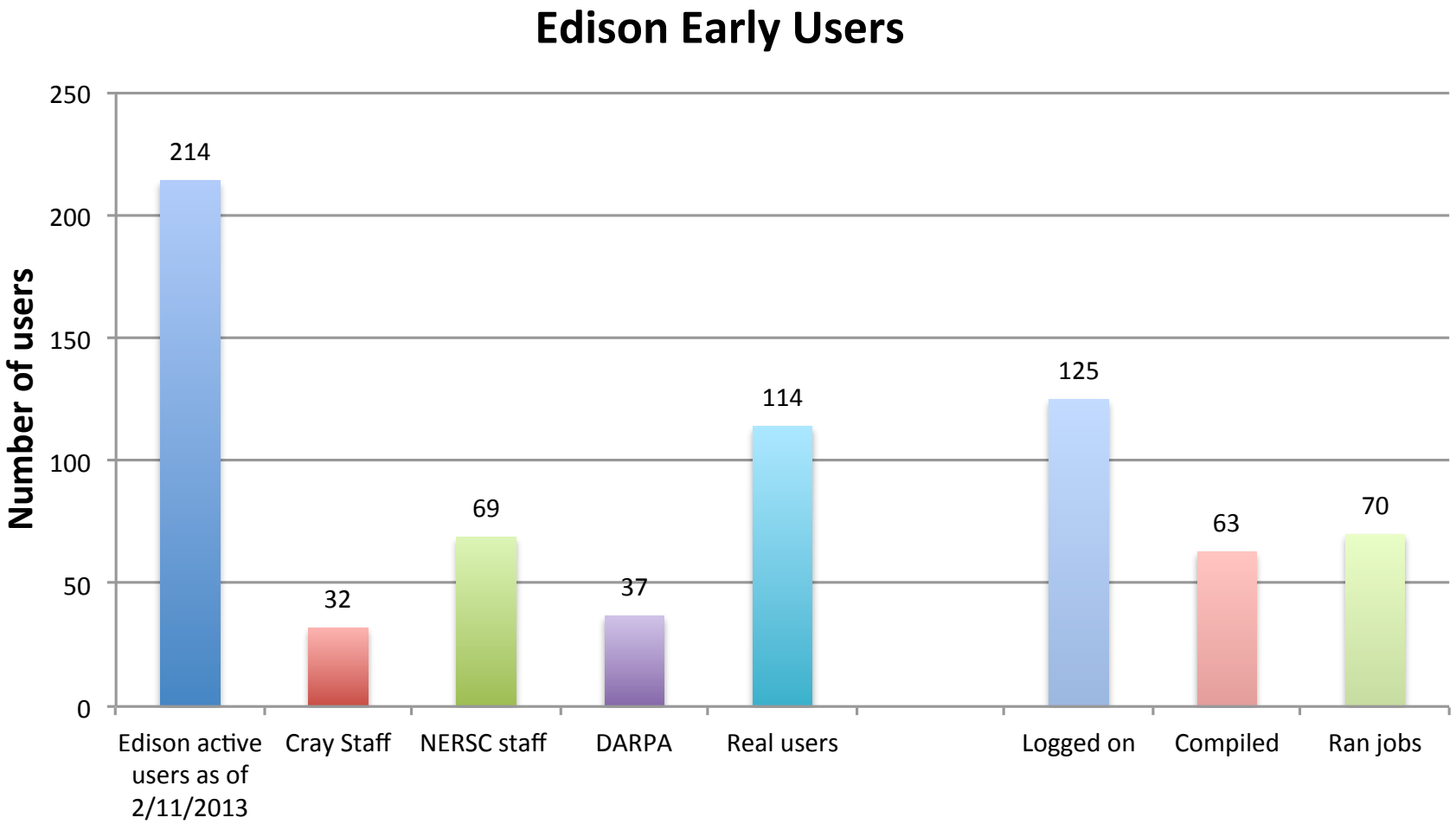
- Cray Aries interconnect
- Hyper-Threading Technology
- Cray Sonexion Storage System
- External Batch Server

Open Issues

- **Intel programming environment is still being configured.**
  - LibSci is not available and users need to use MKL to access LAPCK, SCALAPACK routines.
- **Batch queue issue (Moab)**
  - Some completed jobs stay in the queue long time

Please let us know if you find anything is missing in the Intel programming environment

Please become early users and help us test and configure Edison



We are looking forward to hearing from you!

National Energy Research Scientific Computing Center