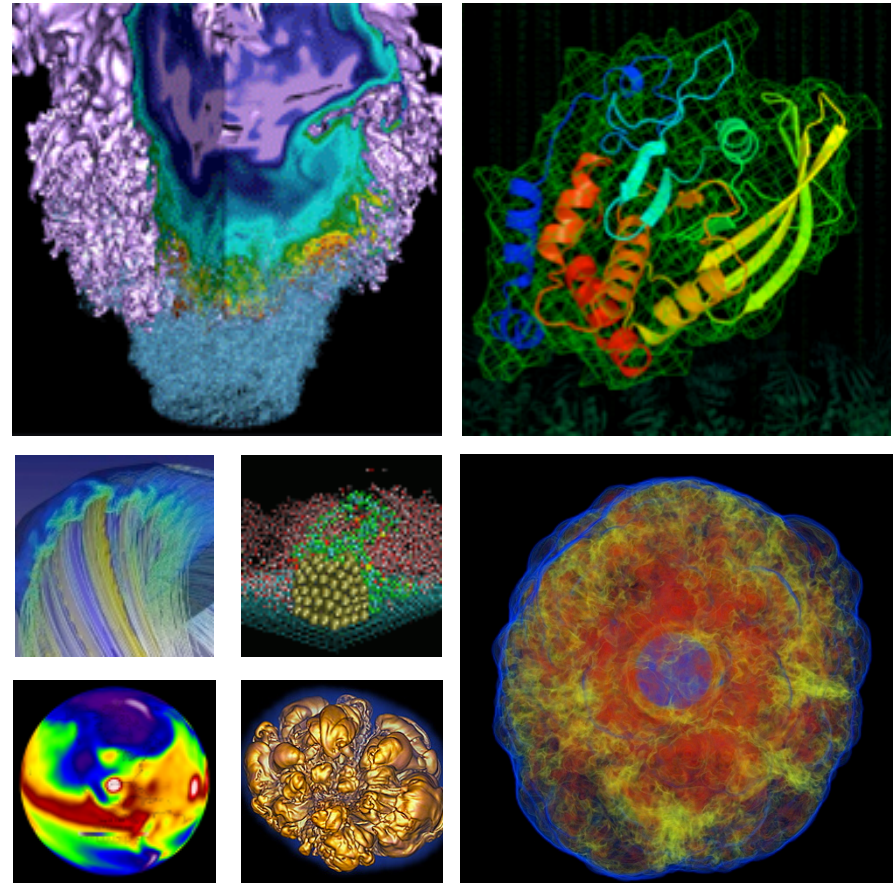


# NERSC Update



**Jeff Broughton**  
**NERSC Deputy for Operations**

February 18, 2014

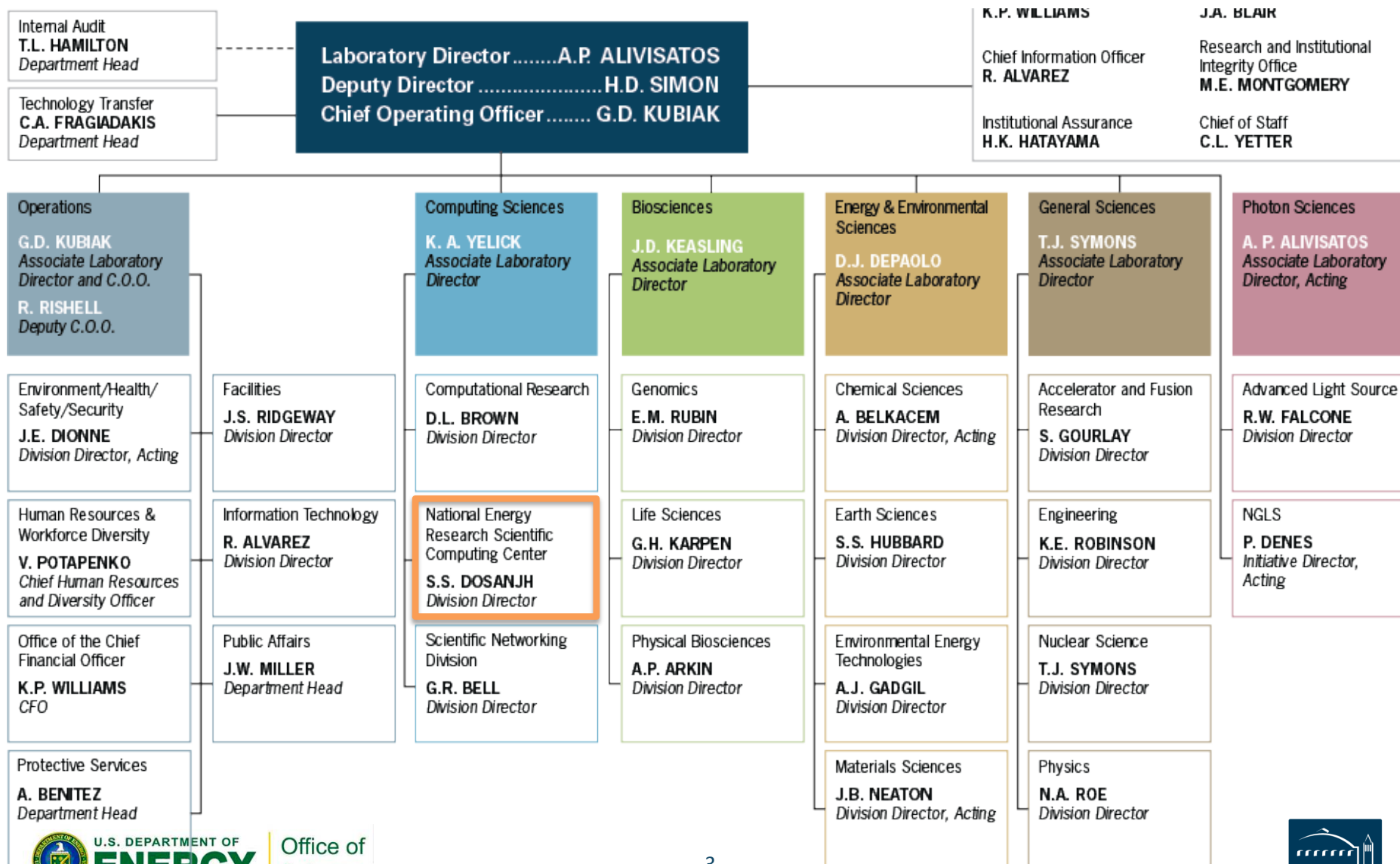
# Agenda

---



- **Organization & Changes**
- **Budget**
- **Edison**
- **CRT and Move-in Plans**
- **NERSC-8**

# Berkeley Lab Org Chart



# NERSC Management Team...



**Sudip Dosanjh**  
NERSC Director



**Francesca Verdier**  
Allocations Manager



**John Shalf**  
CTO

## Systems Department



**Jeff Broughton**  
Deputy for Operations



**Elizabeth Bautista**  
Operations Tech



**Shane Canon**  
Tech Integration



**Brent Draney**  
Network & Security



**Jason Hick**  
Storage Systems



**Jay Srinivasan**  
Computational Systems

## User Services Department



**Katie Antypas**  
User Services DH



**Richard Gerber**  
User Services



**Nick Wright**  
Advanced Tech



**Sheryas Cholia**  
Data & Analytics



**David Skinner**  
Strategic Partnerships



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science





# ... and the People Who Do the Work

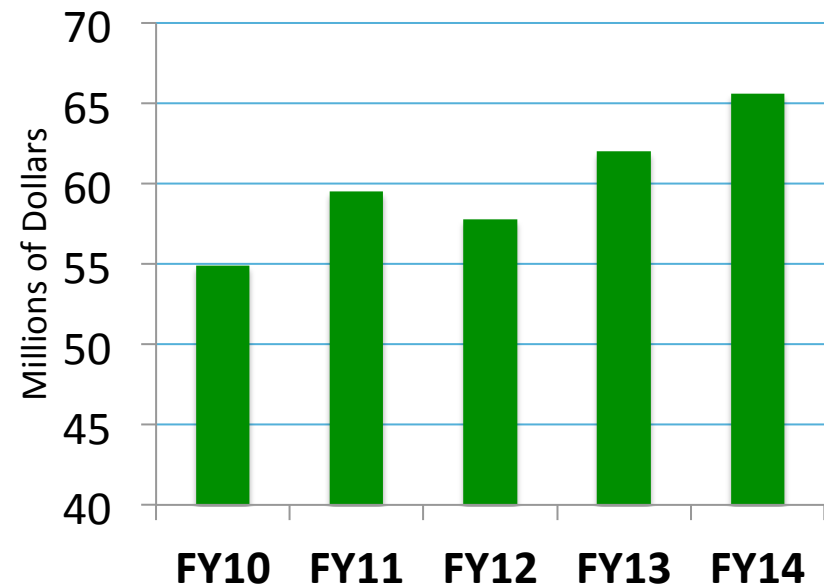


- **Systems Department (41)**
- **User Services (31)**
- **Division Resources (8)**

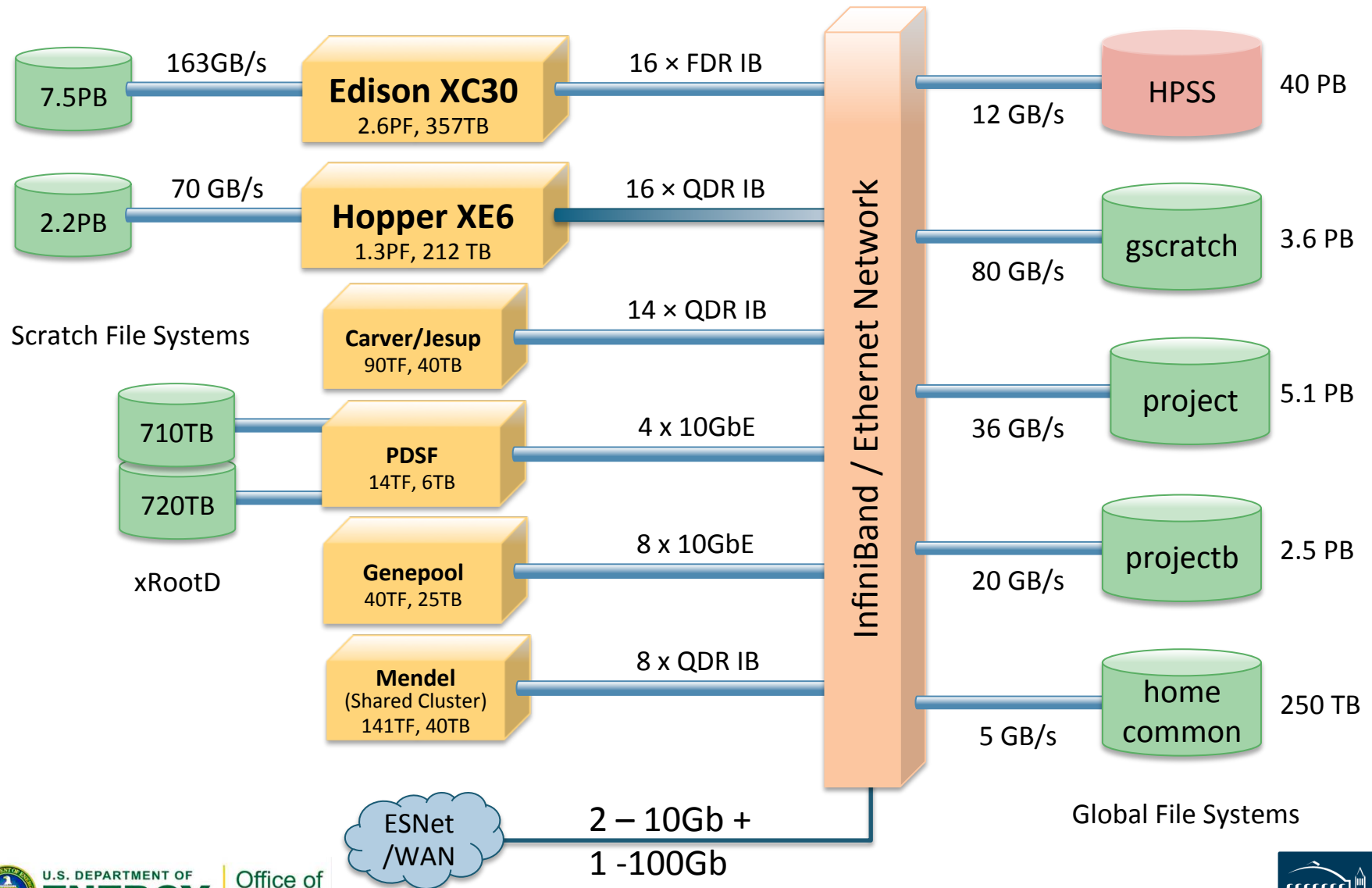
# Budget



- **NERSC has survived the budget wars in relatively good shape**
- **FY14 budget is \$65.6M**
- **Out year guidance is for increased funding**
- **Additional funding for sponsored systems**
  - JGI, Kbase, MatGen, PDSF
- **Additional funding for research**
  - LDRDs, DesignForward, BurstBuffer NRE



# NERSC Systems



# Edison at a Glance



- First Cray XC30
- Intel Ivy Bridge 12-core, 2.4GHz processors
- Aries interconnect with Dragonfly topology for great scalability
- Performs 2-4 x Hopper per node on real applications
- 12 x 512GB login nodes to support visualization and analytics
- 3 Lustre scratch file systems configured as 1:1:2 for capacity and bandwidth
- Access to NERSC's GPFS global file system via DVS
- Easy adoption for users because of similarity to Hopper
- Ambient cooled for extreme energy efficiency



# System Vital Statistics



|                                         | Hopper         | Edison (current) | Edison (late Feb) |
|-----------------------------------------|----------------|------------------|-------------------|
| Cabinets                                | 68             | 28               | 30                |
| Compute Nodes                           | 6,384          | 5,200            | 5,576             |
| CPU Cores ( <i>Total / Per-node</i> )   | 152,408 / 24   | 124,800 / 24     | 133,824 / 24      |
| CPU Frequency (GHz)                     | 2.1            | 2.4              | 2.4               |
| Peak Flops (PF)                         | 1.29           | 2.4              | 2.6               |
| Memory (TB) ( <i>Total / Per-node</i> ) | 217 / 32       | 333 / 64         | 357 / 64          |
| Memory (Stream) BW* (TB/s)              | 331            | 462.8            | 496.3             |
| Memory BW/node* (GB/s)                  | 52             | 89               | 89                |
| File system(s)                          | 2 PB @ 70 GB/s | 7.4 PB @ 163GB/s | 7.4 PB @ 163 GB/s |
| Peak Bisection BW (TB/s)                | 5.1            | 11.0             | 12.6              |
| Power (MW Linpack)                      | 2.9            | 1.9              | 2.0               |

# NERSC is currently located at the Oakland Scientific Facility (OSF)



- 11 MW power (6+5MW)
- 19,000 ft<sup>2</sup>, 250 lbs/ft<sup>2</sup>
- Chilled Air + Water
- PUE ~ 1.3
- Extensive environmental and energy-usage monitoring
- ESnet @ 10Gb/s & 100Gb/s
- Control room staffed 24 × 365
- 6.5 miles from LBNL main campus

# The Computational Research and Theory (CRT) building will be the home for NERSC in 2015



- **Four story, 140,000 GSF**
  - 300 offices on two floors
  - 20K -> 29Ksf HPC floor
  - 12.5MW -> 42 MW to building
- **Located for collaboration**
  - CRD and ESnet
  - UC Berkeley
- **Unique features**
  - Ambient air and water cooling only
  - Seismic isolation for systems
- **Initial occupancy late 2014**



# Ambient (free) cooling provides exceptional energy efficiency

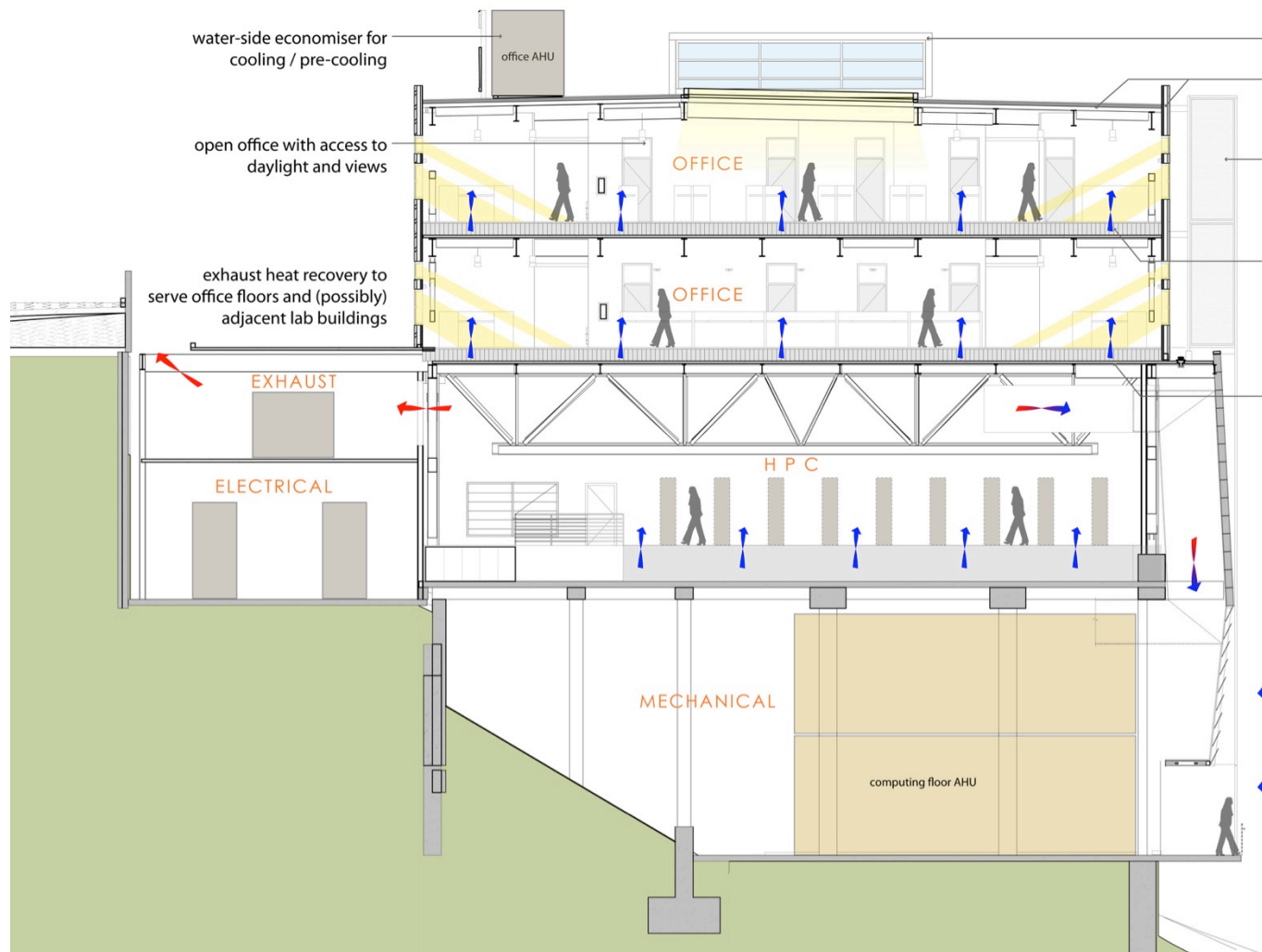
---



- **LBNL's location and the CRT design enables**
  - Power Usage Effectiveness (PUE): 1.1
  - Data Center infrastructure Efficiency (DCiE): 0.91
- **Air cooling**
  - 75°F air year round without chillers
- **Liquid cooling**
  - 74°F water year round without chillers
- **Computer room exhaust heat used to heat office floors**
- **Save ~50% per year on power costs**
  - Free cooling + WAPA power



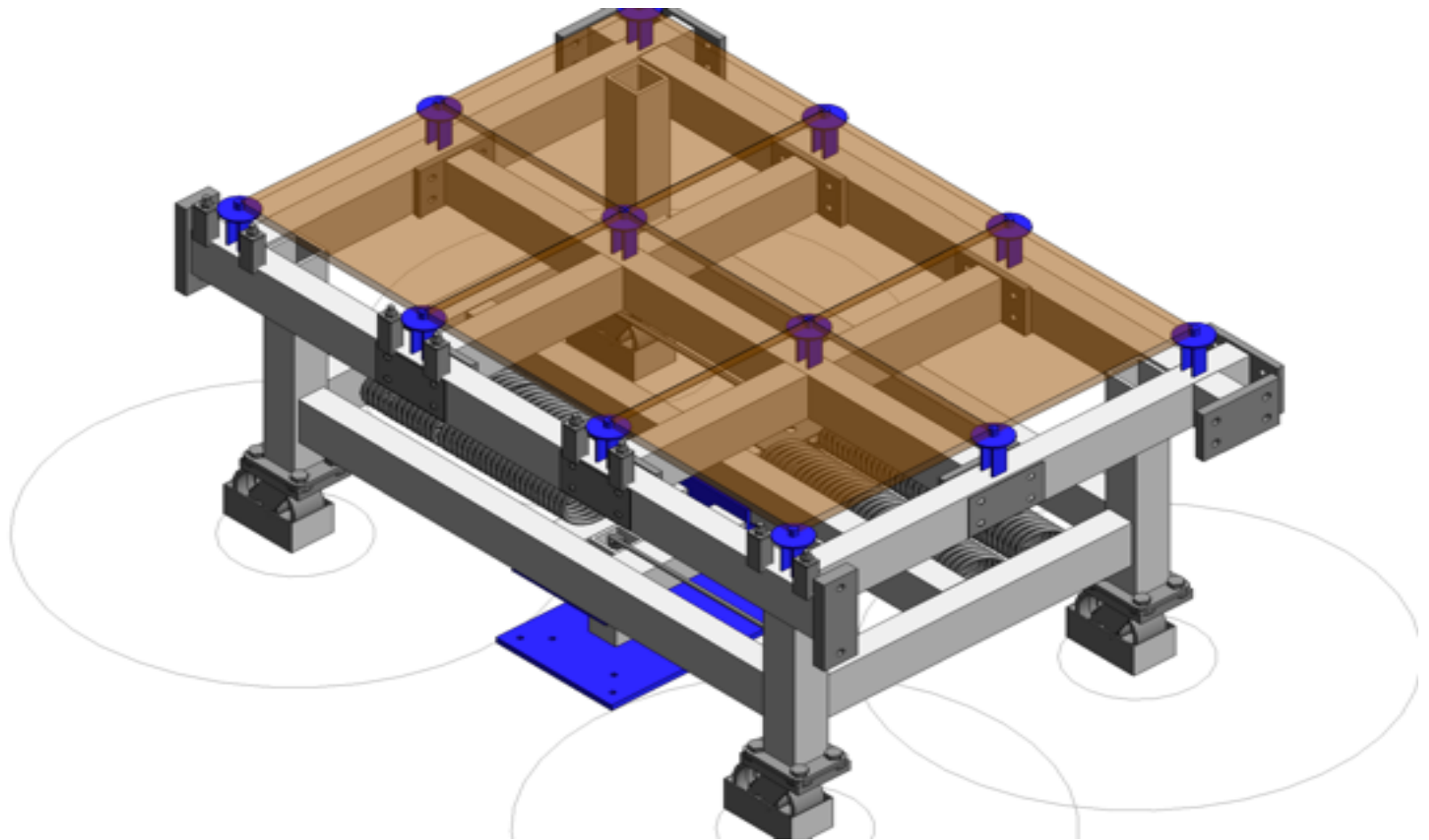
# Building Cross Section



# Machine Room



# HPC Floor Module



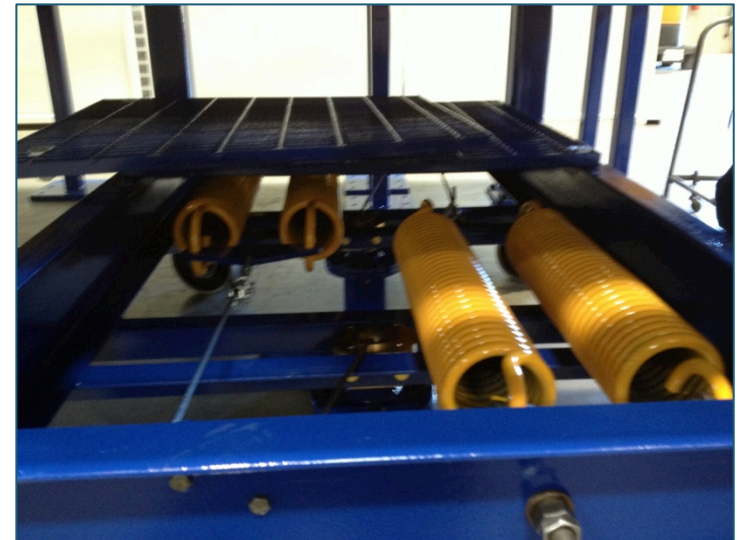
U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science





# Seismic Isolation Floor





# Guiding Principles for the Move



| Goals                                            | Strategies                                                                                                                                                                                                                                                                      |
|--------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Maximize system availability throughout the move | <ul style="list-style-type: none"><li>• Keep Hopper in operation at OSF throughout the move</li><li>• Phased move to avoid extensive downtime</li></ul>                                                                                                                         |
| Preserve users' data ("no bits left behind")     | <ul style="list-style-type: none"><li>• Replicate data (locally/across net) in case of damage in transit</li><li>• Full backups in HPSS</li></ul>                                                                                                                               |
| Minimize number and duration of outages          | <ul style="list-style-type: none"><li>• Network outages (3 outages, &lt; 1 hour total)</li><li>• NGF outages while storage is moved (no outages)</li><li>• HPSS outages while libraries are moved (~2-3 weeks)</li><li>• Plus, individual system outages (~2-6 weeks)</li></ul> |
| Minimize costs pragmatically                     | <ul style="list-style-type: none"><li>• Use planned equipment refresh to pre-stage equipment</li><li>• Duplicate other equipment only when necessary</li><li>• Phased move will reduce need for extra staff</li></ul>                                                           |

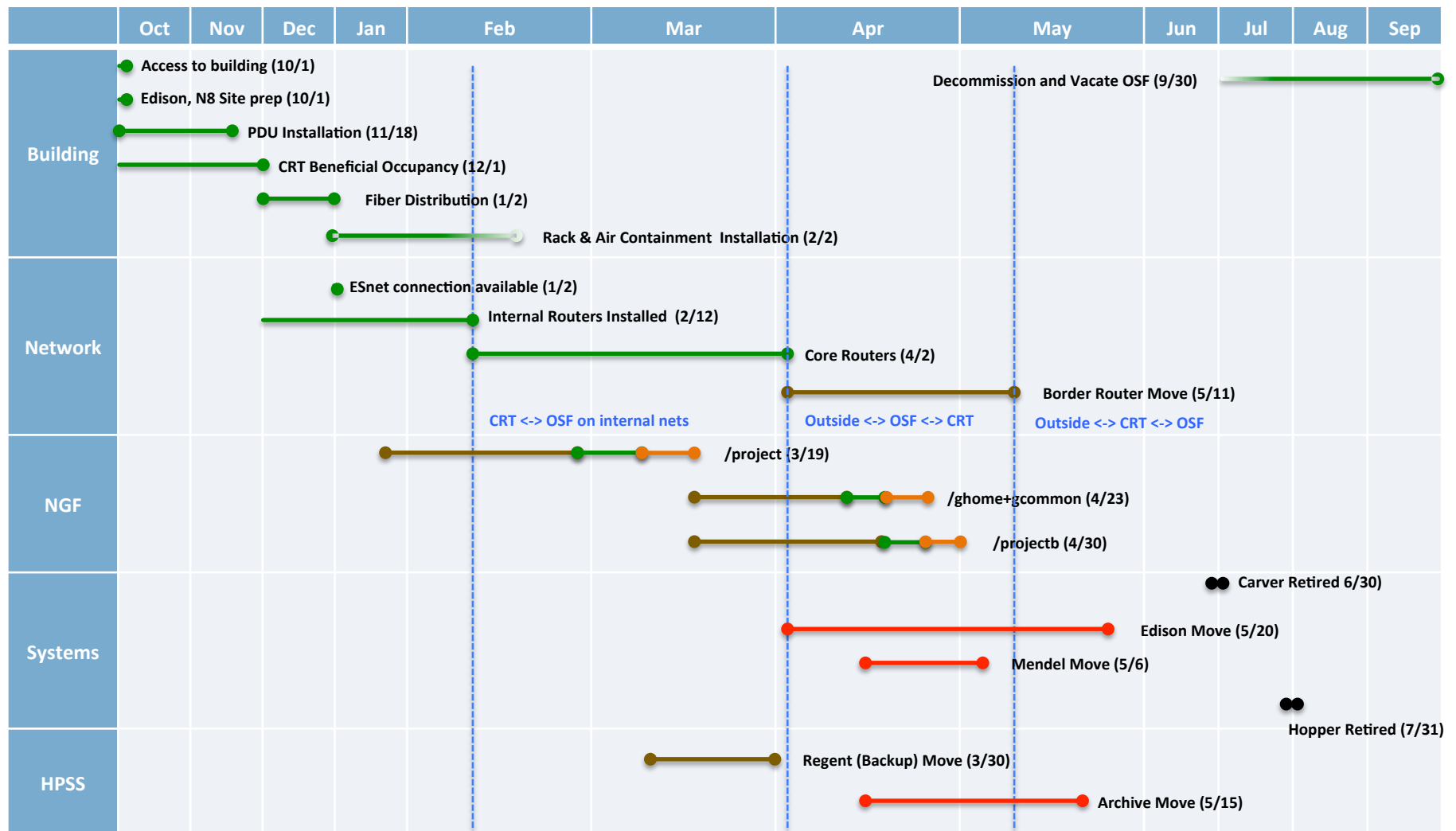
# Very High Level Plan

---



- **Install power distribution and cooling manifolds**
- **Install networking and IB SAN infrastructure**
- **Move file systems (NGF)**
- **Move Edison and sponsored systems (e.g. PDSF, JGI)**
- **Move tape archives (HPSS)**
- **Retire Hopper and Carver at OSF in summer 2015**
- **Vacate OSF by 9/30/2015**
- **Install NERSC-8 in CRT in 2016**

# Startup/Move Schedule





# Bridging the Gap



- **NERSC delivered hours will dip in AY15/16**
  - Edison outage over move
  - Hopper & Carver retirement
  - Timing of NERSC-8 delivery and preproduction time
- **Mitigation**
  - Seeking funding for an interim replacement system
  - Could be standalone or linked to Edison or NERSC-8

# NERSC-8 (Coming 2016)

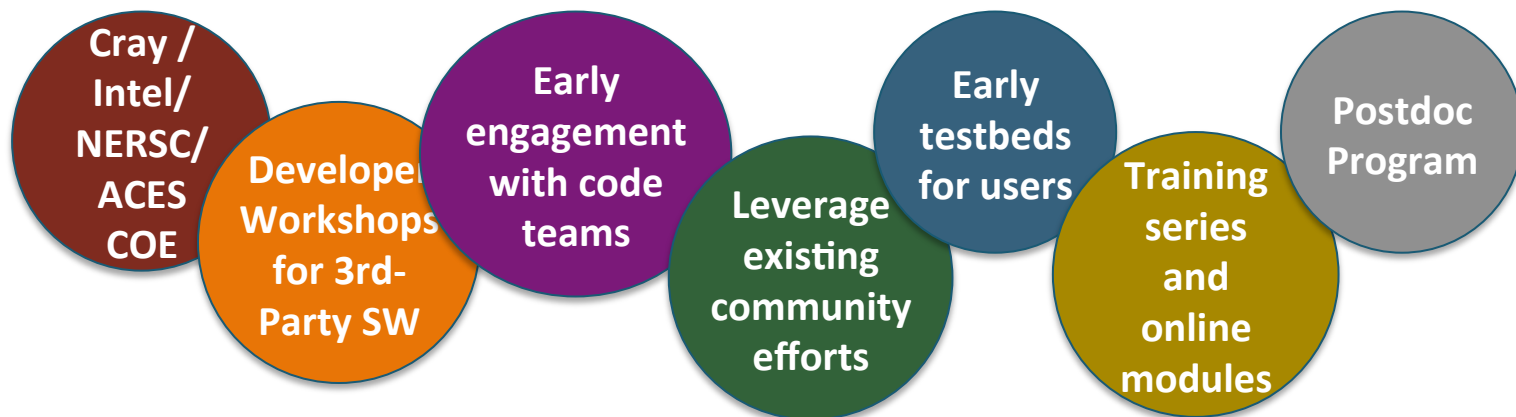


- **Provide ~10 x Hopper performance**
  - Support for legacy code, albeit at less than optimal performance
  - Reasonable performance with MPI+OpenMP
- **Begin the transition to more energy-efficient architectures.**
  - Increased cores per node
  - Increased threads per core
  - Increased vector length
  - Programmer-visible, fast local memories
  - Programmer-visible power-management
- **Exploring option for a “Burst buffer”**
  - Layer of NVRAM between memory and disk to to accelerate application I/O

# NERSC Application Readiness Effort



- NERSC has a robust application transition plan in place which will help transition users to new architectures



*NERSC is committed to helping our users make this transition*



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

