Computing Sciences at Berkeley Lab
CS Student Program Welcome  2 June 2020

David Brown, Director
Computational Research Division
Lawrence Berkeley National Laboratory
Why the nation needs national laboratories

- Discovery science
- Scientific solutions addressing national challenges, especially energy
- Unique scientific capabilities
  - User facilities
- Managed, large research teams
- Important technologies with long, risky R&D paths
- A diverse group of highly trained, creative, and committed scientists and engineers.
Berkeley Lab is one of the 17 U.S. Department of Energy (DOE) National Laboratories

The mission of the Energy Department is to ensure America’s security and prosperity by addressing its energy, environmental and nuclear challenges through transformative science and technology solutions.
Berkeley Lab Changes Science

Radiation Lab staff on the magnet yoke for the 60-inch cyclotron, 1939, including:
E. O. Lawrence
Edwin McMillan
Glenn Seaborg
Luis Alvarez
J. Robert Oppenheimer
Robert R. Wilson

Lawrence introduces big team science 1931  LBNL the first National Lab

Today, Berkeley Lab has:
Over 4000 employees
$1.1B in FY18 funding
13 associated Nobel prizes
Berkeley Lab brings Science Solutions to the World

- Discovered 16 elements
- Unmasked a dinosaur killer
- Identified good and bad cholesterol
- Fabricated the smallest machines
- Turned windows into energy savers
- Confirmed the Big Bang and discovered dark energy
- Explained Photosynthesis
- Revealed the secrets of the human genome

https://www.lbl.gov/program/35-breakthroughs/
Computing Sciences at Berkeley Lab in 2020

NERSC

Computational Research

Scientific Networking: ESnet

Applied Mathematics

Data Science & Technology

Computer Science

Computational Science

Jonathan Carter

Sudip Dosanjh

Inder Monga

David Brown

Inder Monga

Jonathan Carter

Sudip Dosanjh

David Brown
### NOW COMPUTING

A small sample of massively parallel scientific computing jobs running right now at NERSC.

<table>
<thead>
<tr>
<th>PROJECT</th>
<th>MACHINE</th>
<th>NODES</th>
<th>NERSC HOURS USED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guest-host interactions in the gas phase, in aqueous systems and hydrate lattices: Implications for H2 storage and CO2 sequestration (PNNL)</td>
<td>Cori</td>
<td>1,024</td>
<td>636,317.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Event</th>
<th>Presenter</th>
<th>Zoom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun 11</td>
<td>11 AM - 12 PM</td>
<td>NERSC: Scientific Discovery through Computation</td>
<td>Rebecca Hartman-Baker (NERSC)</td>
<td></td>
</tr>
<tr>
<td>Jun 11</td>
<td>1 PM - 3 PM</td>
<td>Introduction to NERSC Resources</td>
<td>Helen He (NERSC)</td>
<td></td>
</tr>
<tr>
<td>Jun 17</td>
<td>10 AM - 12</td>
<td>Crash Course in Supercomputing</td>
<td>Rebecca Hartman-Baker (NERSC)</td>
<td></td>
</tr>
</tbody>
</table>
NERSC’s newest machine Cori supports both the HPC Workload and Data-Intensive Science

- **Cray system with 9,300 Intel Knights Landing compute nodes**
  - Self-hosted, (not an accelerator) manycore processor > 64 cores per node
  - On-package high-bandwidth memory at >400GB/sec

- **Data Intensive Science Support**
  - 10 Haswell processor cabinets (Phase 1) to support data intensive applications
  - NVRAM Burst Buffer with 1.5PB of disk and 1.5TB/sec
  - 28 PB of disk, >700 GB/sec I/O bandwidth in Lustre bandwidth
ESnet is a Unique Instrument for Science

- Connects 40 DOE sites to 140 other networks
- Growing twice as fast as commercial
- 50% of traffic is from "big data"
- First 100G continental scale network
- ANI dark fiber can be leveraged to develop and deliver 1 terabit

Unique capabilities

ESnet designed for large data

**From Simulations to Real-World: Building Deep Reinforcement Learning for Networks**

Jul 23 11 AM - 12 PM

Mariam Kiran (Scientific Data Management Group)
What are the questions driving research in computing?

Can we continue the growth in computing performance through more efficient architectures or new paradigms?

What mathematical models, algorithms and software are needed for increasingly complex scientific theories and experimental data sets?

Can we enable new modes of scientific discovery by applying advanced computing and networking to data from science experiments?
What are the questions driving research in computing?

Can we continue the growth in computing performance through more efficient architectures or new paradigms?

What mathematical models, algorithms and software are needed for increasingly complex scientific theories and experimental data sets?

Can we enable new modes of scientific discovery by applying advanced computing and networking to data from science experiments?
Can new logic devices give us beyond-Moore performance?

Multiple quantum device technologies at Berkeley

Experimentally implement chemical simulations on a 3-qubit platform

Develop model for small user facility to explore device technology

Siddiqi's Quantum Circuit

Quantum Chemistry

Fermi Hubbard

Quantum Ising, Bose-Hubbard, Spin-Boson

Synthetic gauge fields, Relativistic theories

Investigating alternative devices

Use Skirmion "bags" to act as information carriers for multi-valued logic devices

Investigating energy efficient superconducting architectures where information is stored in magnetic flux quanta and transferred with Single Flux Quantum voltage pulses

THE HAMILTONIAN LANDSCAPE FOR QUANTUM SIMULATION
What are the questions driving research in computing?

Can we continue the growth in computing performance through more efficient architectures or new paradigms?

What mathematical models, algorithms and software are needed for increasingly complex scientific theories and experimental data sets?

Can we enable new modes of scientific discovery by applying advanced computing and networking to data from science experiments?
Transforming how we compute:
“smart” math, numerics, HPC: gives unprecedented capability

- Smart math leads to science not possible before
  - Use mathematical properties to build better simulation models
  - Simulation at previously inaccessible scales
  - Exploit matrix structure for faster linear algebra

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Speaker</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun 04</td>
<td>Modeling Antarctic ice with Adaptive Mesh Refinement</td>
<td>Dan Martin (Applied Numerical Algorithms Group)</td>
<td>zoom</td>
</tr>
<tr>
<td>Jun 25</td>
<td>Surrogate Optimization for HPC Applications</td>
<td>Juliane Mueller (Center for Computational Sciences and Engineering)</td>
<td>zoom</td>
</tr>
<tr>
<td>Jul 09</td>
<td>Simulating Supernovae with Supercomputers</td>
<td>Donald Wilcox (Center for Computational Sciences and Engineering)</td>
<td>zoom</td>
</tr>
</tbody>
</table>

Over 100x increase in throughput

MAESTRO simulation near ignition showing flow from center of star and region of high energy generation
What are the questions driving research in computing?

Can we continue the growth in computing performance through more efficient architectures or new paradigms?

What mathematical models, algorithms and software are needed for increasingly complex scientific theories and experimental data sets?

Can we enable new modes of scientific discovery by applying advanced computing and networking to data from science experiments?
The Advanced Light Source (ALS) hosts dozens of different experiments and end station detectors.
Data-driven scientific discovery requires integration of modeling, simulation, analysis, data management.

- **Example: 21st Century Cosmology:**
  - Tight collaboration between astrophysicists and computational scientists to develop new technologies for cosmological data
  - Analysis & simulation of 100s of TeraBytes of data from ground- and space-based observations
  - Modeling & simulation of supernovae and large-scale structure formation
  - Large Scale Structure simulations of cluster formation in the early universe
  - Hydro simulation of a flame front in a thermonuclear supernova explosion

<table>
<thead>
<tr>
<th>Jun 18</th>
<th>High Performance Computing For Cosmic Microwave Background Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 AM - 12 PM</td>
<td>Julian Borrill (Computational Cosmology Center)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jun 30</th>
<th>Efficient Scientific Data Management on Supercomputers</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 AM - 12 PM</td>
<td>Suren Byna (Scientific Data Management Group)</td>
</tr>
</tbody>
</table>

Palomar Transient Factory data-analysis sky-coverage map for the first 3 years of the project

Cosmic Microwave Background Radiation data from Planck
Machine Learning enables new scientific discoveries from massive data sets

- Projects that advance the state-of-art in machine learning with ties to science

- Automated detection and analysis of particle beams in laser-plasma accelerator simulations

- Figure 5. Comparison of particle selection with/without MVEE: extracting the orientation and the axes of an enclosing ellipse from (a) produces (b), increasing the number of particles from 173 to 263. Colors indicate the density of particles, using only $(x, y)$-coordinates, and black dots show potential particles to belong to the beam, according to the different methods.

- Maximum (beam candidate region) per time step. In addition, this is a way of accruing more samples and detecting secondary beams when these are almost as prominent as the primary beam, associated to the maximum of $f$.

- During the searching for values that are approximately equal to $\max(f)$, we keep not only the maximum, but all bins where $f \geq u \ast \max(f)$, where $u$ is an uncertainty or tolerance parameter, here empirically set to 0.85. While this value enables the detection of the main and the secondary beams (when present), lower values of $u$ could be used to control the amount of particles to be selected at a lower accuracy of beam position. From this point, we refer to the subset of particles conditioned to $u \ast \max(f)$ and its adjacency, calculated for each time step, as "beam candidates".

- Figure 4 (top) presents projections of Figure 3.b with their calculated beam candidates emphasized in red. These are the result of our first attempt to improve particle selection by using an algorithm known as minimum volume enclosing ellipsoid as in Khachiyan & Todd (1993), which is able to enclose previously selected particles and to include others based on a geometrically defined polytope. Figure 5 illustrates the algorithm when applied to LWFA data, showing the selected particles as black dots; these particles are not in the most dense region (red) once the colors refers to $(x, y)$-density calculation. When including compactness in $\text{px}$, the most dense region happens further ahead. As distinct from calculating center of mass and forcing an ad hoc diameter or semi-major/minor axes, the minimum volume enclosing ellipsoid (MVEE) algorithm [Khachiyan & Todd (1993); Kumar & Yildirim (2005); Moshtagh (2009)] takes the subset of points and prescribes a polytope model to extrapolate a preliminary sub-selection to other particles likely to be in the bunch. The MVEE algorithm is a semidefinite programming problem and consists of a better approximation to the convexity of subsets of...
Welcome to Berkeley Lab Computing Sciences!