A Superfacility for Data Intensive Science

Kathy Yelick

Associate Laboratory Director for Computing Sciences
Lawrence Berkeley National Laboratory
Professor of Electrical Engineering and Computer Sciences
University of California at Berkeley
Science is poised for transformation
Old School Scientists: The Lone Scientist
Team Science

Lawrence introduces
big team science 1931

LBNL the first
National Lab
New Scientists

17-year-old Brittany Wegner creates breast cancer detection tool that is 99% accurate on a minimally invasive, previously inaccurate test.

Machine Learning + Online Data + Cloud Computing
Experimental Science is Changing

JAX® MICE & SERVICES

JAX® Mice are the highest quality and most-published mouse models in the world. Take advantage of our large inventories of common inbred strains and the convenience of having your breeding and drug efficacy needs met by the leading experts in mouse modeling.
Old School Scientific Workflow
Computing, experiments, networking and expertise in a “Superfacility” for Science

Slot-die printing of Organic photovoltaics

Old School HPC: only for Simulation

Experimentation

Data Analysis

Simulation

Theory

Computing
HPC is equally important in experimentation

Experimentation

Data Analysis

Simulation

Theory

Computing

Growth in Sequencers, CCDs, sensors, etc.
Integration of Simulation and Observational Science

Intermediate Palomar Transient Factory
- Nightly images transferred
- Subtractions, machine learning
- Candidates in database in < 5 minutes
- Simulations aid in interpreting data

Old School Scientific Data Search
Automated Search, Meta-Data Analysis, and On-Demand Simulation

Automated metadata extraction using machine learning

Jobs submitted by “bots” based on queries; algorithms extract informatics for design
Part 2

ASCR Facilities need to adapt
ESnet: Exponential data growth drives capacity

Petabytes/month

Traditional IP
Transatlantic
Big science data

Science DMZ to deliver bandwidth to the end users
OSCARS for bandwidth reservation

100 Exabytes/year by 2024!
ESnet: Discovery Unconstrained by Geography

LCLS/NERSC/Esnet Superfacility demo for Photosystem II
→ 3x network traffic

ESnet-6 Upgrade Options trade off risk and capability

- **Software Defined Networking**
  - Programmable switches may improve cost and speed
  - Adapt lower level network layers for major science flows

- **Packet Optical**
  - Combine hardware for packetization/routing with optical transport
  - Lower cost

- **Current Architecture**
  - Keep packet and optical separate with current fixed routing tables
  - Known technology

Network performance enables efficiency of centralized computing
Systems configured for data-intensive science

NERSC Cori has data partition (Phase 1, Haswell) pre-exascale (Phase 2, KNL preproduction) WAN-to-Cori optimized for streaming data: 100x faster from LCLS to Cori and Globus to CERN
Real-time queue prototyped at NERSC

- In 1998 dedicated hardware; now prototype queue on Cori
- <1% of NERSC allocation
- Cryo-Em, Mass spec, Telescopes, Accelerator, Light sources

Cryo-EM: Image classification
Nogales Lab

PTF: Image subtraction pipeline

ALS: 3D Reconstruction, rendered on SPOT web portal
Containers for HPC Systems

- Data analysis pipelines are often large, complex software stacks
- NERSC Shifter (with Cray), supports containers for HPC systems
- Used in HEP and NP projects (ATLAS, ALICE, STAR, LSST, DESI)
Part 3

ASCR Research challenges are substantial
CAMERA: Math for the Facilities

- Designing mathematical algorithms to allow real-time analysis next to the equipment
- Real-time streaming ptychography—ALS, delivered to NSLS2, LANL, BESSY,
- New algorithms to transform manual into automatic analysis
- Inventing new math and models to match new acquisition technologies
- Robust and reliable codes and data flow: workflow environments
- Cultural and Sociological Challenges
- Compare and integrate multiple analysis tools
- Multi-modal: Building the math that fuses information from multiple experiments
- Automatic image processing for the ALS/GE
- Fluctuation scattering and single particle imaging for the LCLS
- Workflow and access to remote supercomputers: XiCAM for ALS, SSRL, APS, NSLS2
CAMERA: Making the connections

- Probabilistic Graphical Models
- Constrained optimization
- Fast PDE solvers: (Level Set, DG,...)
- Model-based reconstruction
- Iterative Phasing
- Spectral analysis
- Machine learning, feature detection, persistent homology
- New mathematical modeling
- Discrete mathematics/Computational geometry
- Linear Algebra (Selected inversion, fast pseudoinverse approximation,...)

- Materials Design (Zeo++)
- Electronic Structure (PEXSI)
- Image Analysis/Tomography (QuantCT,F3D)
- Ptychography (SHARP)
- GISAXS
- Fluctuation/Single Particle

- BES Functional Electronic Materials
- BES Nanoporous Materials
- EFRC Gas Separation
- CLS (Canada)
- NCEM
- APS
- SSRL
- DIAMOND (UK)
- ESRF (Grenoble)
- PETRA III (Germany)
- LBNL
- LLNL/LANL
- BNL
- ORNL
- ANL
- NERSC
- ALCF
- OLCF
- SIESTA, CP2K, ImageJ, Fiji

Universities: e.g.: Berkeley, Northwestern, Georgia Tech, Rice, UCSD, U.I.C, McMaster, Austin, Stanford,...

Bosch, Samsung, Intel, GE, ...

- ASCR Facilities
- =BES Centers/Internat.
- =BES (and other) Facilities
- =LABS
- =3Rd Party Codes
- =Industry
- =Universities
## Analytics vs. Simulation Kernels:

<table>
<thead>
<tr>
<th>7 Giants of Data</th>
<th>7 Dwarfs of Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic statistics</td>
<td>Monte Carlo methods</td>
</tr>
<tr>
<td>Generalized N-Body</td>
<td>Particle methods</td>
</tr>
<tr>
<td><strong>Graph-theory</strong></td>
<td>Unstructured meshes</td>
</tr>
<tr>
<td>Linear algebra</td>
<td>Dense Linear Algebra</td>
</tr>
<tr>
<td>Optimizations</td>
<td><strong>Sparse Linear Algebra</strong></td>
</tr>
<tr>
<td>Integrations</td>
<td>Spectral methods</td>
</tr>
<tr>
<td>Alignment</td>
<td>Structured Meshes</td>
</tr>
</tbody>
</table>
Machine Learning Mapping to Linear Algebra

- Logistic Regression, Support Vector Machines
- Dimensionality Reduction (e.g., NMF, CX/CUR, PCA)
- Clustering (e.g., MCL, Spectral Clustering)
- Graphical Model Structure Learning (e.g., CONCORD)
- Deep Learning (Convolutional Neural Nets)

- Sparse Matrix-Sparse Vector (SpMSpV)
- Sparse Matrix-Dense Vector (SpMV)
- Sparse Matrix Times Multiple Dense Vectors (SpMM)
- Sparse - Sparse Matrix Product (SpGEMM)
- Dense Matrix Vector (BLAS2)
- Sparse - Dense Matrix Product (SpDM³)
- Dense Matrix Matrix (BLAS3)

Aydin Buluc
Software implementations at scale in pipeline

MicroCT imaging → Segmentation → Topological Analysis

Analysis → Simulation → Visualization
Interactive Analytics using Jupyter

Science notebooks through Jupyter (iPython)

- Widely used in science
- Interactive HPC LDRD

Deployed at NERSC:
- >100 users pre-production
Random Access Analytics

- Genome assembly “needs shared memory”

Global Address Space
- Low overhead communication
- Remote atomics
- Partitions for any structure

Scales to 15K+ cores
Under 10 minutes for human
First ever solution

Data Fusion for Observation with Simulation

• Unaligned data from observation
• One-sided strided updates

Scott French, Y. Zheng, B. Romanowicz, K. Yelick

Hawaii hotspot geology
Productive Programming

Speed
Run programs up to 100x faster than Hadoop MapReduce in memory, or 10x faster on disk.

- High failure rate
- Slow network
- Fast (local) disk

And Spark is still 10x+ slower than MPI
SPARK Analytics on HPC

SPARK on HPC vs. clusters
Network, I/O, and virtualization all key to performance

Chaimov, Malony, Iancu, Ibrahim, Canon, Srinivasan
Architectures for Data vs. Simulation

- Massive Independent Jobs for Analysis and Simulation
- Compute-Intensive Dense LA for Deep Learning and Simulation
- Nearest Neighbor Simulation
- All-to-All Simulation (3D FFTs) and analysis
- Random access, large data Analysis

Different architectures for simulation? Can simulation use data architectures?
Data processing with special purpose hardware

- General trend towards specialization for continued performance growth
- Data processing (on raw data) will be first in DOE

Particle Tracking with Neuromorphic chips
Computing in Detectors
Deep learning processors for image analysis
FPGAS for genome analysis
Extreme Data Science

The scientific process is poised to undergo a radical transformation based on the ability to access, analyze, simulate and combine large and complex data sets.
Superfacility: Integrated network of experimental and computational facilities and expertise

Experimental Facilities

Computing and Data Facilities

User Community

Expertise

ESnet

A single interconnected “facility” where data is acquired, stored, analyzed and served

Methods, models, analytics, and software

Execution plan: one science area at a time