Machine Learning for Science

Kathy Yelick
Associate Laboratory Director for Computing Sciences

U.S. DEPARTMENT OF ENERGY
Office of Science

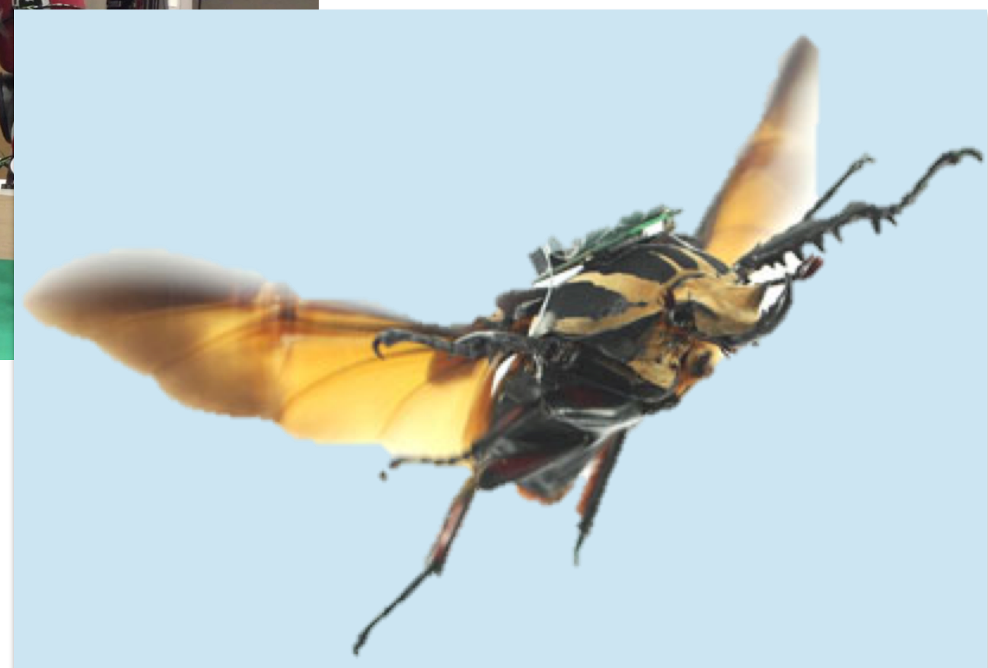
Machine Learning on Images and Videos

Self-driving cars interact with human drivers

Machine Learning on language understanding

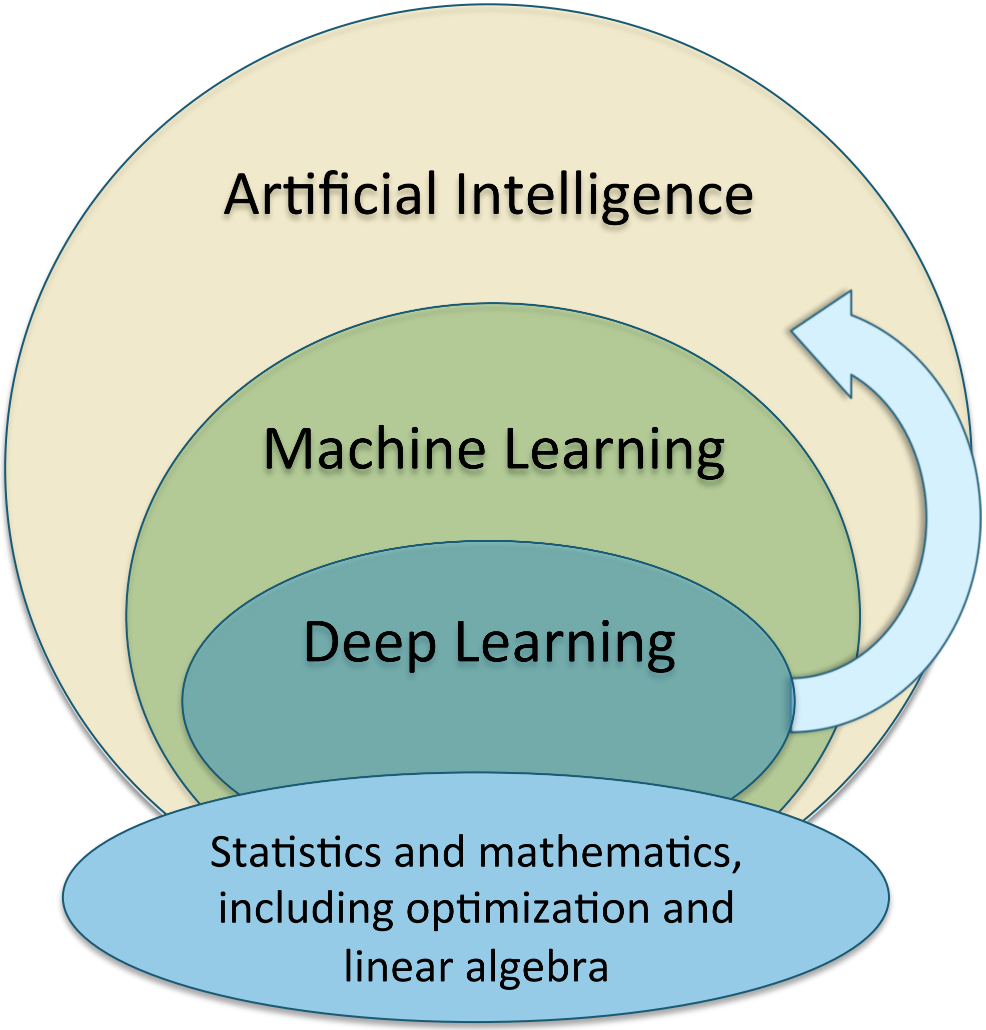
AliBaba's deep learning software beats humans at reading comprehension test

Machine Learning for Robotics



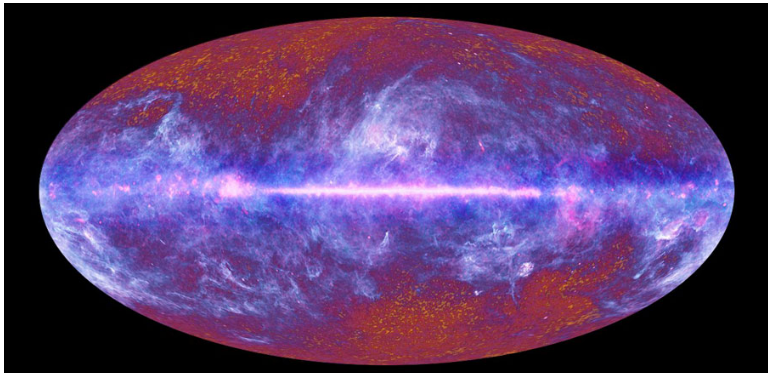
Robots

Artificial Intelligence, Machine Learning and Deep Learning

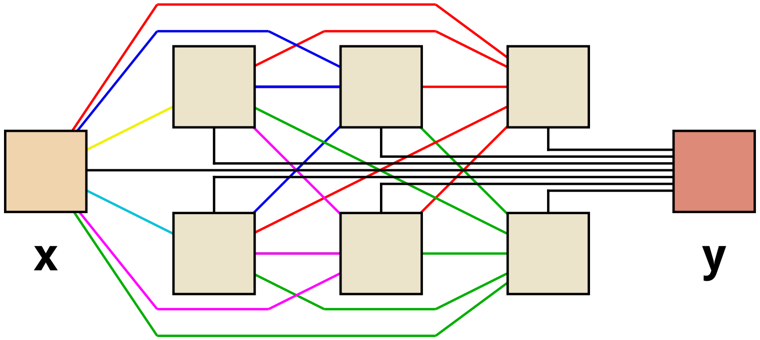


Three ingredients for machine learning

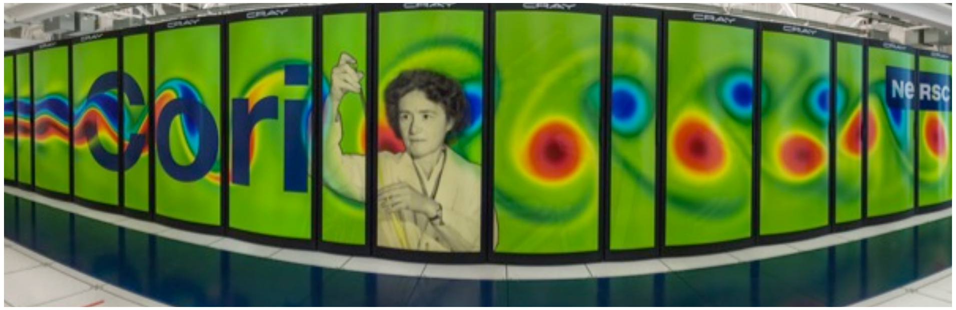
Data



Algorithms



Machines



AI Revolution in reasoning

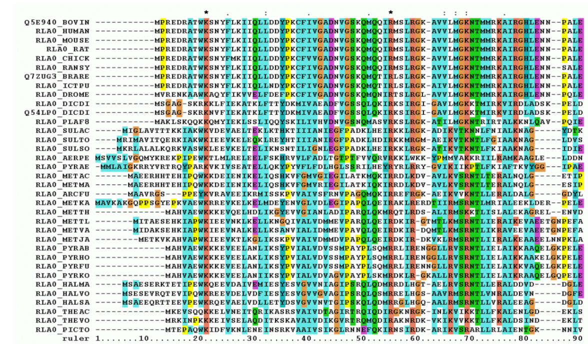
Google's AlphGo Zero beats humans after self-training for 70 hours

The Data

Scientific data needs machine learning

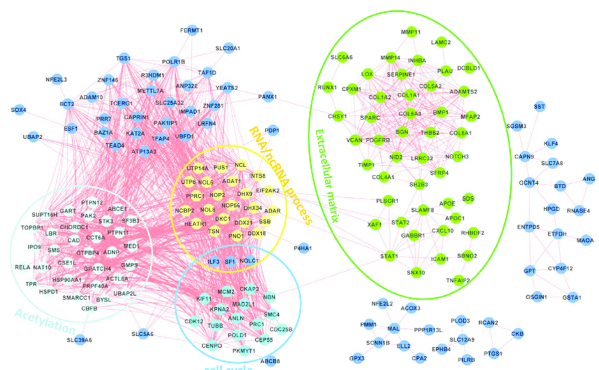
Image / Video Processing

Text

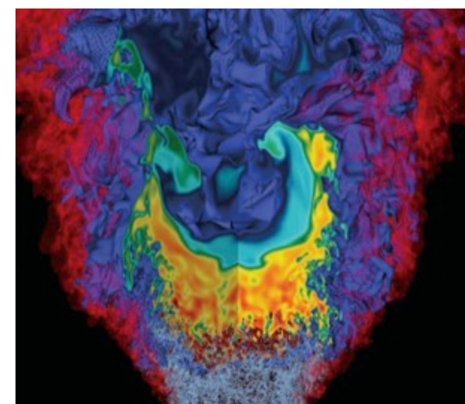


Genomics

Signal Processing

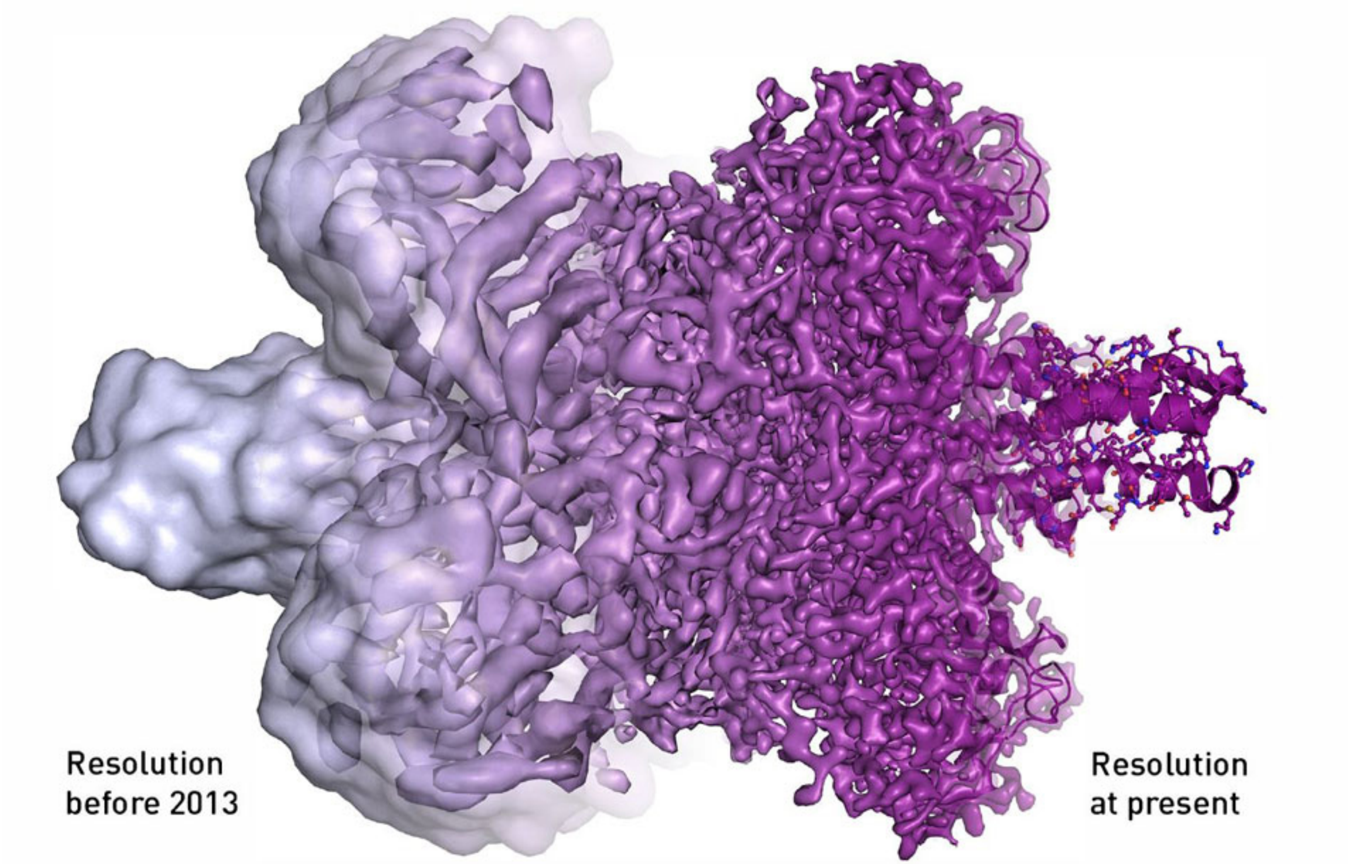


Graph Analytics (Relationships)



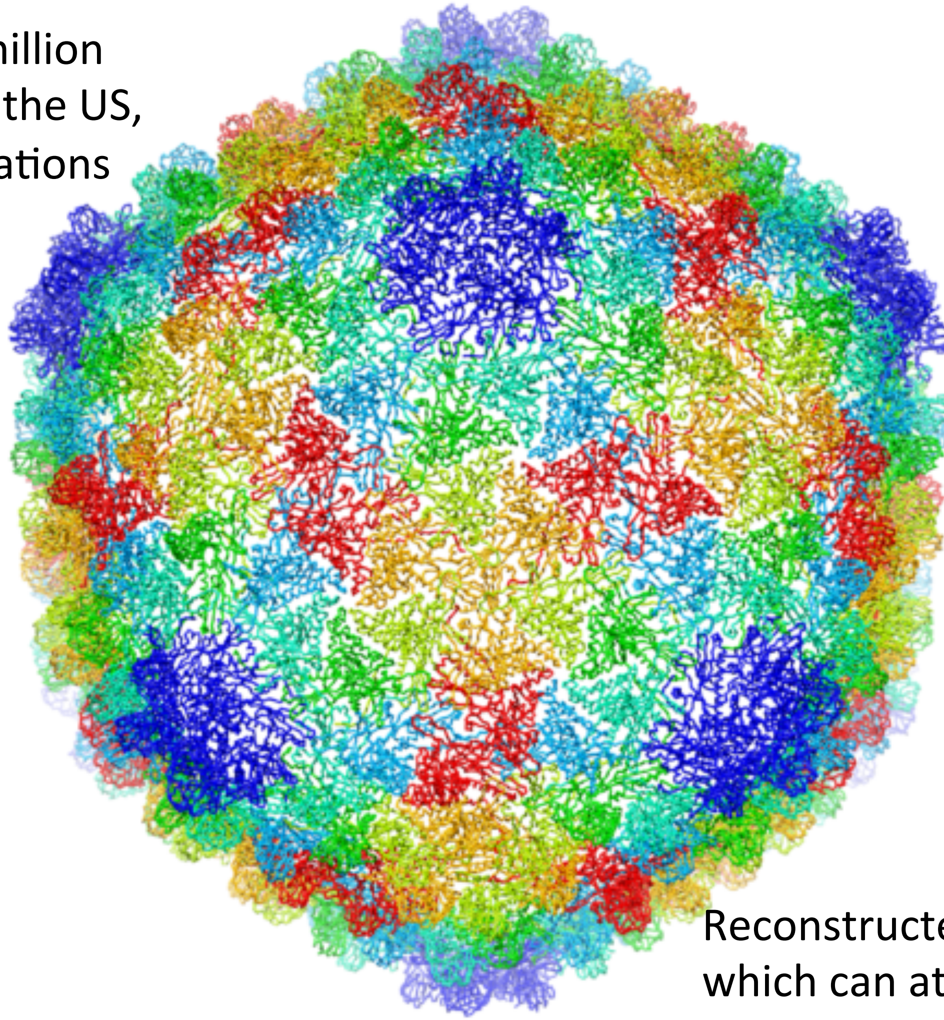
Simulation Analytics

Detectors: the "sensory system" for science



Berkeley Lab advances detector technology for many fields of science, including (above CryoEM) biology, cosmology, material science, physics, and more.

Salmonella causes 1 million foodborne illnesses in the US, with 19,000 hospitalizations and 380 **deaths**.



Reconstructed Bacteriophage P22 which can attack Salmonella in foods

Embedded sensors in infrastructure

Fiber-optic cables can be used a sensors for urban seismic hazard analysis, to monitor soil layer changes, detect nuclear explosions, and do global seismic imaging

Embedded sensors in infrastructure

Data analysis of power grid micro sensors trained to identify intrusion and other events

Climate Simulations

Preliminary CAM5 hi-resolution simulations (0.25°, prescribed aerosols)

Michael Wehner, Prabhat, Chris Algieri, Fuyu Li, Bill Collins
Lawrence Berkeley National Laboratory

Kevin Reed, University of Michigan

Andrew Gettelman, Julio Bacmeister, Richard Neale
National Center for Atmospheric Research

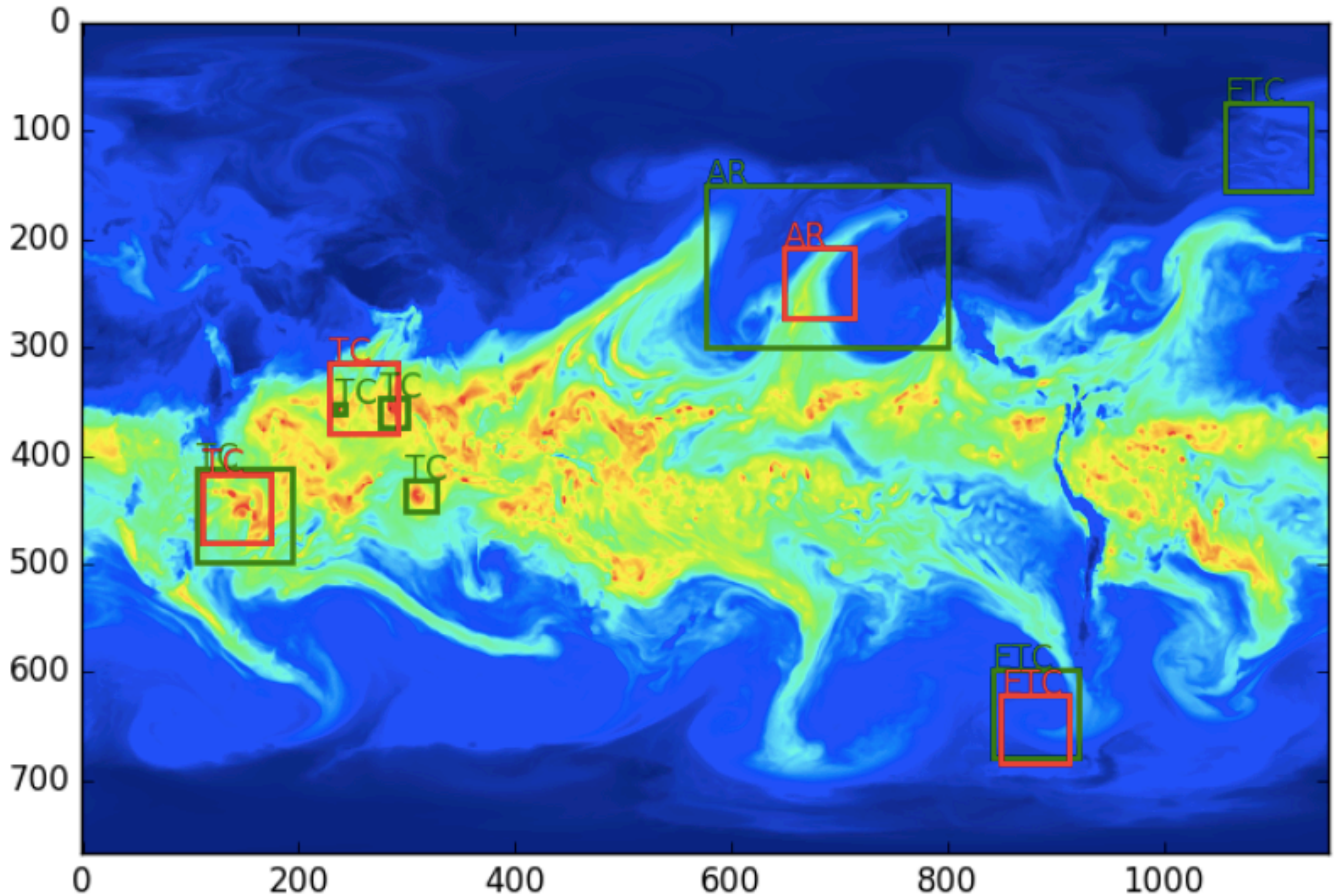
June 1, 2011

BERKELEY LAB
Lawrence Berkeley National Laboratory

NERSC

The Algorithms

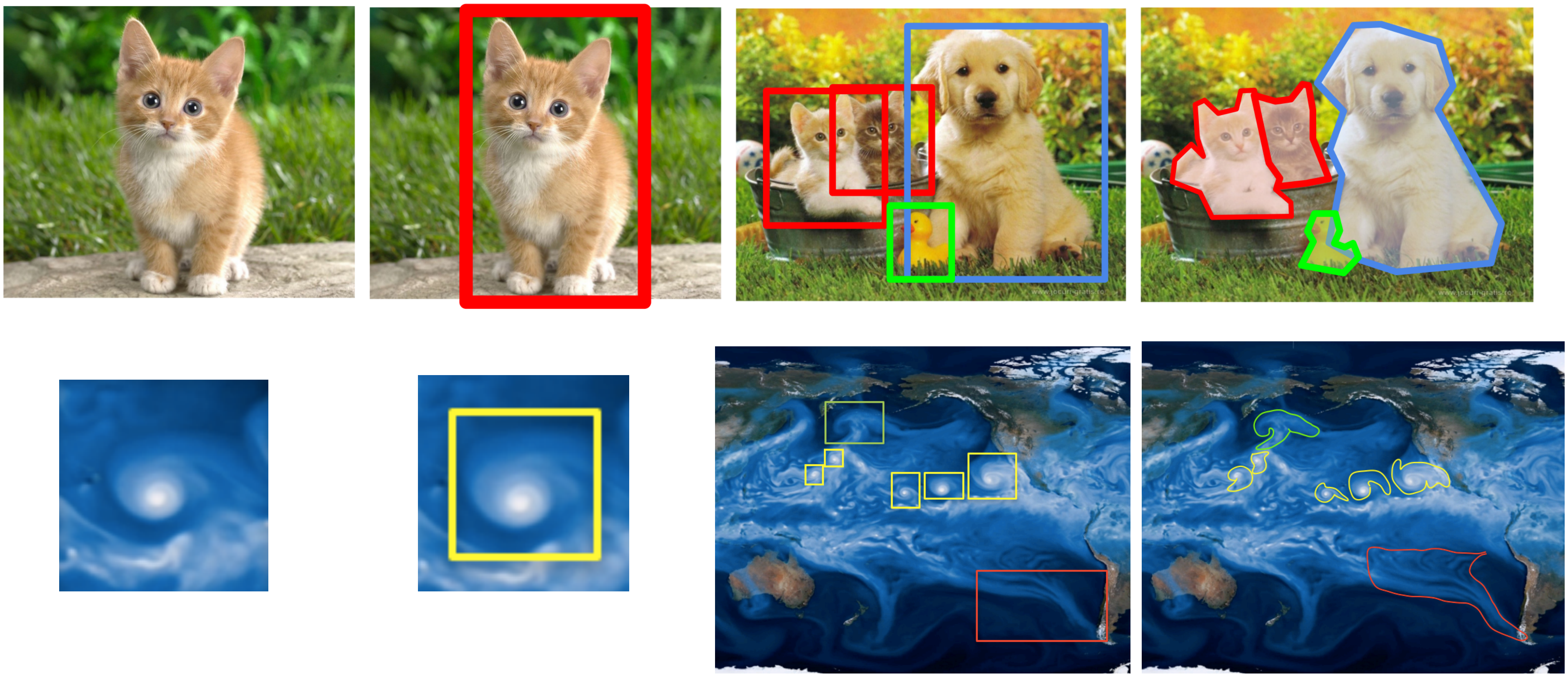
Finding Storms in Simulations



Ground Truth
Prediction

Climate Science Tasks

Classification	Classification + Localization	Object Detection	Instance Segmentation

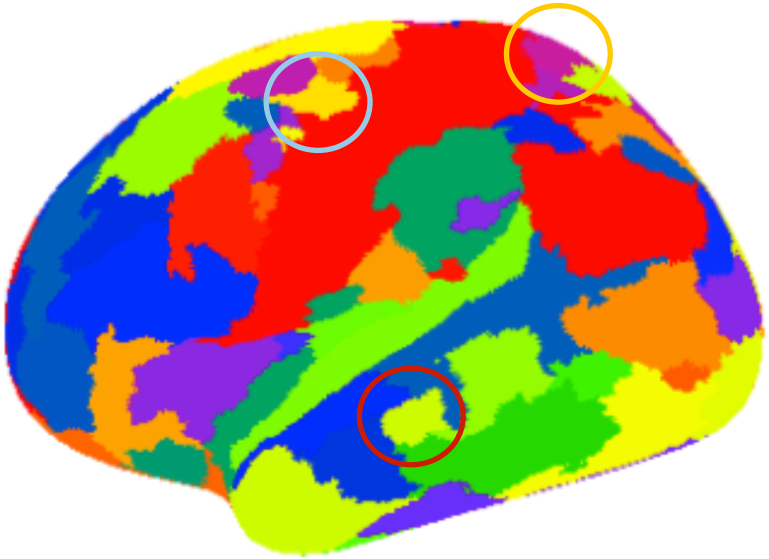


Learn the relationship between features with Graphical Model Estimator

Source: https://media4.s-nbcnews.com/i/newscms/2017_25/958456/150401-dna-strand-mn-1645_9d74198e59853eb79be3124a876ad4fd.jpg

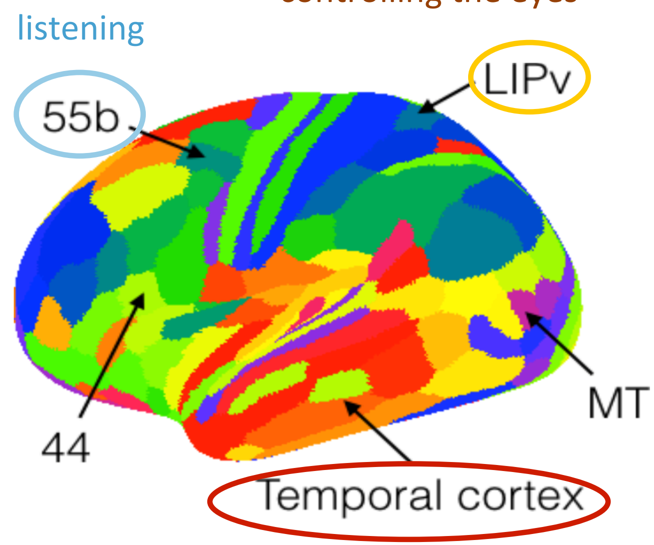
New Algorithm for HPC discovering regions and co-regions

Automatic parcellation from fMRI data alone



$\lambda_1 = 0.48,\ \lambda_2 = 0.39,\ \epsilon = 3,$
$\%$ of best score $= 100$

Baseline parcellation from Glasser
[Glasser et al. 2016]

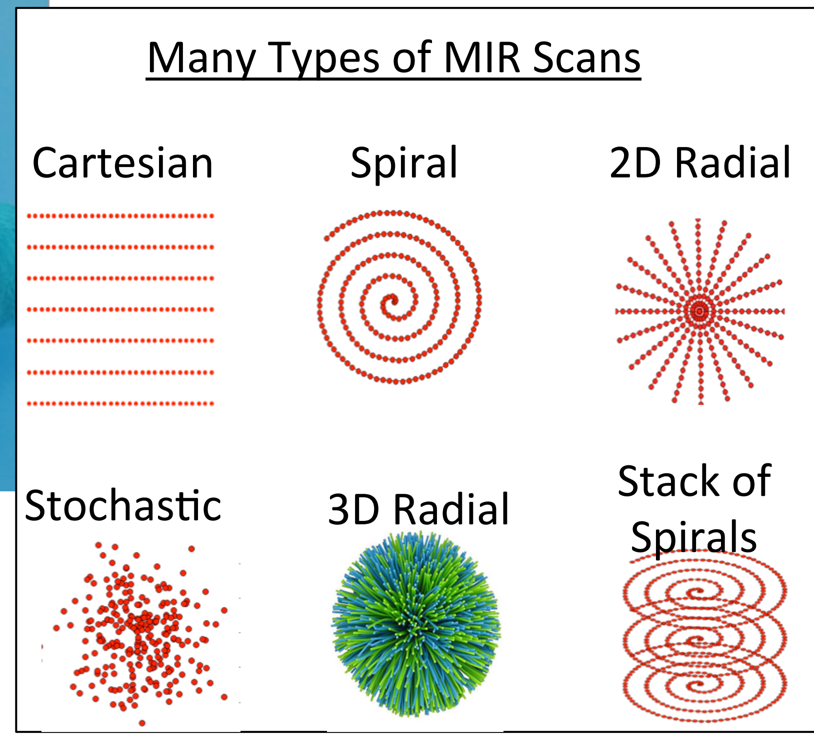


First of kind analysis at this scale using new algorithm and high performance computing at LBNL

Koanantakool, Oh, Buluc, Morozov, Oliker, Yelick, AISTAT 2018, to appear.

Wearable MRI sensors + HPC Analytics

Wearable MIR sensors [Arias UCB]

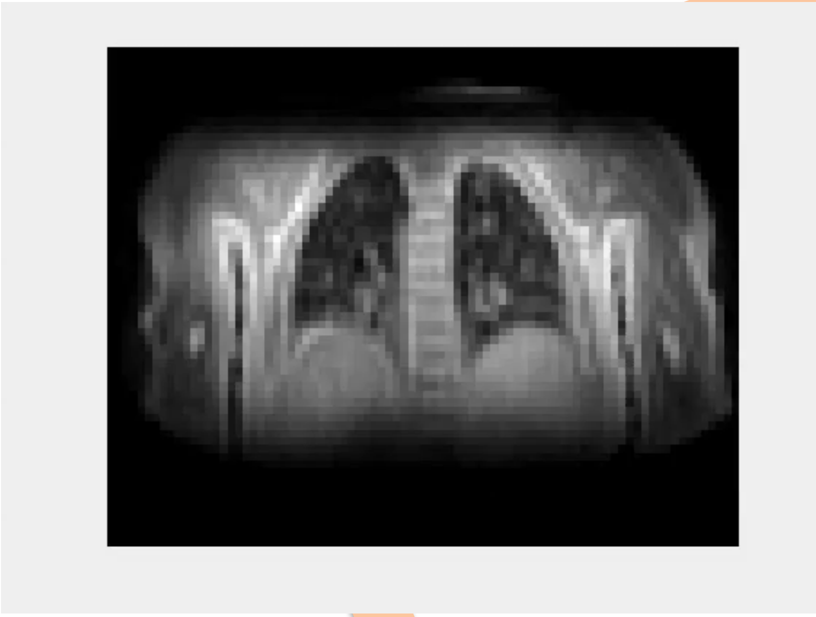


Compressed sensing algorithms [Lustig, UCB]

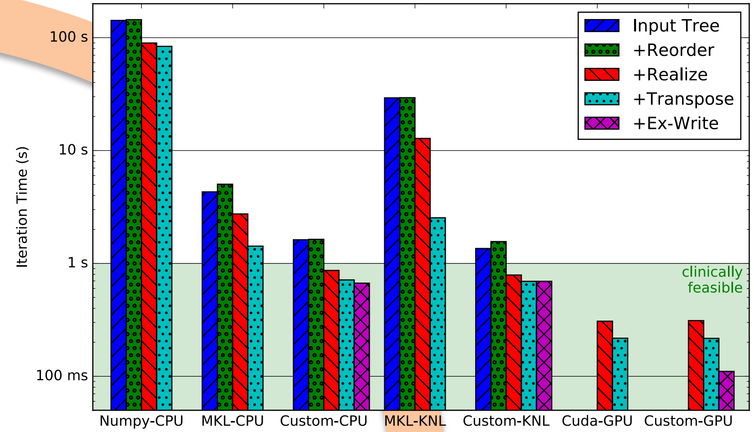
Goals:

1) **reduce time in MRI**
2) **improve patient experience**
3) **better quality of images**

Real-Time Analytics in Health



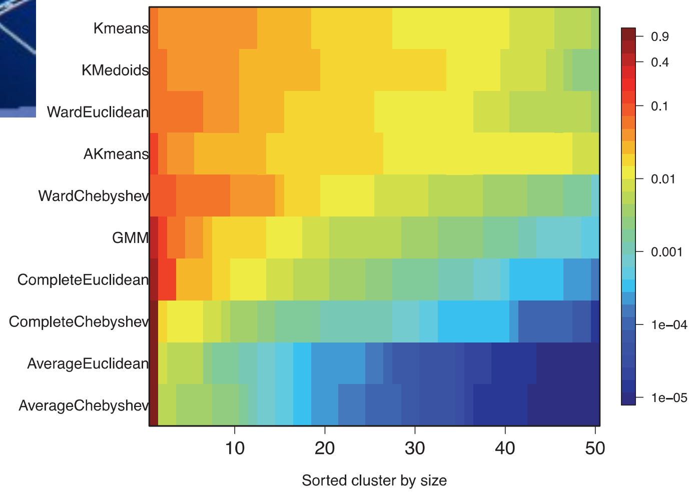
Compressed Sensing Approach by Mike Lustig et al
MRI results Wenwen Jiang



3 min goal (1 sec/iteration)
Michael Driscoll HPC optimization

Machine Learning for Behavior in Energy Use

Understanding energy and human behavior

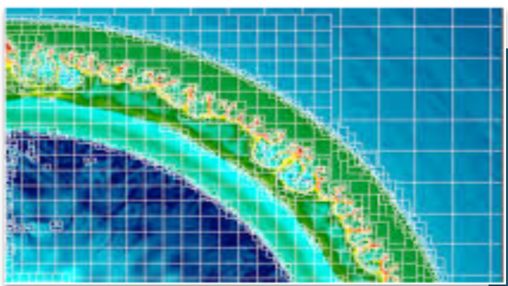


How well do various algorithms work?

What are the particular challenges in science?

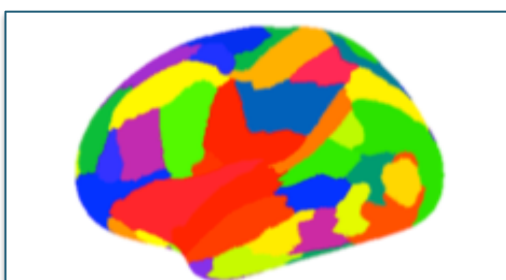
Scale

- Data rates from detectors
- Machine scale, novelty and performance



Complexity

- Adaptive, hierarchical
- Multi-modal, noisy



Interpretability

- Explainable, understandable, robust
- Physically realizable

The Machines

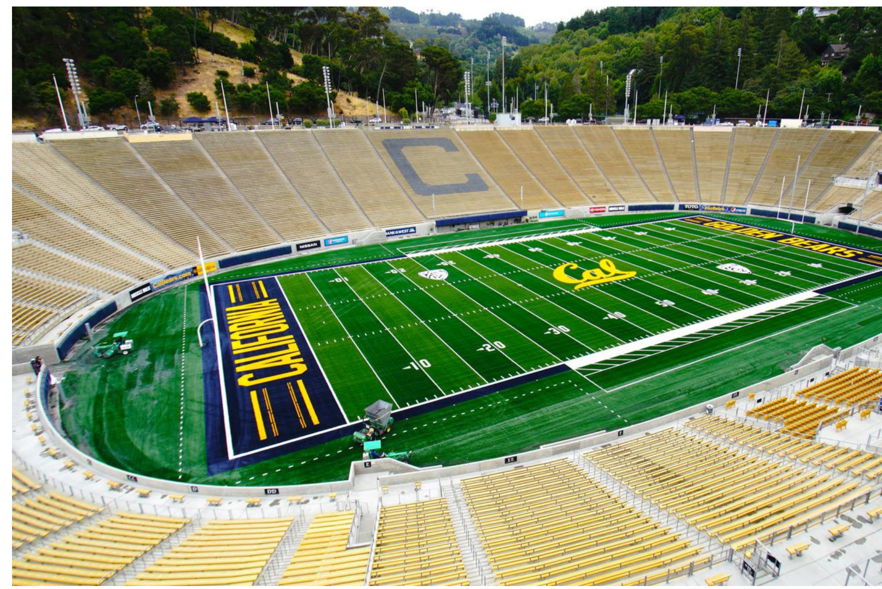
NERSC Supercomputing for Science and Energy

State-of-the art computing for the broad DOE science community – over 7000 users, 700 applications

Deep Learning at 15 Petaflops

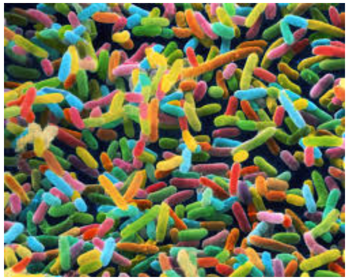
Berkeley Lab scientists new parallel algorithm for deep learning on climate and particle physics data at 15 Petaflops

= 4 x



Microbiome analysis uses high performance computing

Environment



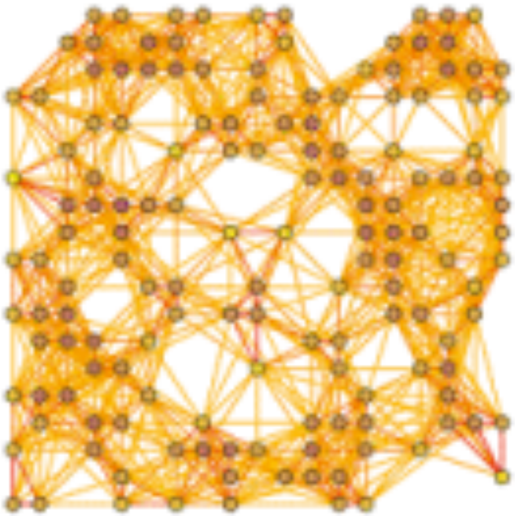
Health

Bio-Energy

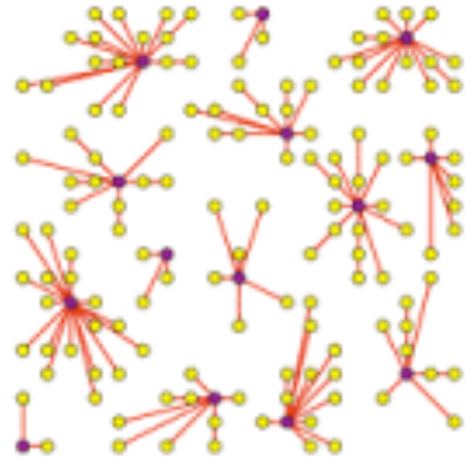
Bio-Manufacturing

Microbiome analysis uses high performance computing

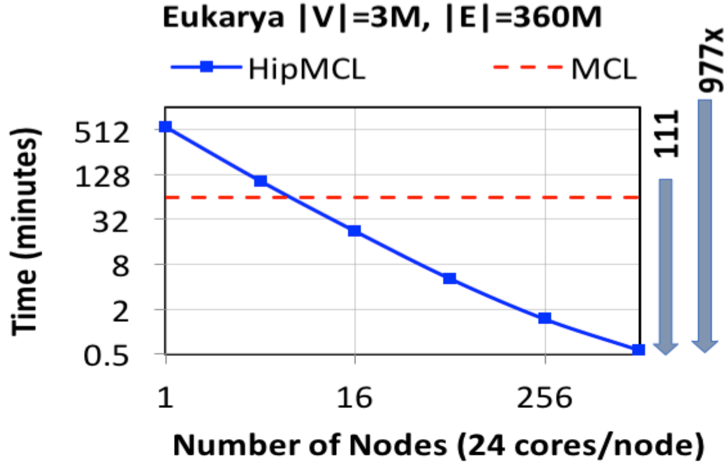
Microbiome analysis with machine learning



Similarities between genes (proteins)

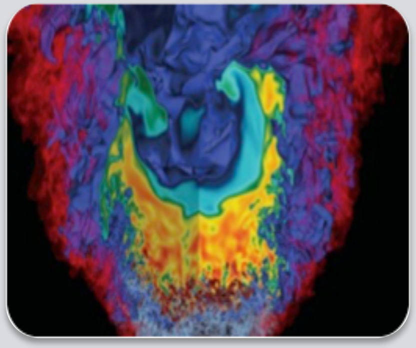


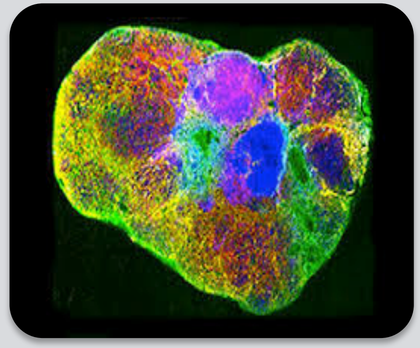
Clusters of related ones

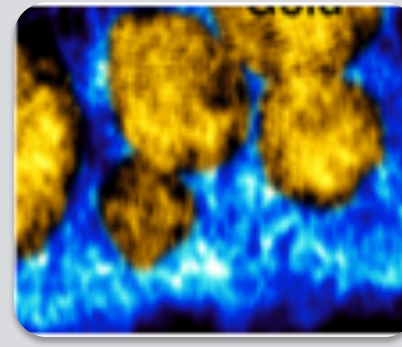


New science, impossible without HPC

The Berkeley National Lab Advantage in ML for Science







HPC and networking

Systems and Expertise

Applied Math

Driven by Science

Team Science

End-to-end Solutions

Advanced Detectors

Data source

AI is also revolutionizing science revolution