Computational Science at BNL

Francis J. Alexander
Deputy Director
Computational Science Initiative

BROOKHAVEN NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

BNL is a Data Driven Science Laboratory: HEP Data is the largest component

- BNL provides Data-rich Experimental Facilities:
  - RHIC - Relativistic Heavy Ion Collider - supporting over 1000 scientists world wide
  - NSLS II - Newest and Brightest Synchrotron in the world, supporting a multitude of scientific research in academia, industry and national security
  - CFN - Center for Functional Nanomaterials, combines theory and experiment to probe materials
  - ATF - Accelerator Test Facility
  - LHC ATLAS - Largest Tier 1 Center outside CERN
  - ARM - Atmospheric Radiation Measurement Program - Partner in multi-side facility, operating its external data Center
- BNL supports additional large scale Experimental Facilities:
- Belle II - Computing for Neutrino experiment
  - QCD - Computing facilities for BNL, RIKEN & US QCD communities

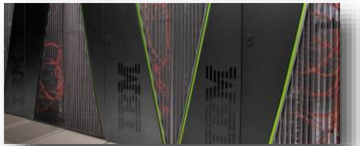
RHIC

NSLS II

CFN

ATLAS

QCD



Computational Science Initiative (CSI)

- Established December 2015
- **Vision**
- Translating leading Computer Science and Applied Mathematics research into measurably improved scientific discovery processes
- **Focus**
- Data analysis, numerical modeling, support for experiments, reusable knowledge repositories
- **Organizational Structure**
- Scientific Data and Computing Center
- Computer Science and Applied Math Research
- Computational Science Laboratory
- Center for Data Driven Discovery
- Computing For National Security

Approaching ~90 staff and students

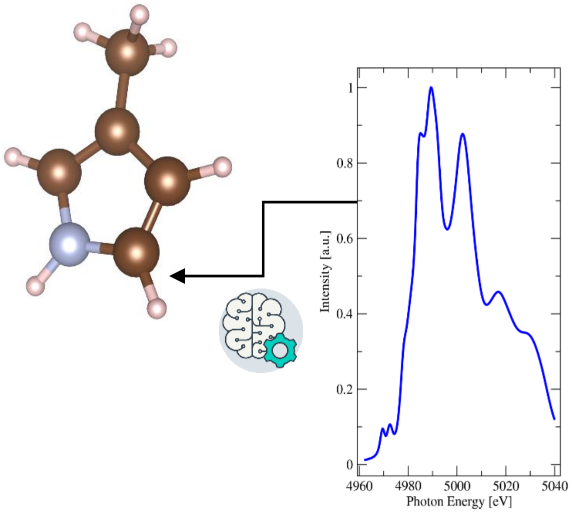
CSI's new home, including proposed new Data Center (CFR)

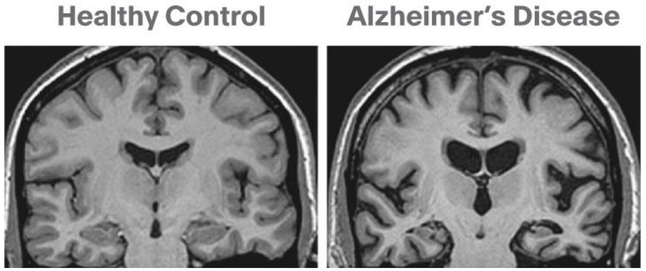
Key research initiatives:
Making Sense of Data at Exabyte-Scale and Beyond

- **Real-Time Analysis of Ultra-High Throughput Data Streams**
    - Integrated, extreme-scale machine learning and visual analytics methods for real time analysis and interpretation of streaming data.
    - New in situ and in operando experiments at at large scale facilities (e.g., NSLS-II, CFN and RHIC)
    - Data intensive science workflows possible in the Exascale Computing Project
    - Analysis on the Wire
- **Autonomous Optimal Experimental Design**
    - Goal-driven capability that optimally leverages theory, modeling/simulation and experiments for the autonomous design and execution of experiments
    - Complex Modeling Infrastructure
- **Interactive Exploration of Multi-Petabyte Scientific Data Sets**
    - Common in nuclear physics, high energy physics, computational biology and climate science
    - Integrated research into the required novel hardware, system software, programming models, analysis and visual analytics paradigms

Impactful Machine Learning

- Research Leaders in extreme scale, science specific new Machine Learning Methods
- Joint Research Center for **Machine Learning @ Scale** with Carnegie Mellon University
- Successful applications
  - Advanced analysis deployed at several NSLS-II beamlines and CFN instruments
    - Decipher in real time the size and shape of catalysts from X-ray absorption spectra – enable catalyst optimization during experiment
  - BNL algorithm originally developed for ASCR achieved nearly 100% Alzheimer's detection accuracy and 83% prediction accuracy for early-stage Alzheimer's. The algorithms were trained on magnetic resonance imaging and brain phenotyping clinical data. Further application to VA and BER Kbase.





Artificial Intelligence-driven Optimal Experimental Design under Uncertainty and with Limited Resources

Example problems

- Improve biofuel yield for algal systems
- Determine phase diagrams for complex materials
- Use HPC resources efficiently for costly biomolecular simulations
- Optimize RHIC Beam Energy Scan data taking strategy

DOE/VA Pilot Projects benefit both Departments

HPC, modeling/simulation and large scale data analysis to VA data to improve healthcare for our Veterans
Develop scalable machine learning algorithms for challenging problems in DOE

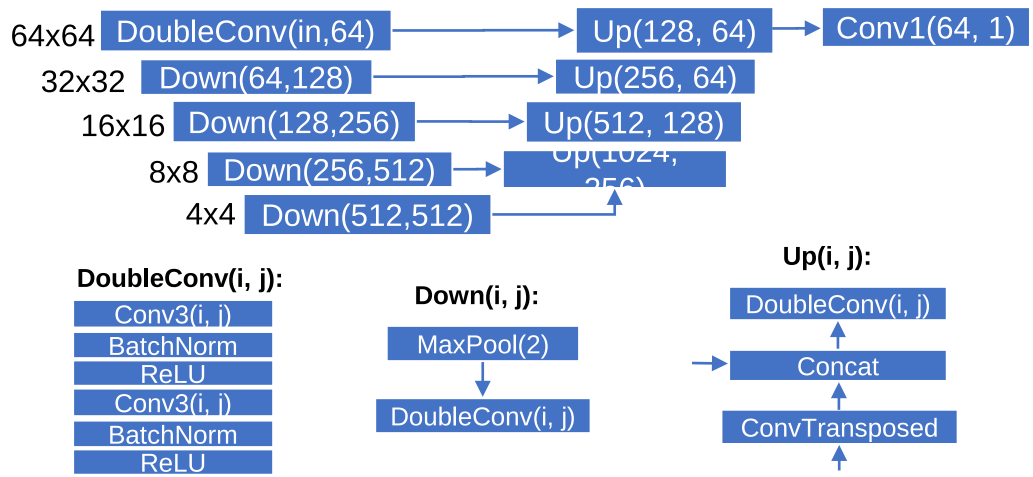
- ***Enhanced prediction and diagnosis of Cardiovascular Disease* (CVD)**
  - Develop methods to inform individualized drug therapies to prevent, pre-empt and treat CVD.
  - Enhance prediction, diagnosis and management of major CVD subtypes in Veterans
- ***Precision discrimination of lethal from non-lethal Prostate Cancer***
  - Build improved classifiers to distinguish lethal from non-lethal prostate cancers.
  - Reduce unnecessary treatments /provide an increased quality of life for patients
- ***Patient-specific analysis for Suicide Prevention***
  - Provide tailored and dynamic suicide risk scores for each Veteran at risk.
  - Create a clinical decision support system that assists VA clinicians in suicide prevention efforts, and helps to evaluate effectiveness of various prevention strategies.

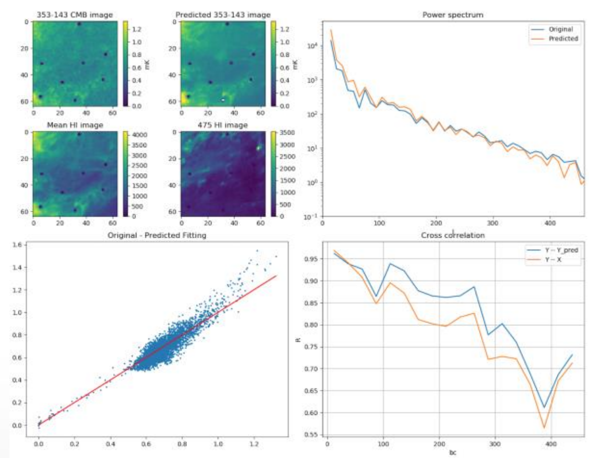
- ***Scalable Algorithms for***
  - ***Binary and Multiclass classification***
  - ***Data Imputation (for missing data)***
  - ***Imbalanced data problems with constrained resources***
- ***Integrating large multimodal data sets (>20M patients)***
  - ***Images***
  - ***Mechanistic biological models***
  - ***Full Genomes***
  - ***Longitudinal Data***
- ***UQ and error analysis (skill assessment)***
- ***Potential Applications to many DOE programs in***
  - ***Genomics***
  - ***Climate state assessment***
  - ***Prediction of Complex Systems Behavior***

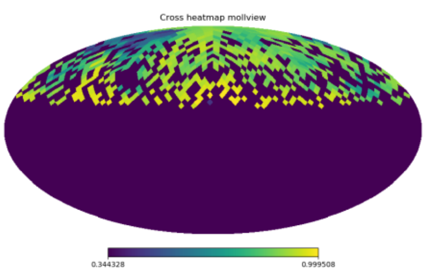
SciDAC HEP project

- **Task**: Predict CMB foreground using 21cm Hydrogen line.
- Assumption: CMB foreground is dominated by dust thermal radiation. The distribution of dust and Hydrogen atoms in the cosmos are almost the same.
- Input: 50 Hydrogen maps of different frequencies
- Output: CMB foreground
- The result is shown with Fourier power spectrum.
- U-Net based model is proposed with loss function defined as negative cross correlation of the Fourier power spectrum plus image L1 Loss as regulation.

U-Net:







The dark spots are point source mask, including pulsar, galaxy.

U.S. DEPARTMENT OF ENERGY
BROOKHAVEN NATIONAL LABORATORY

Preparing Lattice QCD for Exascale Computing: Software and Algorithms

Chulwoo Jung, Meifeng Lin
with contributions from **Peter Boyle, Yong-Chull Jang** and **Christoph Lehner**

BROOKHAVEN NATIONAL LABORATORY

ENERGY

ECP EXASCALE COMPUTING PROJECT

BROOKHAVEN NATIONAL LABORATORY

Lattice QCD at Exascale

Lattice QCD simulations are numerically demanding, and require sustained use of a large portion of the existing DOE leadership-class computers. With exascale computing resources, we expect to

- Increase the precision of critical calculations to understand the fundamental symmetries of the universe and the structure of matter, such as CP violations, muon g-2 (muon anomalous magnetic moment), the internal structure of protons, etc.
- Extend the calculations of the light nuclei and multi-nucleon systems in nuclear physics with quark masses that are closer to their values in nature. For example, first-principle simulations of the Helium nucleus.

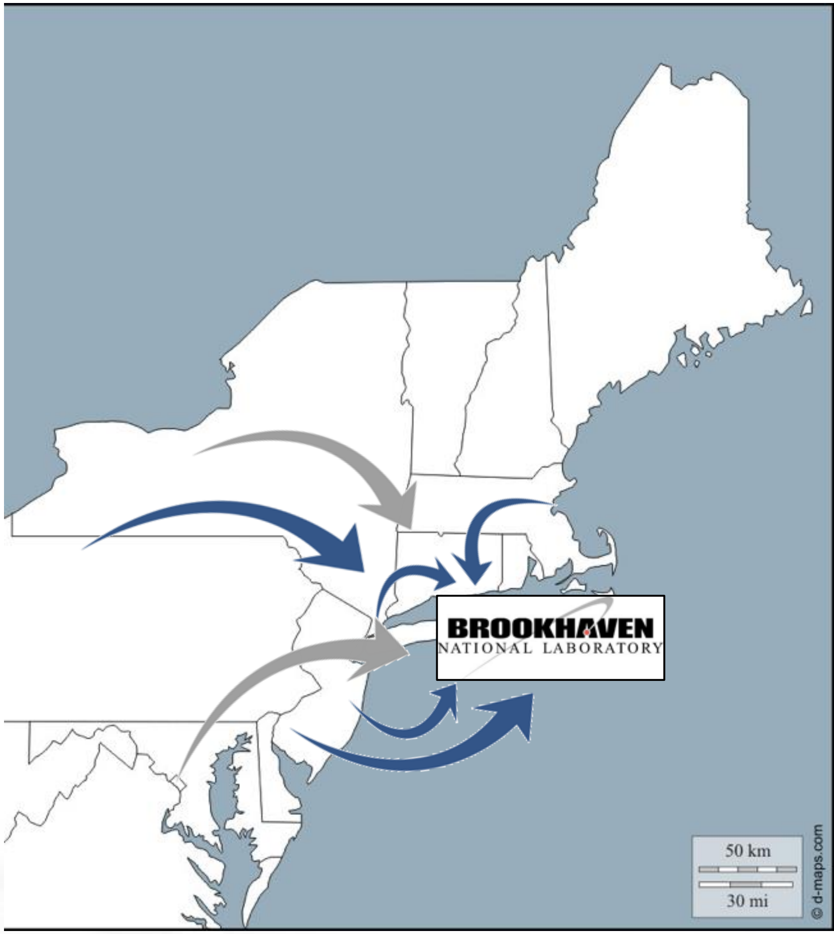
Software Requirements

- Efficiency: Should be able to effciently exploit the expected multiple levels of parallelism on the exascale architectures. Need to conquer the communication bottleneck.
- Flexibility: Should be flexible for the users to implement di↵erent algorithms and physics calculations, and can provide easy access to multi-layered abstractions for the users.
- Performance Portability: Should be portable to minimize code changes for different architectures while maintaining competitive performance.

The vision: BNL as a hub for next-generation quantum information science in the Northeast

NeQSys: Northeast Quantum Systems

Provide an collaborative environment for research integration across the spectrum of activities from materials to computational applications



Potential activities

- Joint Appointments
- Visiting Faculty
- Student/postdoc exchanges RE's

- Workshops
- Colloquia
- Summer School

- University inter-movement
- Masters Degree in QC
- Problem Market

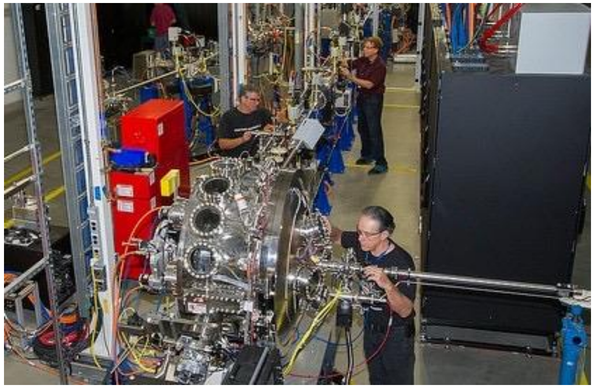
U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

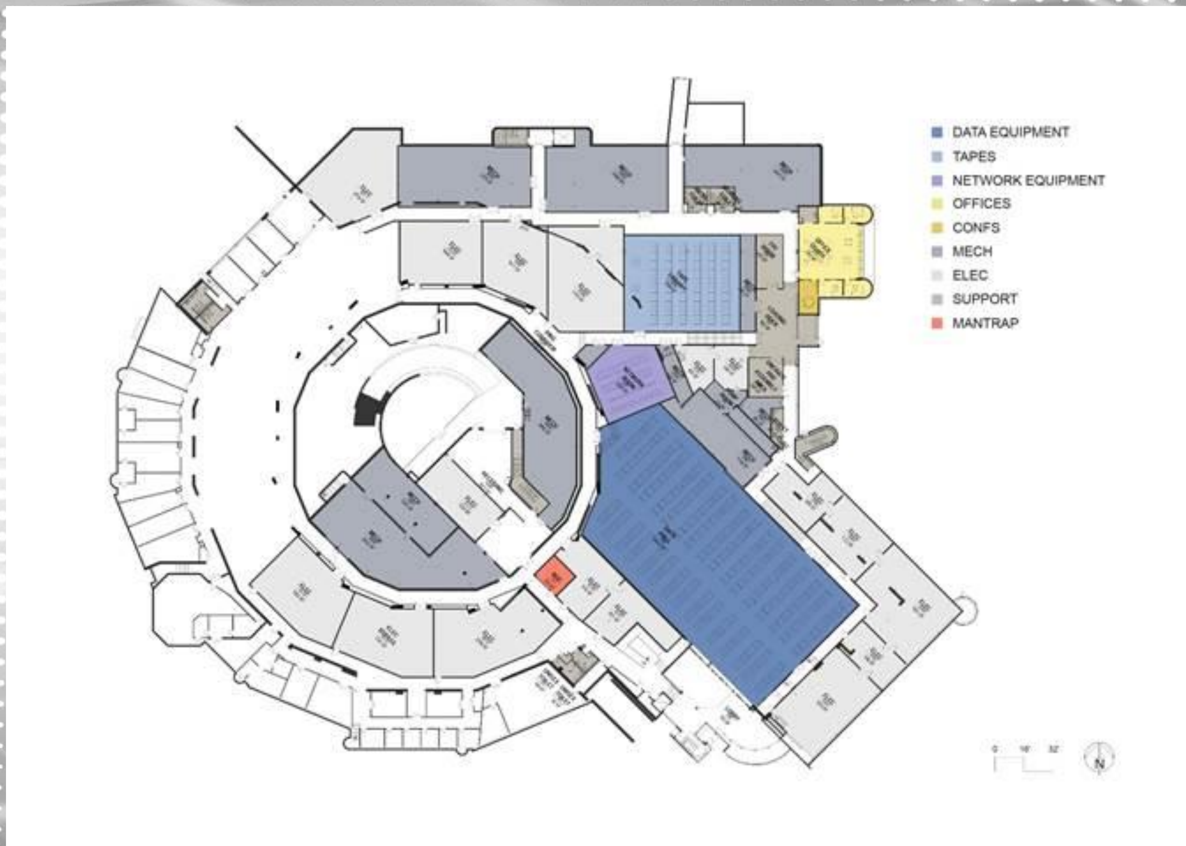
New York State has made Strategic Investments in Brookhaven National Lab

- $65 million over five years to support three transformative projects:
  - Build high-energy x-ray beamline at NSLS-II, enabling researchers/industry to map the three-dimensional structure of batteries and energy-storage materials in real time, under real operating conditions ($25 million)
  - Advance Lab's "Big Data" computation capabilities with focus on NSLS-II, which will generate enormous amounts of data ($15 million)
  - Build prototypical ERL facility for the technical design of a future electron-ion collider, transforming RHIC into eRHIC ($25 million)

- RecentAdditions:
  - A new $15M grant to develop a Cryo Electron Microscopy Center with NSLS-II for biological studies
  - A $20M addition to the LIRR budget to relocate the Yaphank station to BNL



The CSI Compute Center upgrade will receive $30M in FY18



BROOKHAVEN NATIONAL LABORATORY

Secure Computing (Various levels incl PHI)

Laboratory Space of a variety of kinds

Potential for 14MW of Power

U.S. DEPARTMENT OF ENERGY

- Thank you