Computational Science Initiative

U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

70 YEARS OF DISCOVERY

A CENTURY OF SERVICE

BNL is a Data Driven Science Laboratory

DOE has 27 User Facilities - 6 are operated by BNL

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 opened 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:

- **QCD -** Facilities for BNL, RIKEN & US QCD communities
- **Belle II – Computing for Japanese Neutrino Experiment**

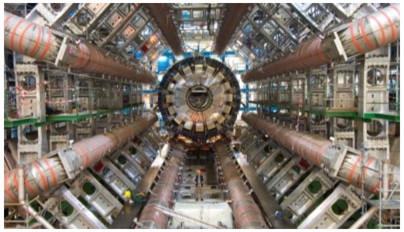
Science Today is Data Driven Discovery

RHIC

NSLS II

CFN

ATLAS



QCD

70 YEARS OF DISCOVERY
A CENTURY OF SERVICE

U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

BNL Data Statistics

- **2017 Milestone: Over 100PB**, of catalogued data archived, long term, frequent reuse
- **2nd largest scientific archive in the US**, 4th largest in the world (ECMWF, UKMO, NOAA)
- **2016 - 400PB analyzed, 2017 -expected 500PB**
- **2016 37 PB exported**
- **Computing** – HPE Broadwell CPU + K80 and P100 GPUs (will reach 200 nodes total this year), 144 node Intel Knights Landing, in the process of being purchased new Skylake Cluster, also IBM Blue Gene Q, HTC computing for RHIC and Atlas, BNL Cloud

U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

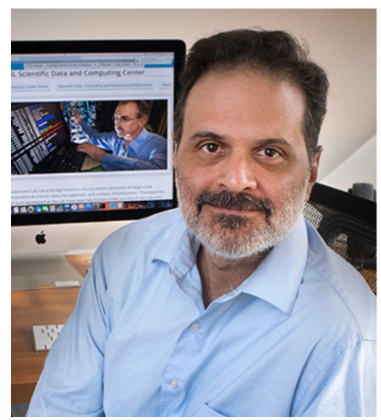
70 YEARS OF DISCOVERY

A CENTURY OF SERVICE

Computational Science Initiative

- Established in 2015
- An umbrella to bring together computing and data expertise across the lab
- Aims to foster cross-disciplinary collaborations in areas of computational sciences, applied math, computer science and data analytics.
- Aims to drive developments in programming models, advanced algorithms and novel computer architectures to advance scientific computing and data analysis.

Director:
Kerstin Kleese van Dam



Deputy Director:
Francis Alexander

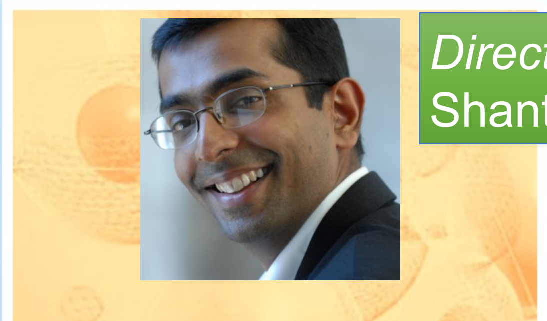
Computational Science Initiative

Computer Science & Mathematics

Director:
Barbara Chapman

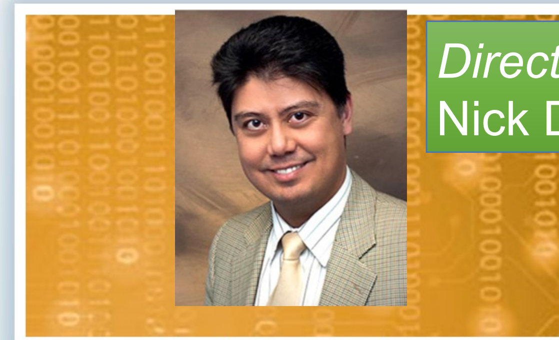
Scientific Data & Computing Center

Director:
Eric Lancon



Center for Data-Driven Discovery

Director:
Shantenu Jha



Computational Science Lab

Director:
Nick D'Imperio

Director:
Adolfy Hoisie

Computing for National Security

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

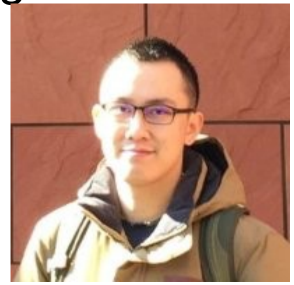
CSI Support for and Collaboration with User Communities

- **Scientific Data and Computing Center (SDCC)** hosts the GPU-accelerated institutional cluster (part of USQCD resources), the Intel Knights Landing (KNL) cluster, new Skylake system and other computing facilities (Linux farm, storage systems etc.) to meet the BNL high-throughput computing needs.
- **Computational Science Lab** provides support for application development, porting, optimization and benchmarking.
- **Computer Science and Mathematics** department collaborates with groups to study and develop performance portable programming models and improve compilers.
- **C3D** collaborates on novel big data analysis and visualization solutions.

Example Collaborations with ECP-QCD Team

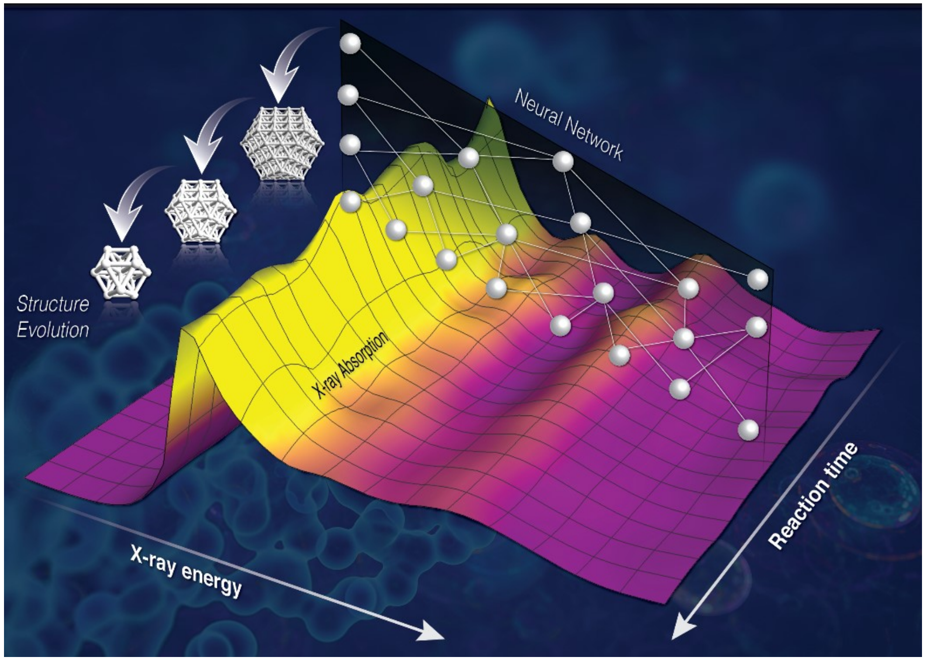
- Programming models and performance portability
  - Explore directive-based programming (OpenMP/OpenACC) for performance portability between CPUs and GPUs
  - QCD miniapp and benchmark developments
  - Collaborators: Barbara Chapman, Martin Kong, Lingda Li





- Workflow and performance profiling tools
  - Explore potential QCD use cases for performance tools such as TAU
  - Explore workflow management tools developed by CODAR
  - Collaborators: Kerstin Kleese van Dam, Line Pouchard


Solving the structure of nanoparticles by machine learning



A sketch of the new method that enables fast, "on-the-fly" determination of three-dimensional structure of nanocatalysts. The colored curves are synchrotron X-ray absorption spectra collected in real time, during catalytic reaction (in operando). The neural network (white circles and lines) converts the spectra into geometric information (such as nanoparticle sizes and shapes) and the structural models are obtained for each spectrum.

Scientific Achievement

A new method was developed to decipher the structure of nanoparticle catalysts on-the-fly from their X-ray absorption spectra.

Significance and Impact

Tracking the structure of catalysts in real working conditions is a challenge due to the paucity of experimental techniques that can measure the number of atoms and the distance between metal atoms with high precision. Accurate structural analysis at the nanoscale, now possible with the new method in the harsh conditions of high temperature and pressure, will enable new possibilities for tuning up reactivity of catalysts at the high flux and high energy resolution beamlines of NSLS-II and other advanced synchrotron facilities.

Research Details

Using methods of machine learning the previously "hidden" relationships between the features in the X-ray absorption spectra and geometry of nanocatalysts were found.

J. Timoshenko, D. Lu, Y. Lin, A. I. Frenkel,
***J. Phys. Chem. Lett.*, DOI: 10.1021/acs.jpcc.7b06270**

Event Driven Adaptive Sampling based on Manifold Learning

Scientific Achievement

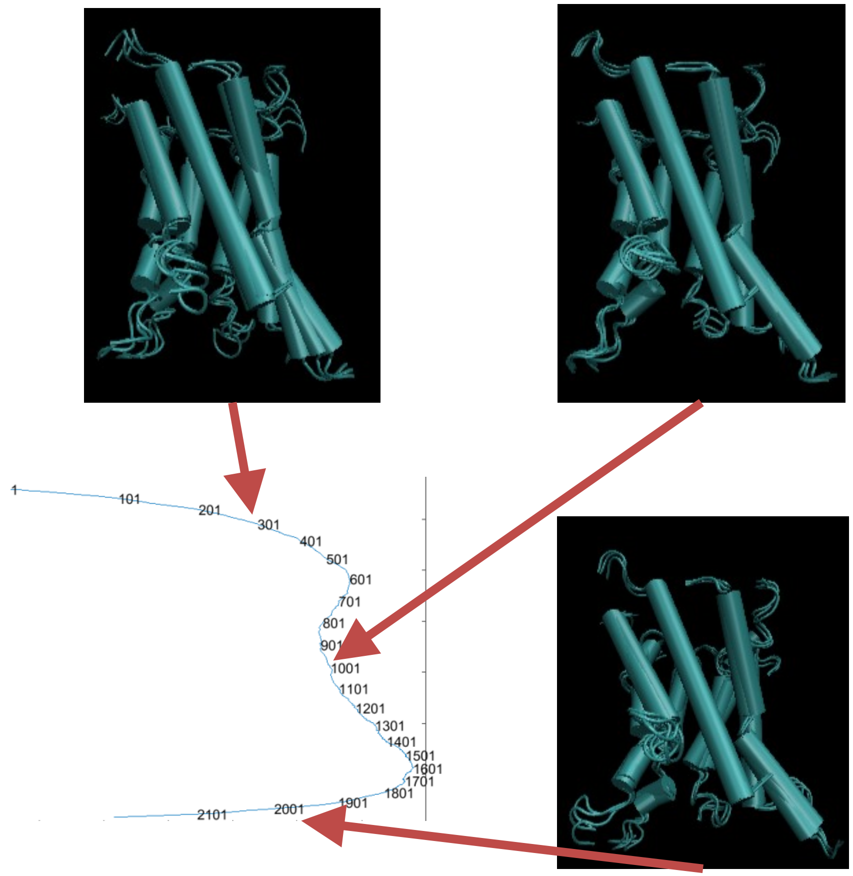
Preliminary batch adaptive sampling study on MD (Molecular Dynamics) trajectories from manifold change detection method

Significance and Impact

Since a single MD resulting trajectory would generate easily 32 PB of data, our goal is developing generic methods for online data analysis detecting relevant or significant events to preserve during the simulation or experiments.

Research Details

- Projection of MD trajectories to manifold space (dimensionality reduction) across of time into two dimensional space
- Change detection on manifold space, which is much more robust than original full coordinate space as it removed local vibrational noises
- Adaptive sampling strategy was implemented based on accumulated changes of trajectories



Low dimensional manifold projection of different state of MD trajectories

Deep Learning Classification of X-ray Scattering Data

Scientific Achievement

Modern deep learning methods were applied to the automated recognition and classification of x-ray scattering data

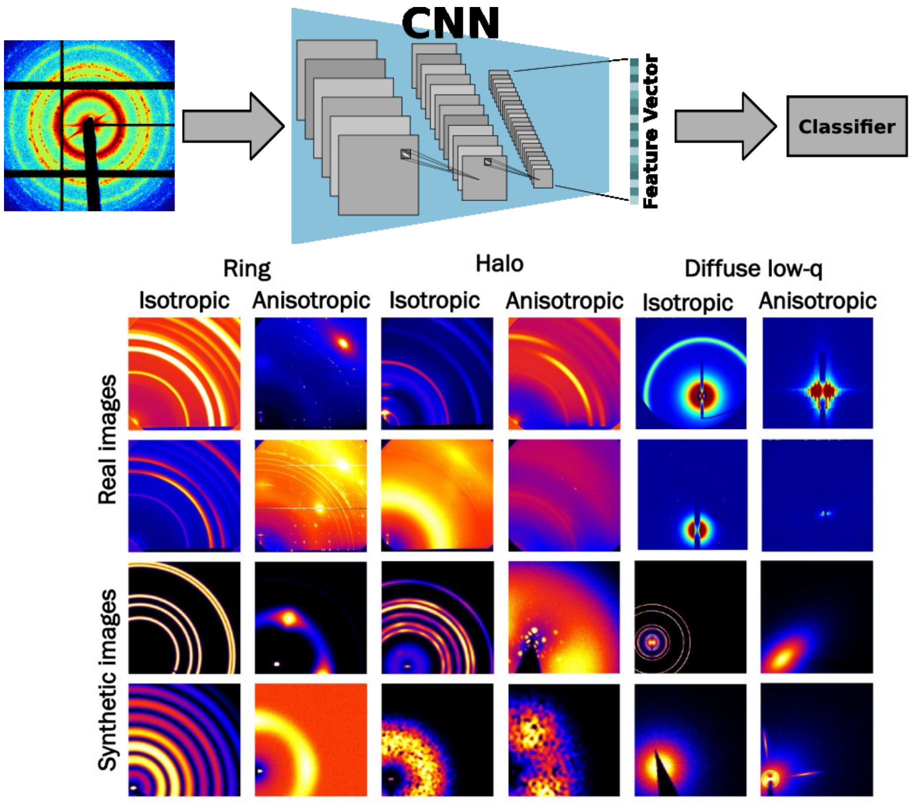
Significance and Impact

Experimental scientific data can be automatically and accurately classified, allowing autonomous experimental decision-making

Research Details

- Modern x-ray synchrotrons generate data at a massive rate; automated data classification is critically required to deal with data-rates
- Modern deep learning methods were used to recognize and classify x-ray scattering data
- Neural network training was augmented by inputting a large corpus of synthetic images; because the experimental physics are understood, these images are highly realistic and allow the network to implicitly learn the underlying physics

Boyu Wang, et al. *New York Scientific Data Summit*, 1 (2016)
Boyu Wang, et al. *Applications of Computer Vision (WACV)*, 1, 697 (2017).
Jiliang Liu, et al., *IUCrJ* 4, 455 (2017).



Deep learning for scientific data: (Top) A convolutional neural network is used to convert input x-ray scattering images into a concise descriptor that can be used for classification. (Bottom) X-ray scattering images are complex and diverse. Owing to the limited amount of *tagged* experimental data, we generate synthetic highly-realistic synthetic data to train the CNN.

Hackathons and Trainings

- Parallel Programming 101 – provided on demand
- Feb 14/15 Performance Analysis and Modeling Hands on Workshop
- Feb 28 – March 1st – KNL Hackathon for application teams
- July – GPU Hackathon for application teams
- August – New York Scientific Data Summit
- August – ModSim – Performance Modeling and Simulation Workshop
- Planned Machine Learning Workshop

- New Training space finished in Bldg 725

Acceleration of Radar Simulator Code for Cloud Research

Scientific Achievement

Accelerated Cloud resolving model radar simulator code (CR-SIM) from 18 hours to 6 minute execution time.

Significance and Impact

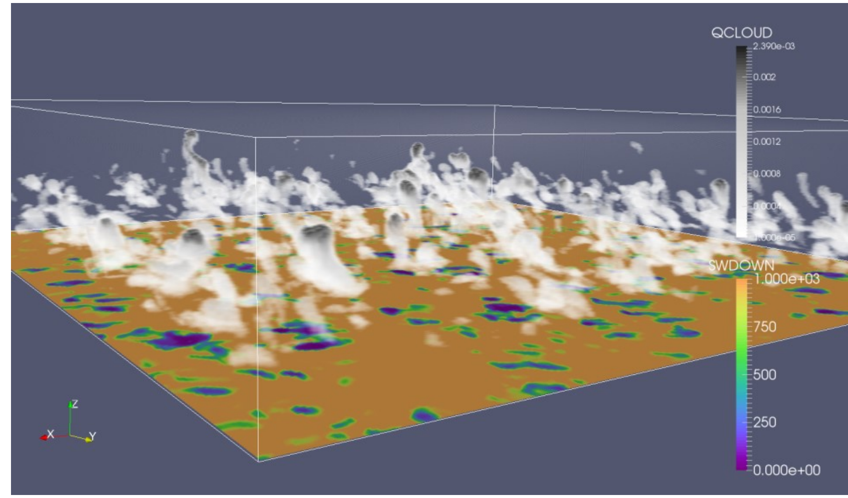
Optimization allows now for effective comparison of CRM and LES models to real observations.

Research Details

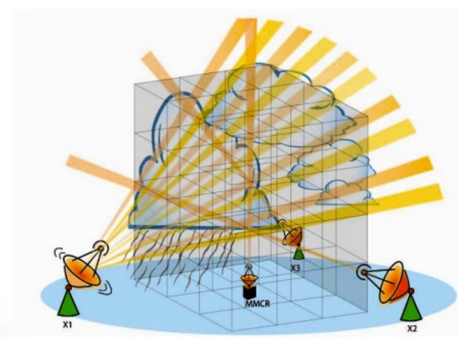
Creates Virtual Cloud Observations accounting for various sensor limitations

Using Data Intensive Programming Model Features for code Optimization

Users: BER ARM LASSO to test Large Eddy Simulations (LES), Climate Model Development and Validation (CMDV) teams. Now also international interest.



LES is commonly used to simulate clouds and the planetary boundary layer (lowest part of atmosphere). Shown here are the cloud water content (QCLOUD) and the resulting shadows that impact the sunlight reaching the ground. Credit: ARM Climate Research Facility.



U.S. DEPARTMENT OF ENERGY

BROOKHAVEN NATIONAL LABORATORY

70 YEARS OF DISCOVERY
A CENTURY OF SERVICE


BNL
Brookhaven
National Laboratory
OPERATED BY BROOKHAVEN SCIENCE ASSOCIATES UNDER CONTRACT WITH THE UNITED STATES DEPARTMENT OF ENERGY

Thank You!