Los Alamos
NATIONAL LABORATORY

Real-time Computational Workflows for Experiments at Particle Accelerators

Christine Sweeney, LANL

August, 9, 2022

LA-UR-22-28184

Talk Outline

- Data Management
  - Dynamic Diamond Anvil Cell (dDAC)
  - Shock compression
- High Performance Computing
  - Single Particle Imaging
- Control
  - Guided user experiments

LANL in spring

Data and Computational Workflow Challenges for Experiments at Accelerators

- Interdisciplinary effort
- Real-time component
- Connection to resources
- Diversity of applications
- Large data volumes and velocity, in some cases
- Portability
- Sustainability

Tools for Real-time Dynamic Diamond Anvil Cell Experiment Data Analysis

Data Science Thrust Area within:

Novel in situ Probes of Mesoscale Materials Dynamics, 2019-2021

LDRD Directed Research Project, PI: Dmitry Yarotski, Co-PI: Blake Sturtevant


Christopher Biwer


Andres Quan

Larissa Huston


John Lazarz


Ye Jin Choi


Dmitry Yarotski

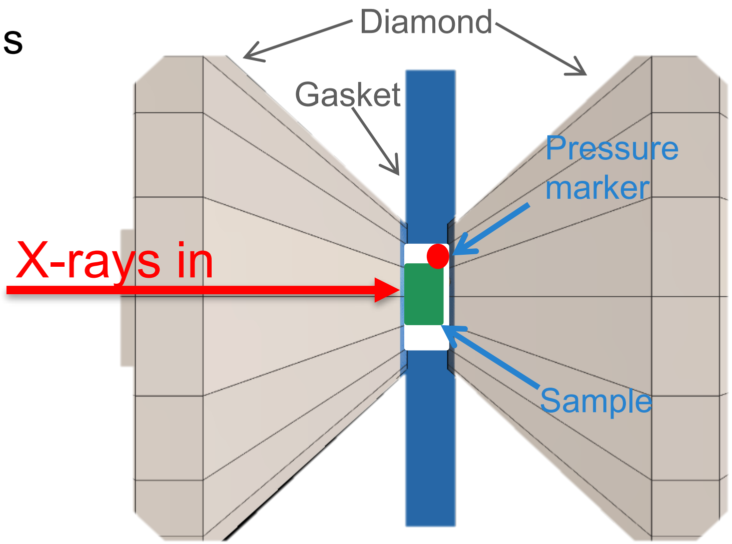

Blake Sturtevant


Christine Sweeney

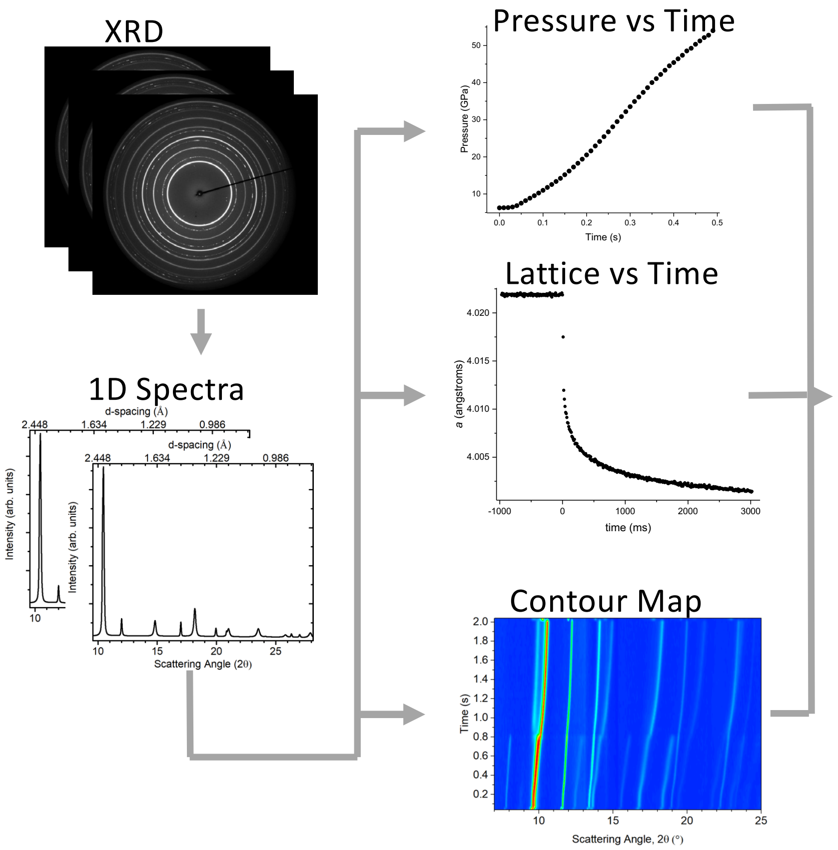
dDAC Analytics Requirements

- Use at HPCAT at APS, PETRA-III at DESY and ultimately at the EuXFEL
- Typically acquire 30-1000 images with 2kHz detector
- Could go to 4.5 MHz at EuXFEL
- Integrated tools, no time for using separate tools
- Help drive experimental design

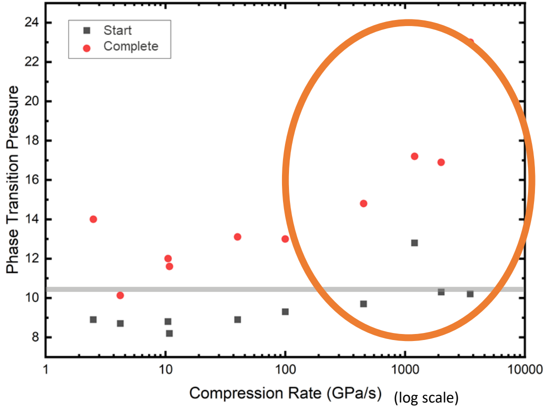
In situ X-ray Diffraction allows for time-resolved material phase and density determination in novel DAC.



Motivation/Background – Analyzing Experimental Data Quickly Enables Collection of Better Quality Data



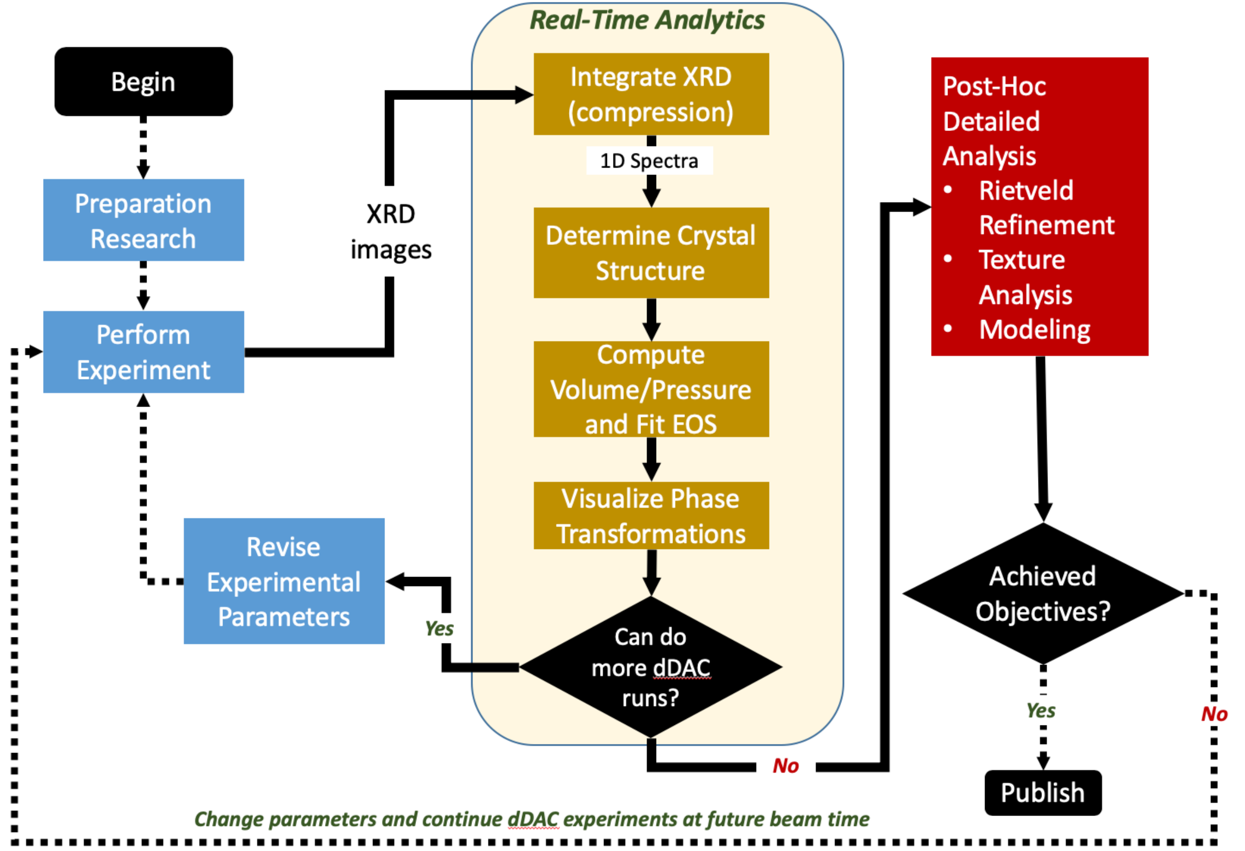
Phase transition pressures (start and end) for different strain rate compressions.



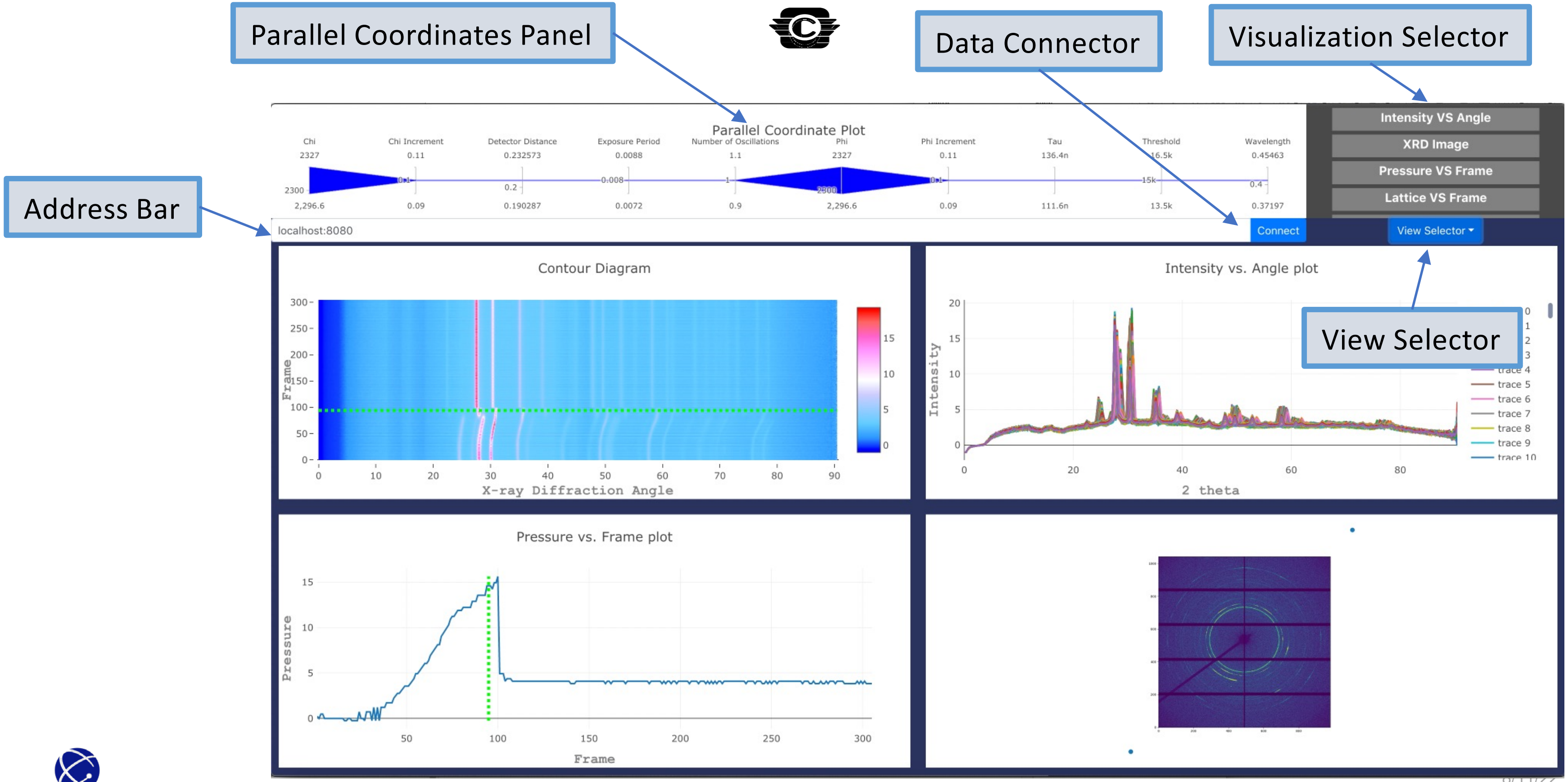
Ti phase transition. Courtesy B. Sturtevant

Should have collected more data in the gold area, which has more interesting results.

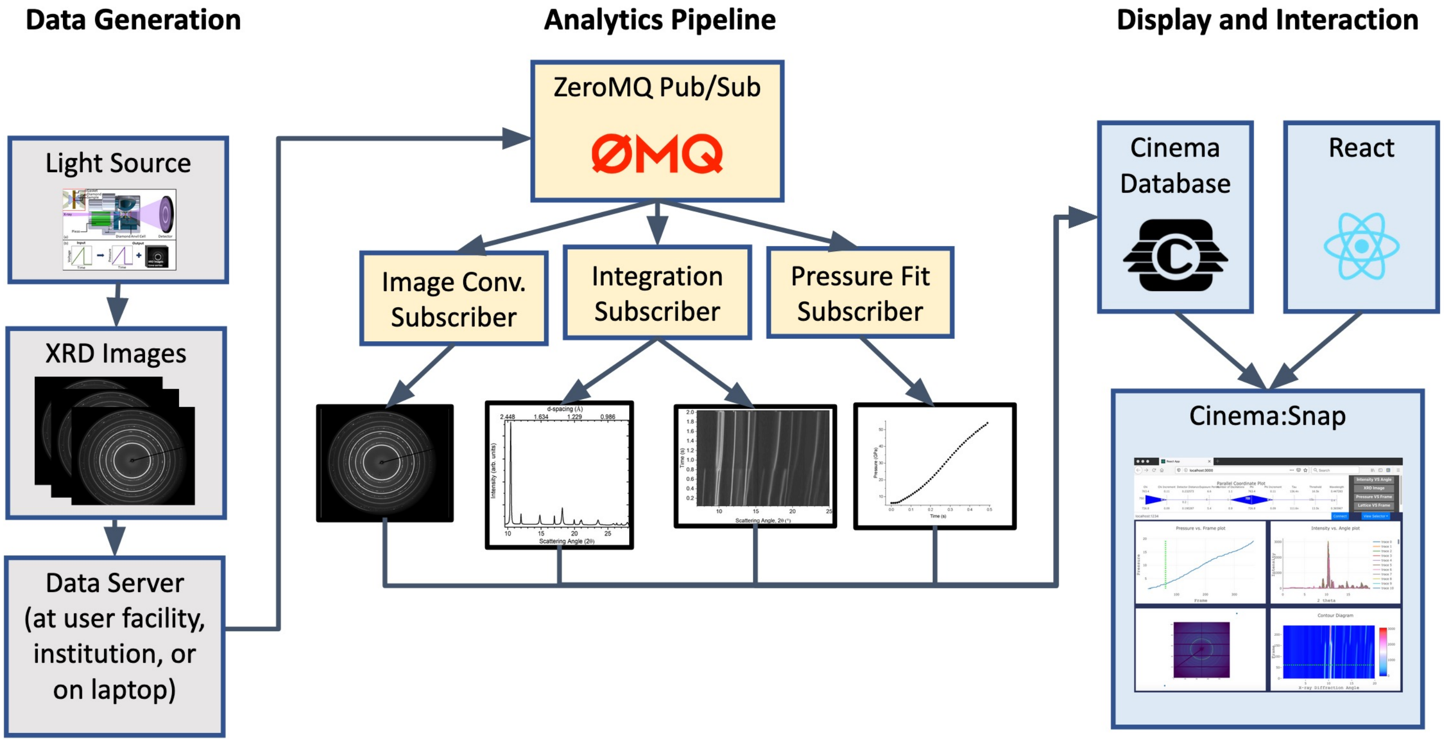
Revised Analytics Workflow



Cinema:Snap Layout



Cinema:Snap and Automated Workflow Benefits



- Open source: https://github.com/cinemascience/cinema_snap
- Open source (future): https://github.com/lanl/analytics_pipeline

Dynamic Compression Experiment Data Management and Real-time Data Analysis

Workflow and Visualization Staff for:
Real-time Adaptive Acceleration of Dynamic Experimental Science, 2016-2019
ASSIST Project
LDRD Directed Research Project, PI: James Ahrens, Co-PI: Cynthia Bolme


Dan Orban


Divya Banesh


Christopher Biwer


Ayan Biswas


Christine Sweeney


Richard Sandberg

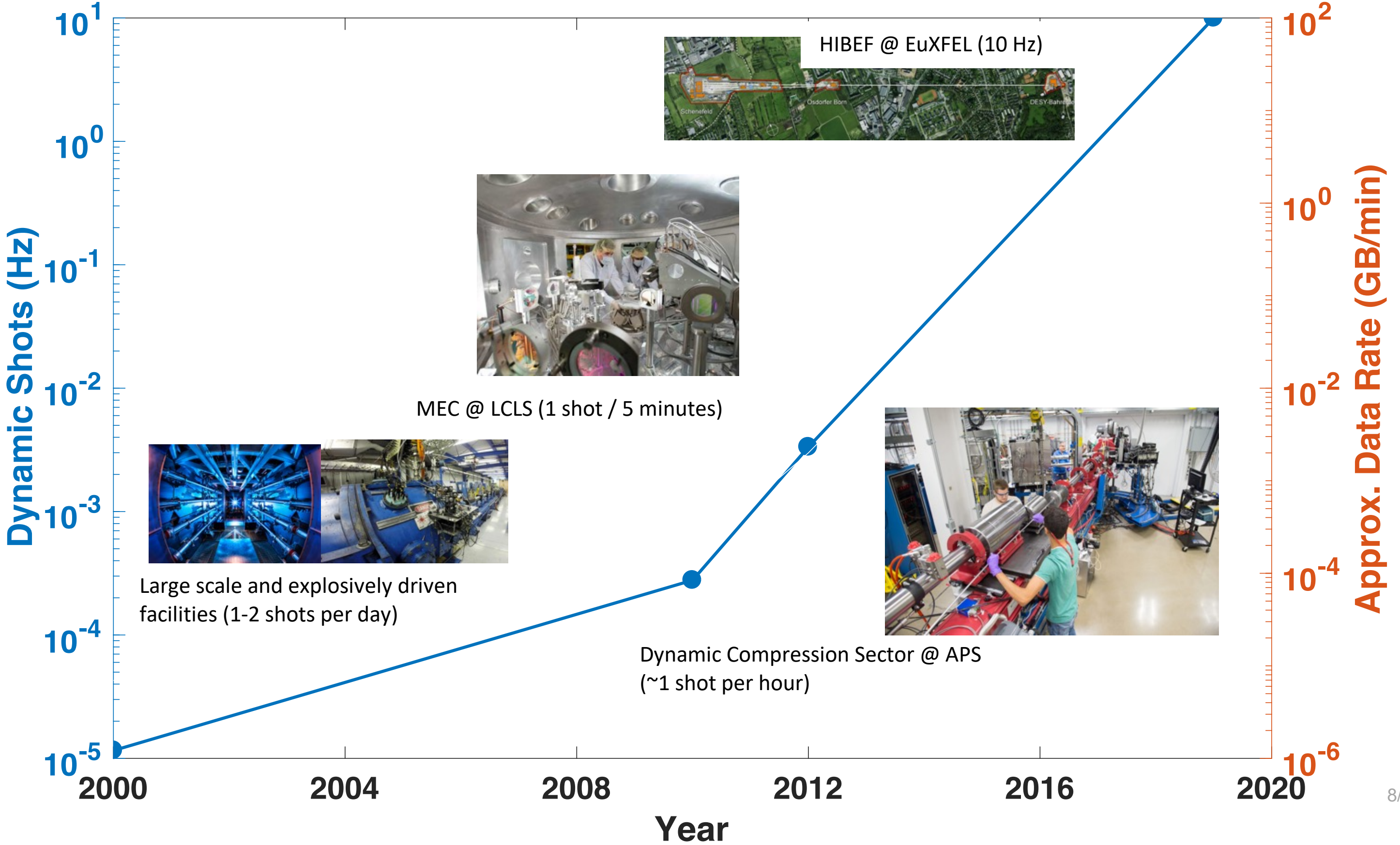

Cynthia Bolme


James Ahrens

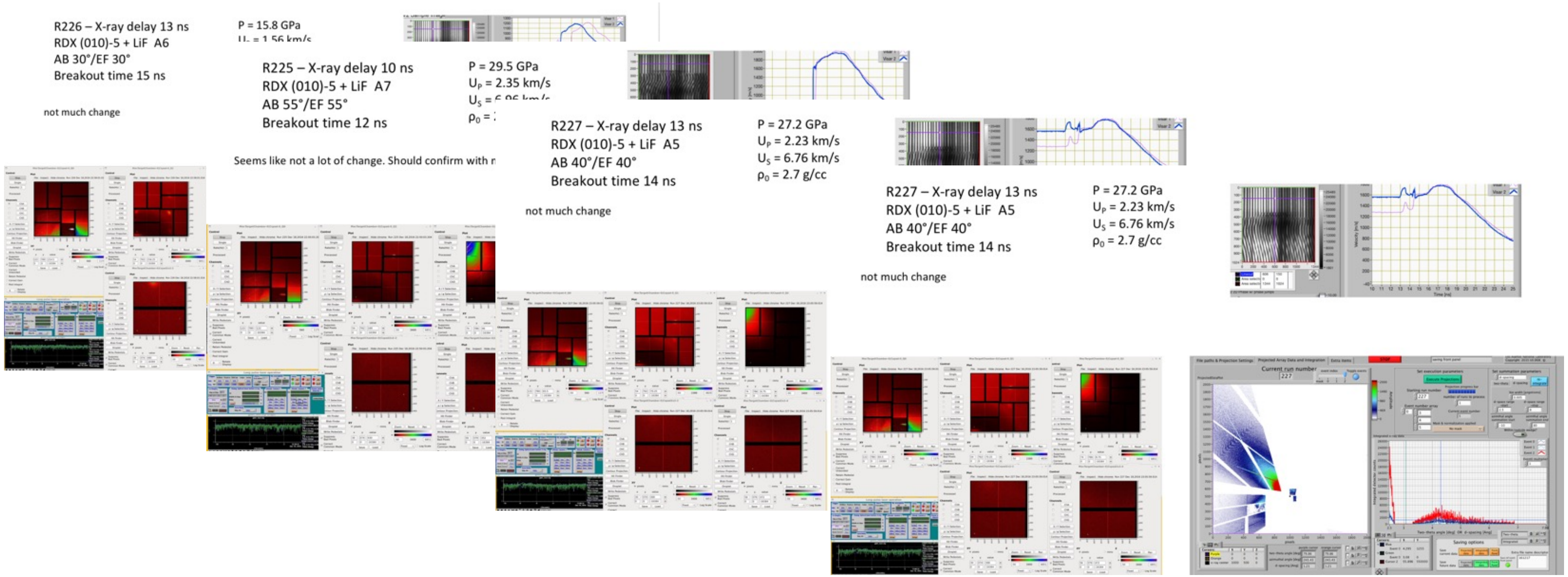

David Rogers

Not pictured: C. Tauxe, R. Saavedra

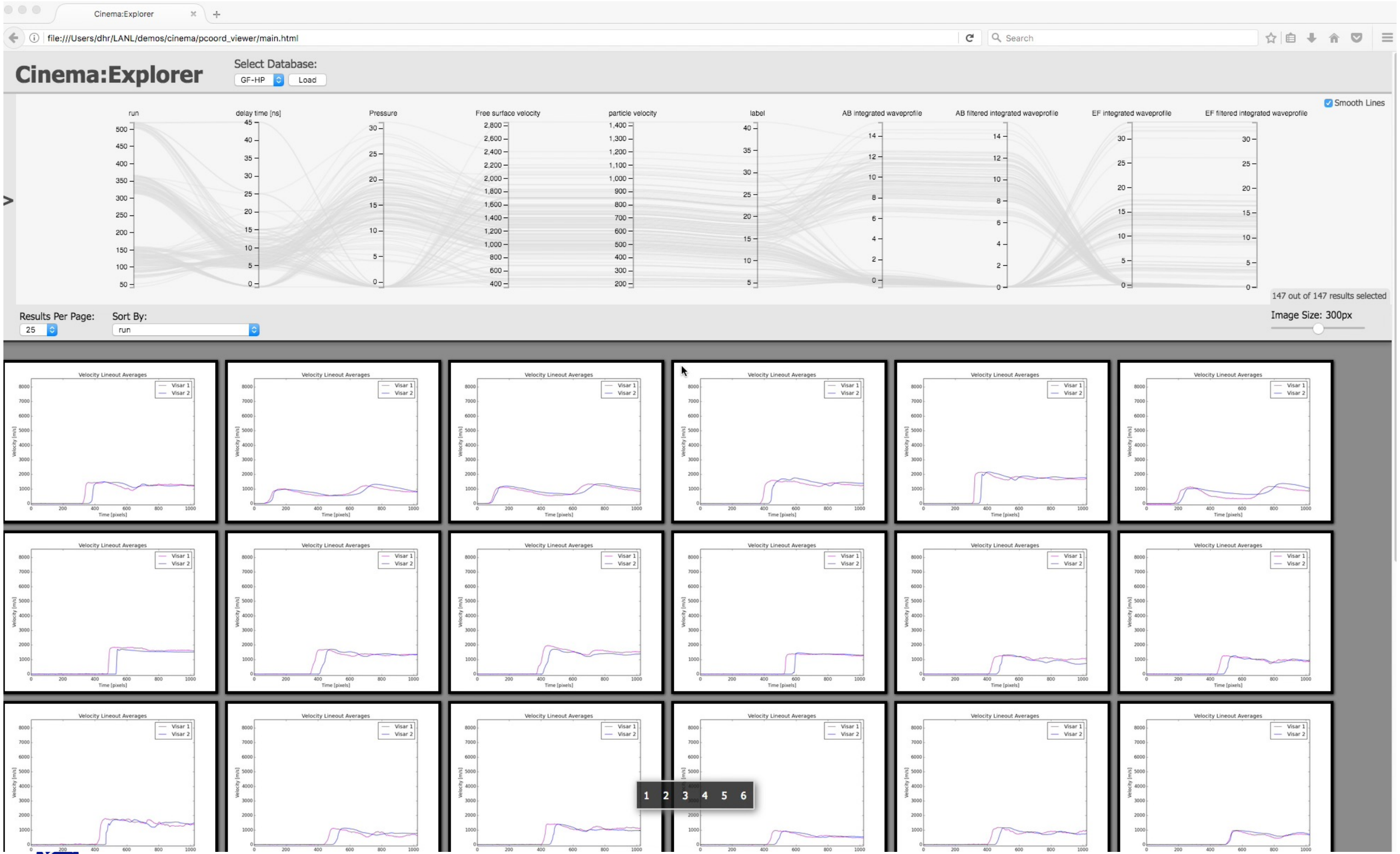
Big Data Problem at Light Sources, Especially in Dynamic Compression/HEDP Science



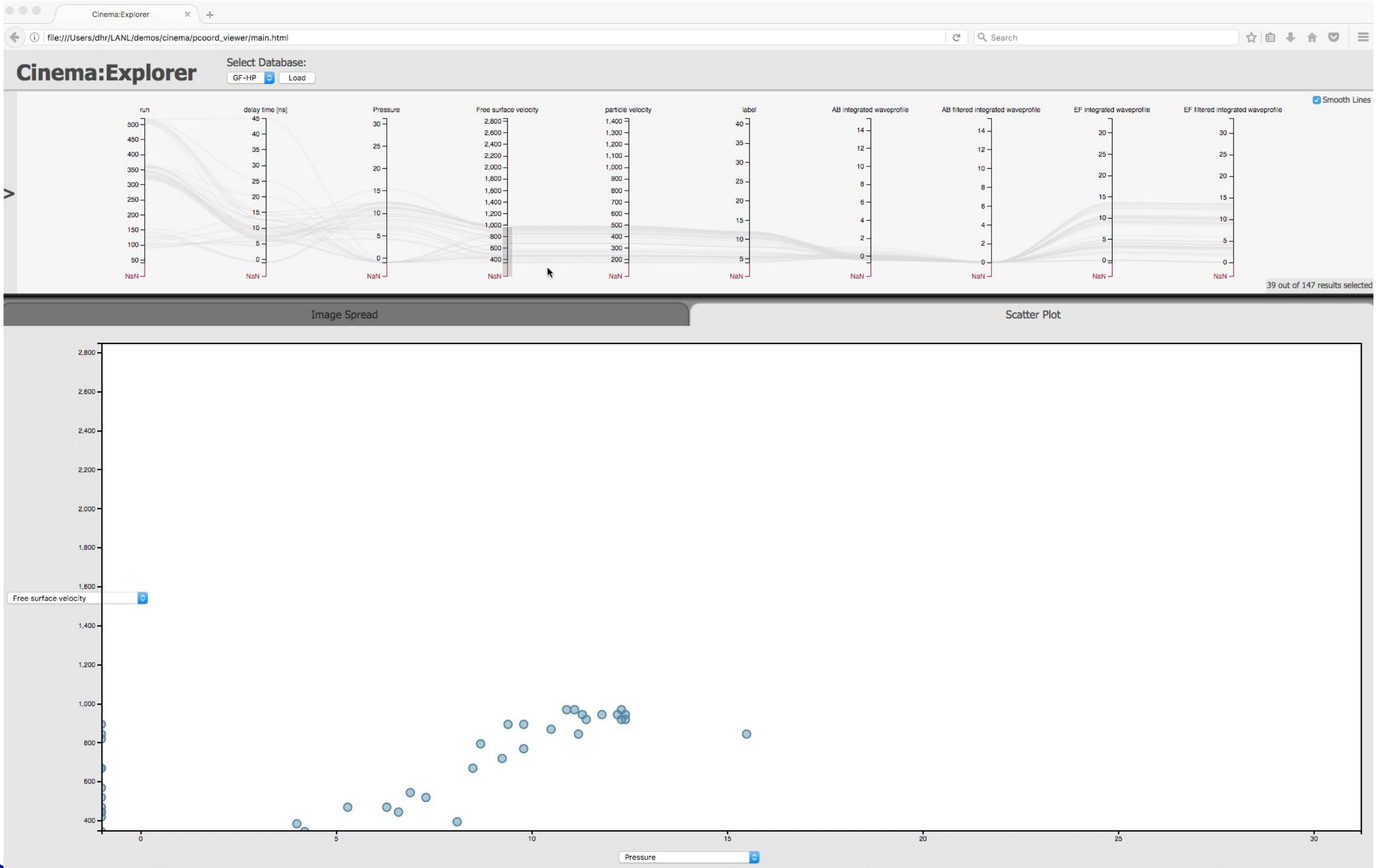
Previous State of the art for data management - PPTX



Repeat for each run, one slide per shot, up to several hundred shots over an experiment (500+)

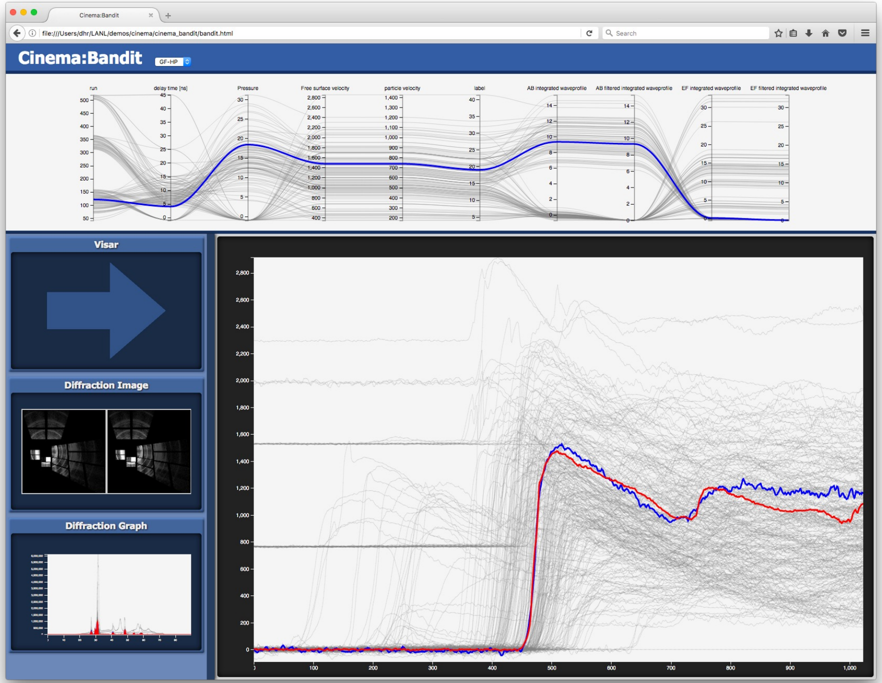


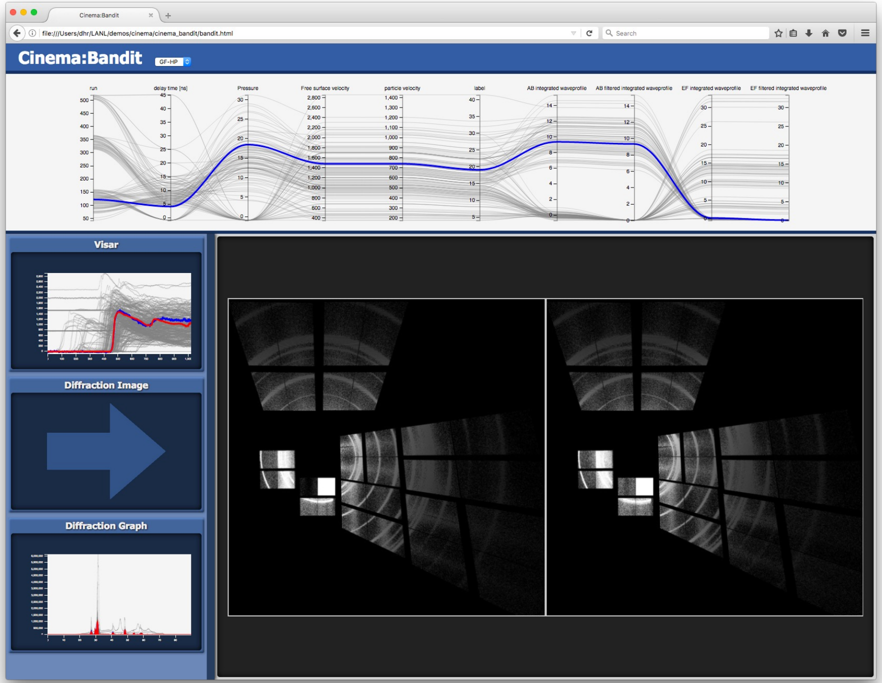
Slide courtesy
David Rogers



Slide courtesy
David Rogers

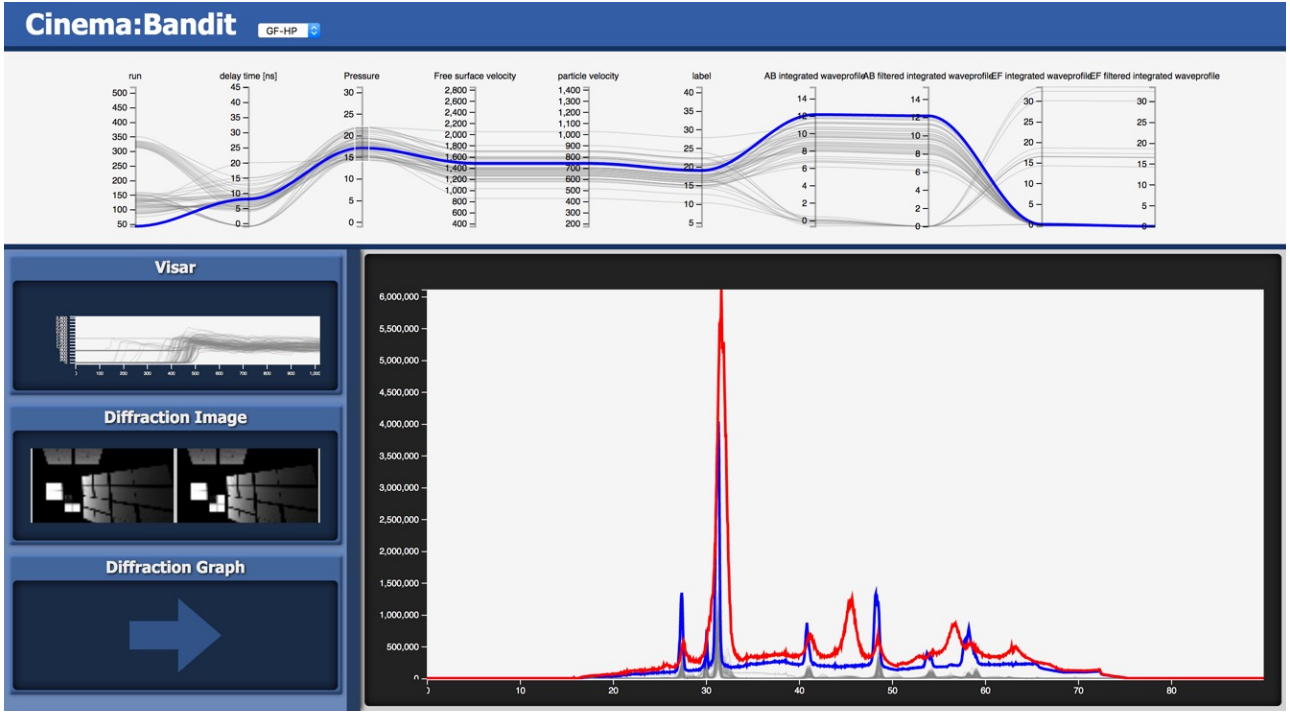
Cinema:Bandit multi-data viewer





Cinema-Bandit for Database Visualization and Curation

Real time display of data populating with runs in real time

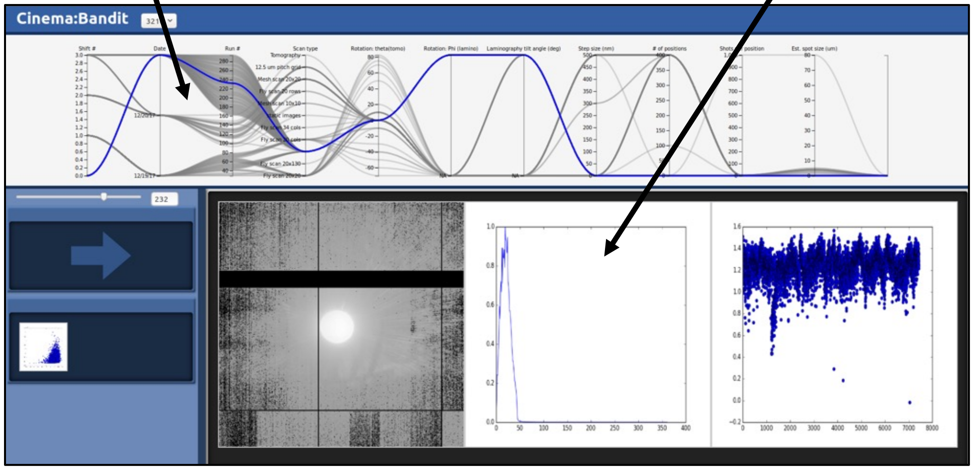


Example dataset from MEC-LCLS experiment on shocked high-pressure phases of titanium (C. Bolme PI)

Application of Tools to Beam Line Monitoring Data

Example from December 2018 LCLS Imaging Beamtime

Cinema
Database

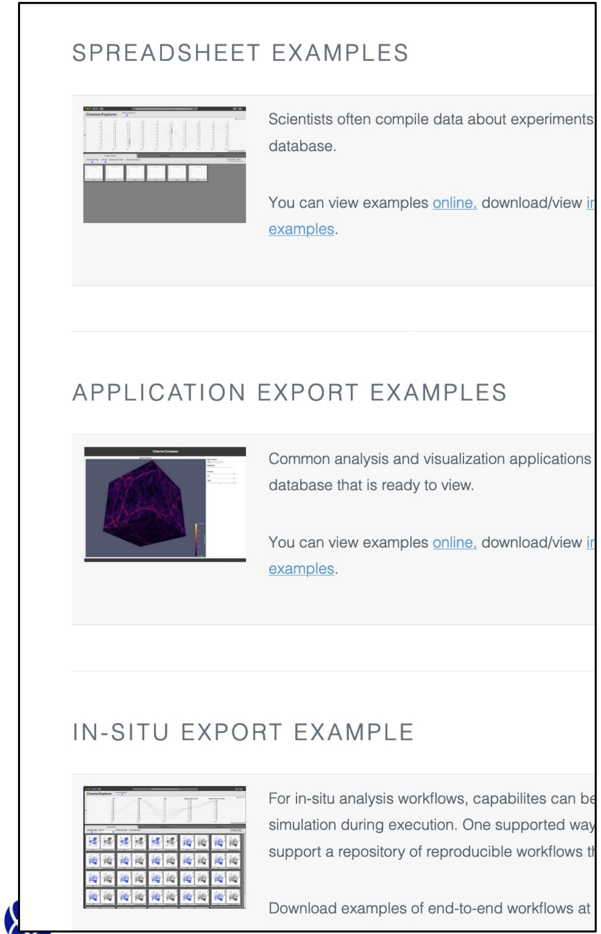


Cinema
Bandit

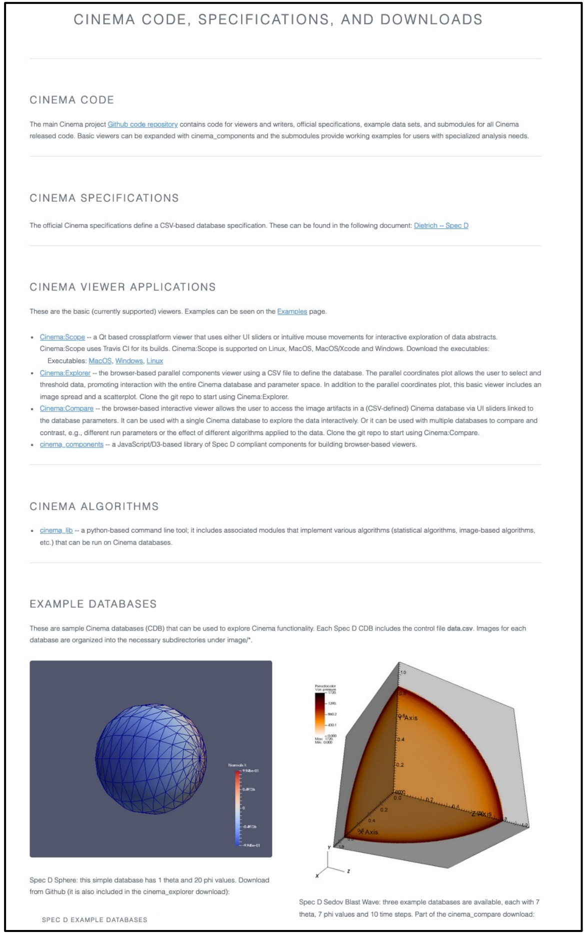
ASSIST team members present: R. Sandberg, C. Sweeney, D. Banesh, D. Orban

Cinema Released Open Source

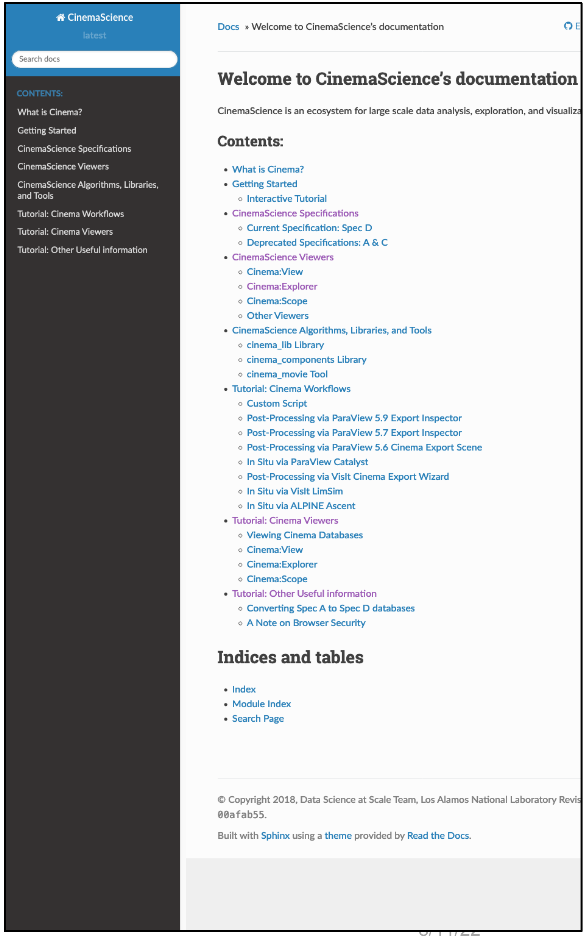
cinemascience.github.io



Examples



Downloads



Documentation

High-Performance Single-Particle Imaging Reconstruction on Pre-Exascale Computing Platforms

ExaFEL: Data Analytics at the Exascale for Free Electron Lasers, 2017 - present

SLAC National Accelerator Laboratory

ExaFEL PI: Amedeo Perazzo

Hsing-Yin Chang

Antoine Dujardin

Seema Mirchandaney

Ariana Peck

Elliott Slaughter

Monarin Uervirojnangkoorn

Chun Hong Yoon

Lawrence Berkeley National Laboratory

Johannes Blaschke

Jeffrey Donatelli

Petrus H. Zwart

Los Alamos National Laboratory

Pranay Kommera

Vinay Ramakrishnaiah

Christine Sweeney

ExaFEL Resource Orchestration Team:

Lawrence Berkeley National Laboratory

ESnet:

Chin Guok

Thomas Lehman

Alexander Sim

NERSC:

Deborah Bard

Damian Hazen

Ashwin Prabhu Selvarajan

SLAC National Accelerator Laboratory

Mark Foster

Wilko Kroeger

Amedeo Perazzo

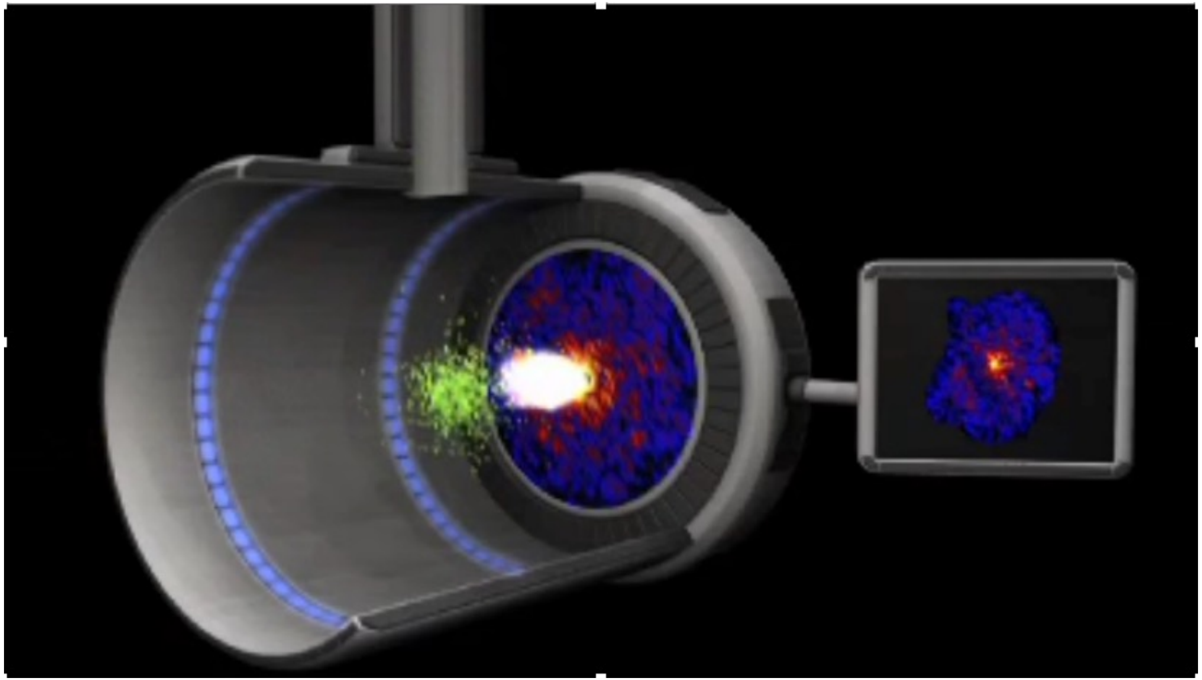
Frederic Poitevin

Murali Shankar

Cong Wang

This research was supported by the Exascale Computing Project (17-SC-20-SC), a joint project of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration, responsible for delivering a capable exascale ecosystem, including software, applications, and hardware technology, to support the nation's exascale computing imperative.

Single-Particle Imaging is Performed via High Repetition Rate X-ray Free Electron Lasers (XFELs)

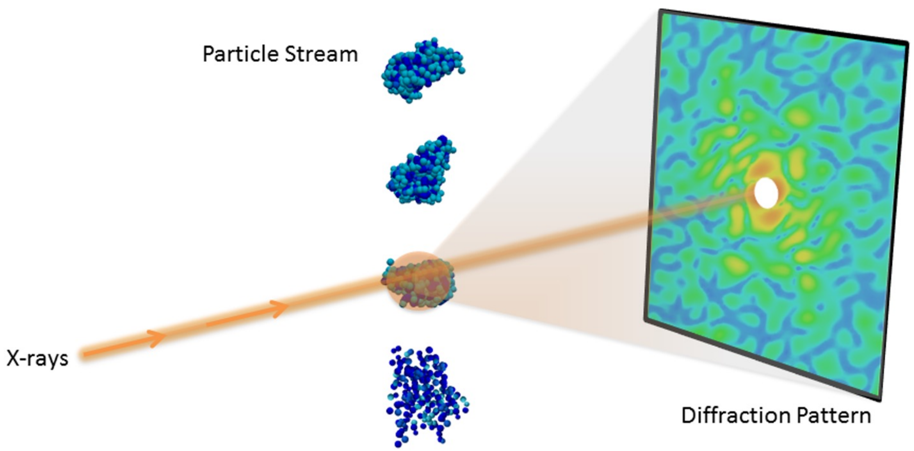


Single-particles such as ribosomes or proteins are placed in front of the x-ray beam and diffraction patterns are collected on a detector behind the sample

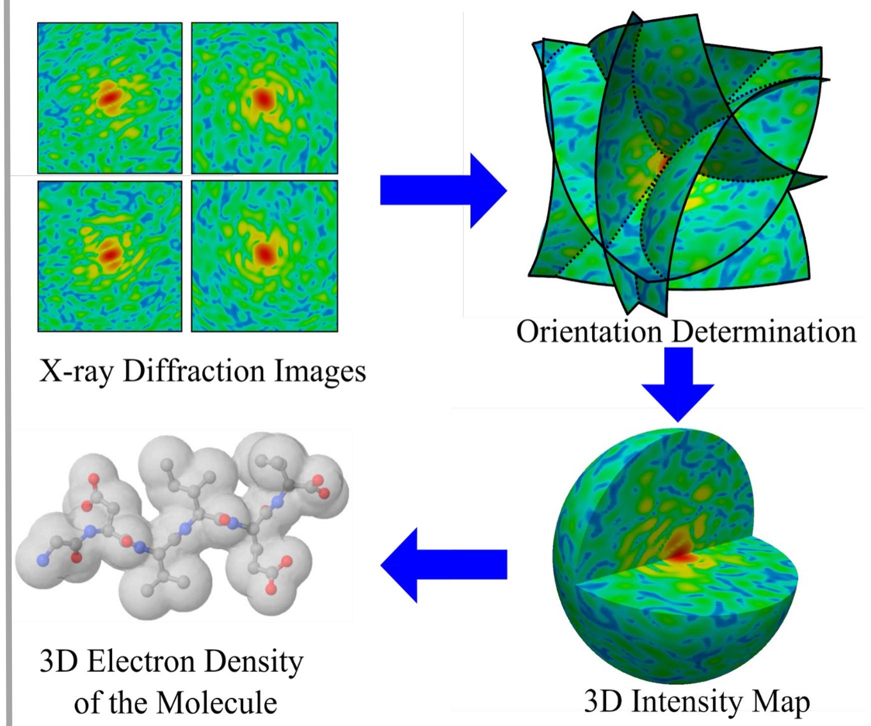
- **Ultrafast X-ray pulses** are used like flashes from a high-speed strobe light, yielding data that can produce stop-action movies of atoms and molecules.
- These experiments are performed at X-ray Free Electron Laser (XFEL) facilities like LCLS-II at SLAC National Accelerator Laboratory.

Single-particle Imaging Data Requires Image Reconstruction

Image Collection



3D Reconstruction

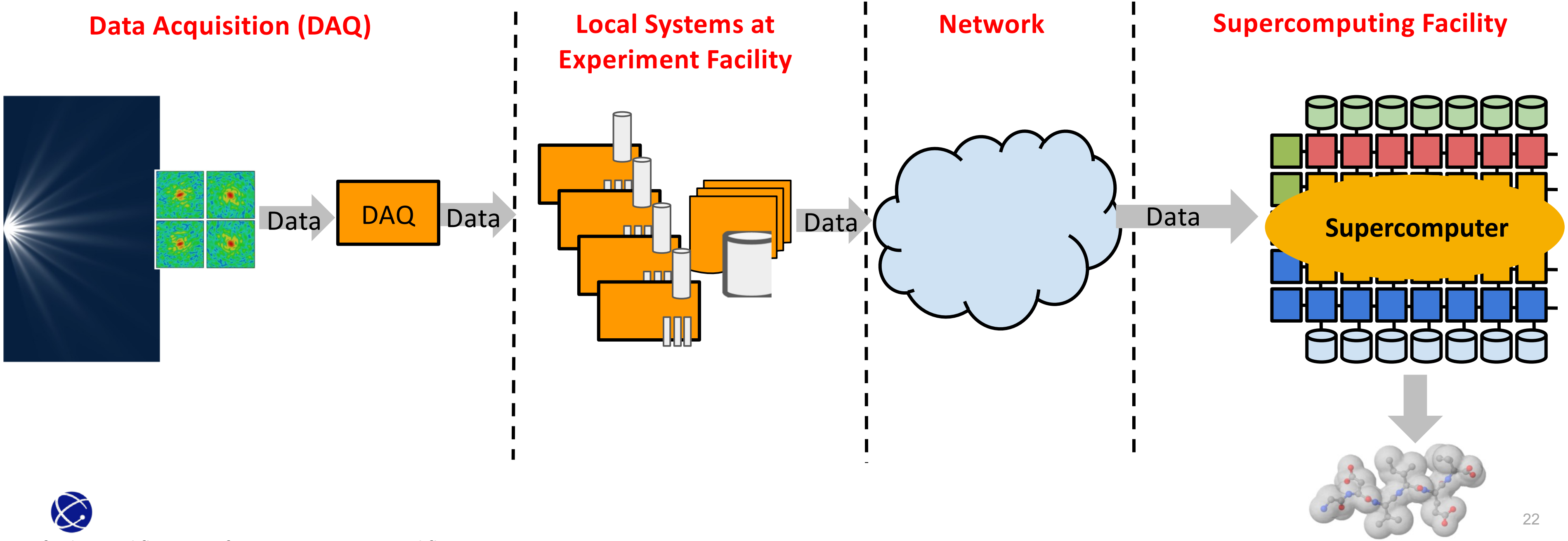


ExaFEL:
Data Analytics at the Exascale for Free Electron Lasers

Application Project within Exascale Computing Project (ECP)

Two Thrusts: Serial Femtosecond Crystallography and Single-Particle Imaging

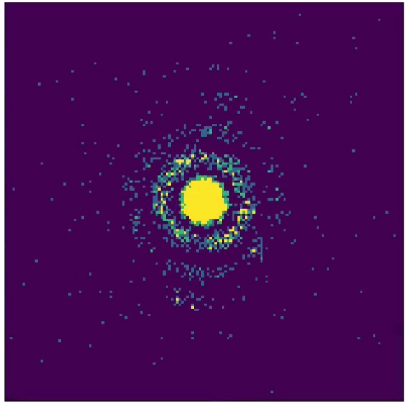
PI: Amedeo Perazzo, SLAC. Co-PI: Nicholas Sauter, LBNL



Interfacility workflow icons from ExaFEL project workflow

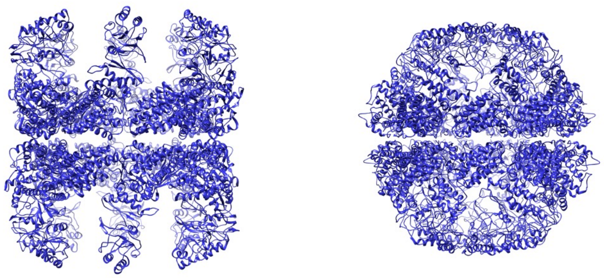
ExaFEL Single-Particle Imaging Workflow Vision

- Produce detailed **3D protein structure and a movie** of the protein functioning at room temperature.
- Demonstrate this exascale capability at **5kHz rate of 20 minutes data collection, totalling 6 million snapshots.**
  - On simulated realistic data and real data previously collected.
- (Stretch goal) Include additional complexity from **conformational heterogeneity** (obtained when molecule is excited through an optical laser, or in equilibrium experiments) in the analysis.
- Run multiple 3D electron density reconstruction instances for each conformation on each dataset and choose the best.



Simulated diffraction pattern

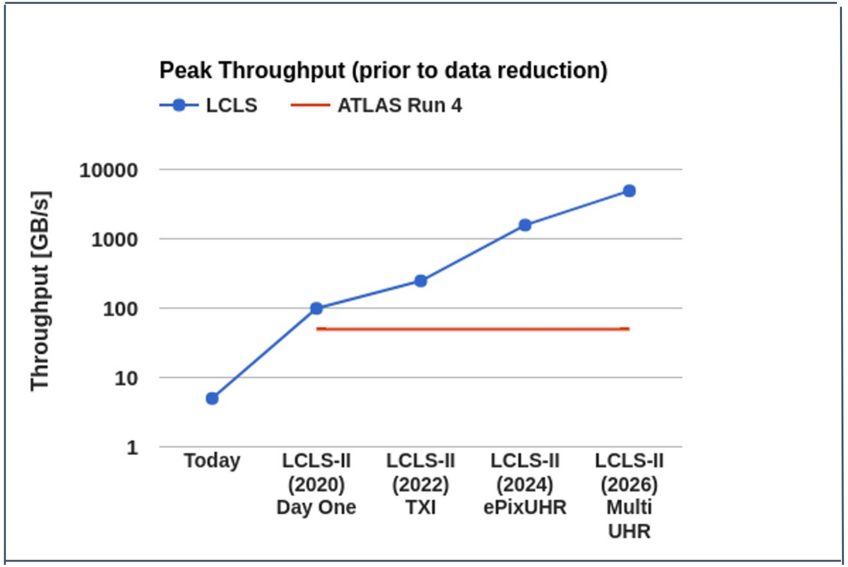
Conformational heterogeneity



Challenging Characteristics of LCLS Data Regime

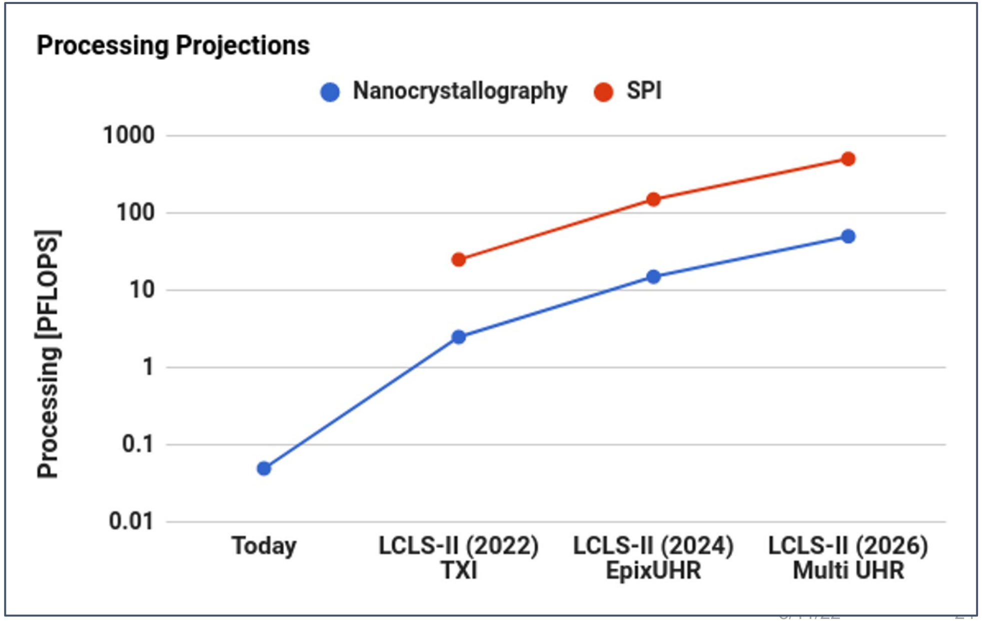
Example data rate for LCLS-II (early science)

- 1 x 4 Mpixel detector @ 5 kHz = **40 GB/s**



Example LCLS-II and LCLS-II-HE (mature facility)

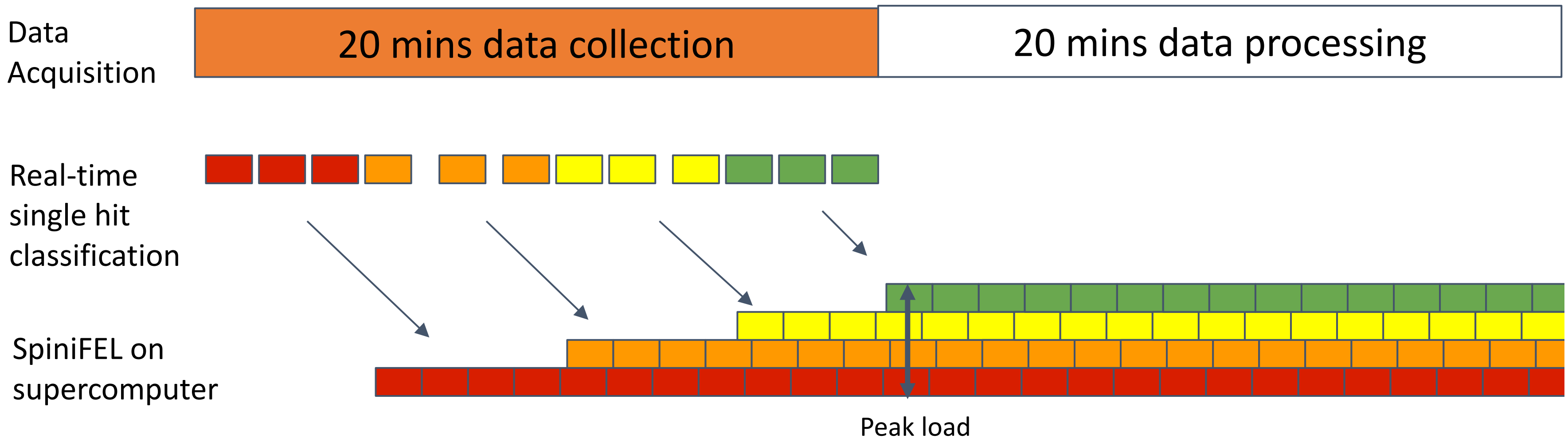
- 2 planes x 8 Mpixel ePixUHR @ 50 kHz = **1.6 TB/s**



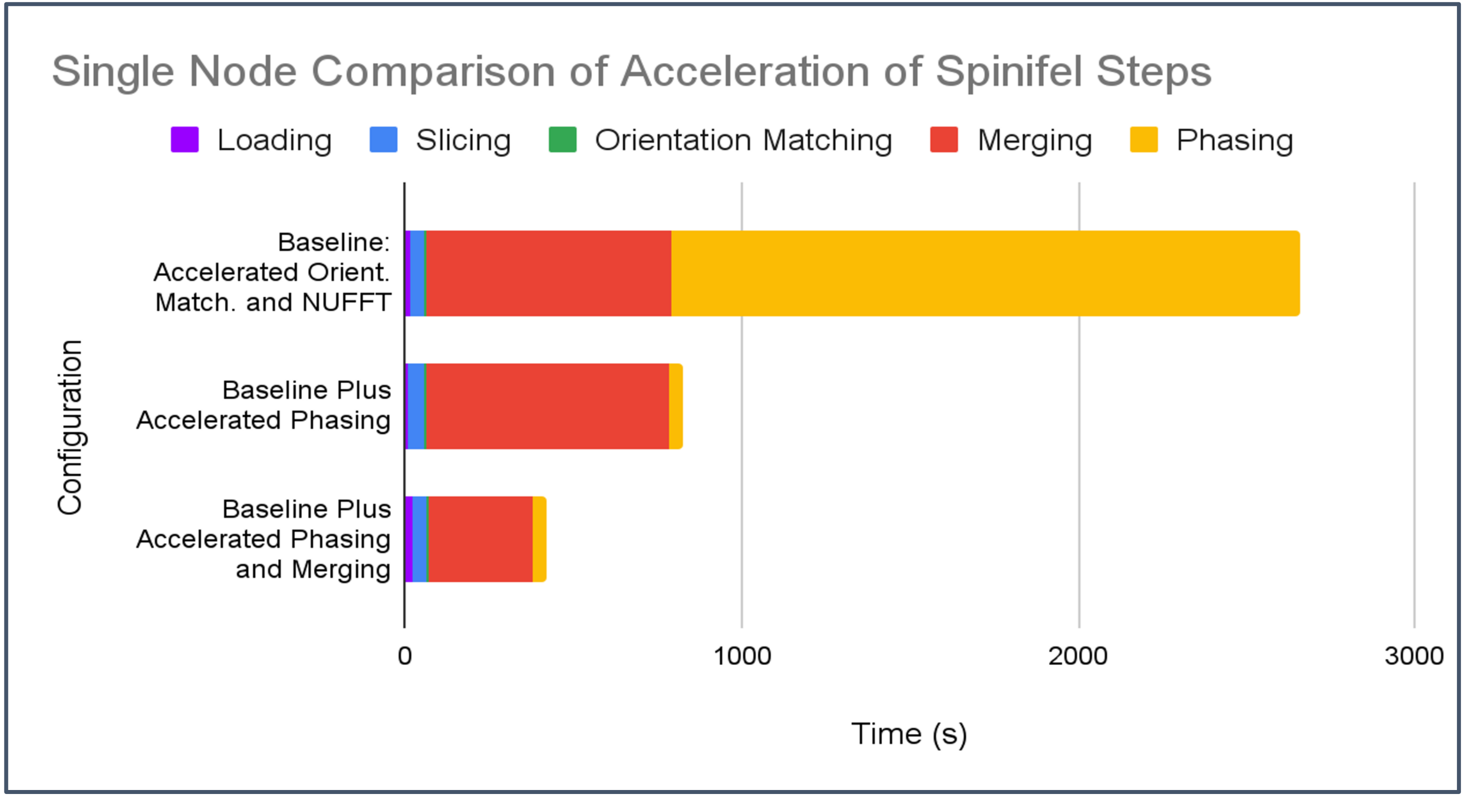
ExaFEL Puts Computational Load on Supercomputers

Goal: Ingest 5kHz data for ~20 minutes (typical run length), totalling 6M images

Scenario	# Nodes	Protein	Resolution	# Orient	# Images	# Conformations
Low-end	3,300	3IYF	14 Angstroms	20k	198k	30
High-end	10,000	2CEX	4 Angstroms	60k	12k	500



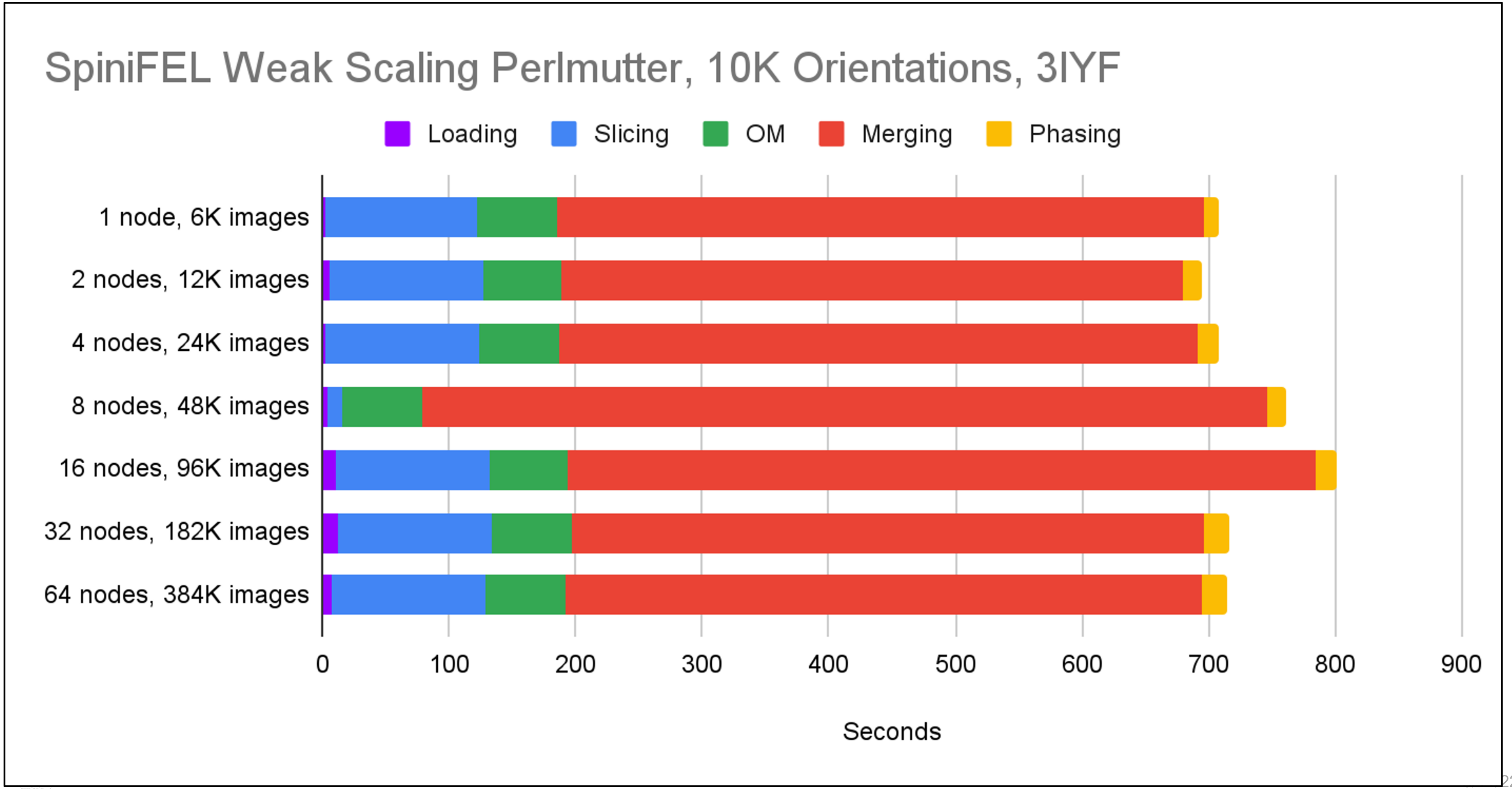
GPU Acceleration Shortens SpiniFEL Run Time



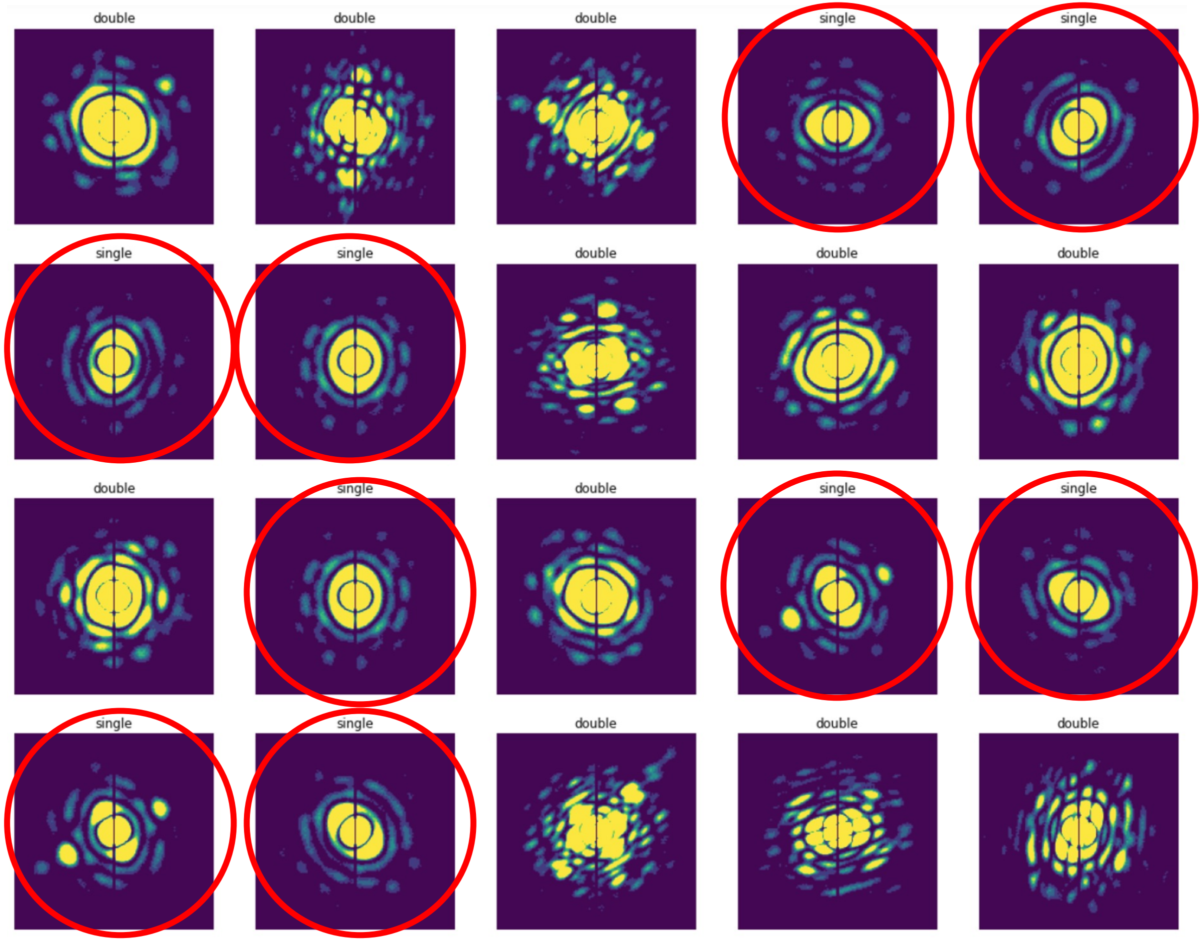
Parameters used:

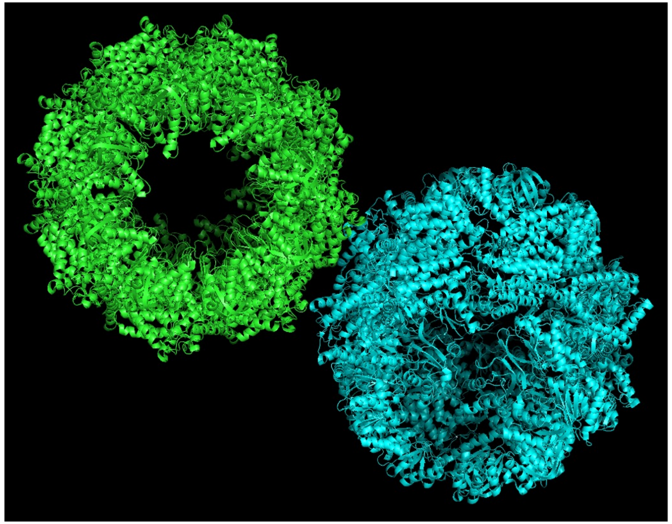
- 1000 images/rank, 2000 orientations, 6 ranks/node, 7 CPUs/rank, 6 GPU/node on Summit
- Dataset: 2CEX, 128 x 128 pixels
- 10 generations (main loop iterations)

Image Scalability Crucial for Meeting Real-time Constraint



AI-based Classifier Accurately Classifies Images





Multiple particles in the x-ray beam need to be excluded from SpiniFEL reconstruction.

Achieves **95% accuracy**, f1 score.

ExaFEL Automates Data Path and Workflow

- **Resource orchestration** streams and distributes data, and reduces job startup time.
- Software-defined network (SDN) allows selection of uncongested ESnet paths.
- Results in **1.5-2x increase in bandwidth throughput for the orchestrated dedicated path versus the path set up using normal routing protocols.**
- Now automating resource coordination, data flow, and analytics pipeline.

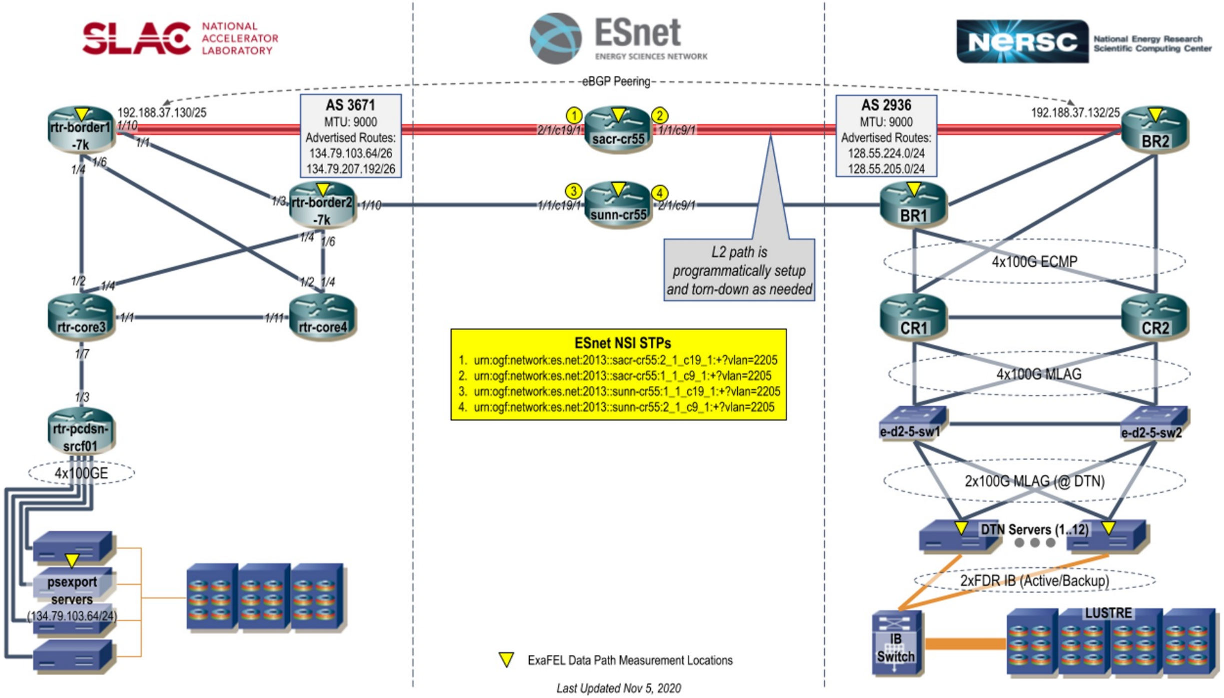


Image courtesy Chin Guok (ESnet, Lawrence Berkeley

Machine Learning for Control

ExaLearn Control Pillar

ExaLearn Co-Design Center for Exascale Machine Learning Technologies, 9/18 –

Project within Exascale Computing Project, ExaLearn PI: Frank Alexander, BNL

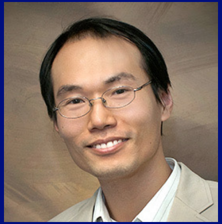
This research was supported by the Exascale Computing Project (17-SC-20-SC), a joint project of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration, responsible for delivering a capable exascale ecosystem, including software, applications, and hardware technology, to support the nation's exascale computing imperative.

Asher Mancinelli, PNNL

Vinay Ramakrish-naiah, LANL

Malachi Schram, PNNL

Josh Suetterlein, PNNL



Shinjae Yoo, BNL

Christine Sweeney, LANL

Frank Alexander, PNNL

Not pictured: S. Ghosh, PNNL, Y. Huang, PNNL, A. Kagawa, BNL, J Mohd-Yusof, LANL, D. Vrabie, PNNL, P. Welch, LANL

How to Solve Control Problems?

- Controllers
  - Good if control algorithm is simple/straightforward and experts are available, otherwise could be intractable
- Optimization
  - Can utilize HPC and try many solutions, works well for problems for which there is a response surface
  - Doesn't work well for more complicated problems
- Supervised Learning
  - Good for problems that have a fixed and labeled data set. Like having a supervisor watch and tell which action agent should have taken. Provides an exact answer.
- Reinforcement Learning
  - Problems where agent must learn by interaction with environment, self-teaching, no need for expert control engineer or labelled data.
  - Used when can formulate problem in terms of Finite Markov Decision Process (described later)

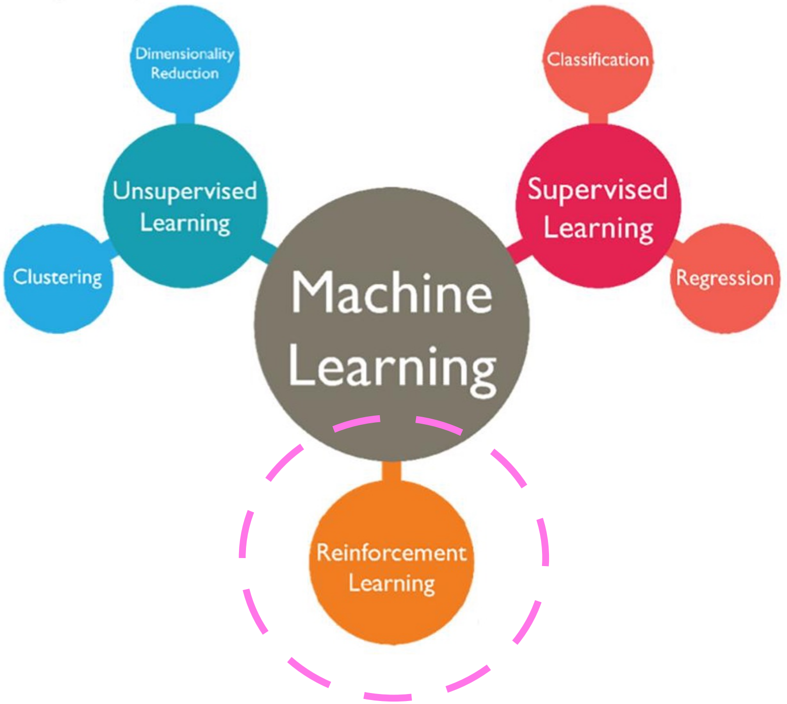
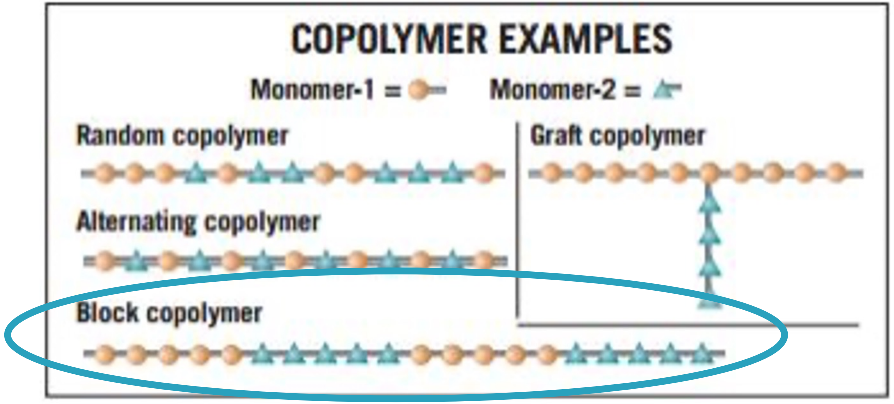


Image courtesy Vishakha Jha
https://www.techleer.com/articles/203-machine-learning-algorithm-backbone-of-emerging-technologies/

Light Source Experiment Control Use Case: Block Copolymer Self-assembly

What is a Block Copolymer (BCP)?

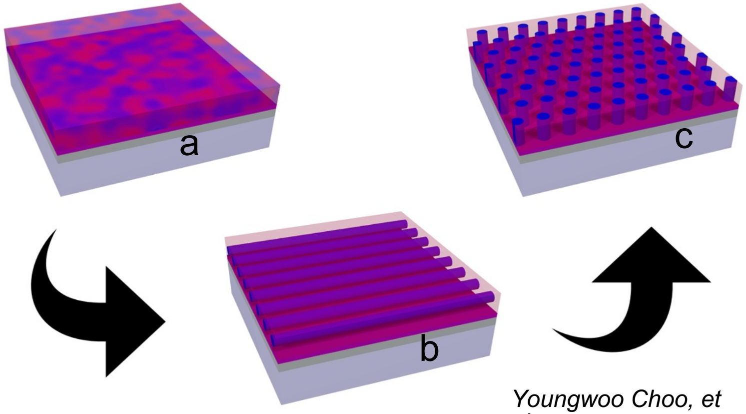


https://www.particlesciences.com/news/technical-briefs/2011/glossary-of-polymer-terms.html

Why do we care about BCPs?

Combining protein and synthetic polymers can create functional biomaterials **useful for catalysis, sensors, nanotechnology and renewable energy**.

How is self-assembly of block copolymers directed?



Youngwoo Choo, et al.

a) Block copolymer is in complete disorder.
b) Laser “orders” BCP into horizontal tubes.
c) High-temp annealing helps achieve desired morphology while maintaining previous order.

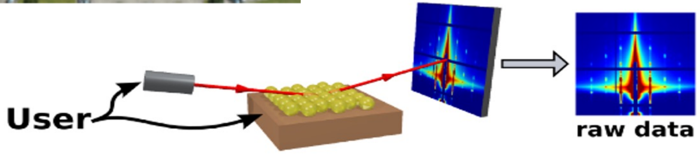
This process could take hundreds of experimental trials to get right!!

Challenges for Block Copolymer Experiments

- BCP experiments are performed at DOE light source user facilities.
- Temperature is adjusted to direct the formation of the block copolymers to a target morphology.

- GISAXS technique is used to detect BCP morphology during directed self-annealing process
- Light source beam shining on sample at small grazing incidence angle produces a diffraction pattern

- The multi-dimensional energy landscape underlying directed block copolymer self-assembly requires engineering a convoluted pathway in order to obtain a target morphology.



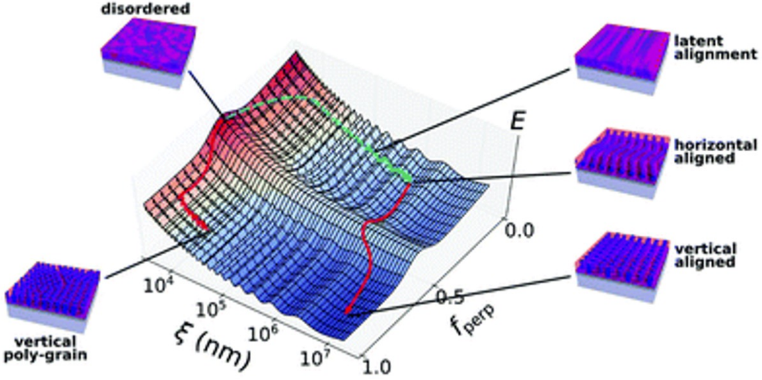
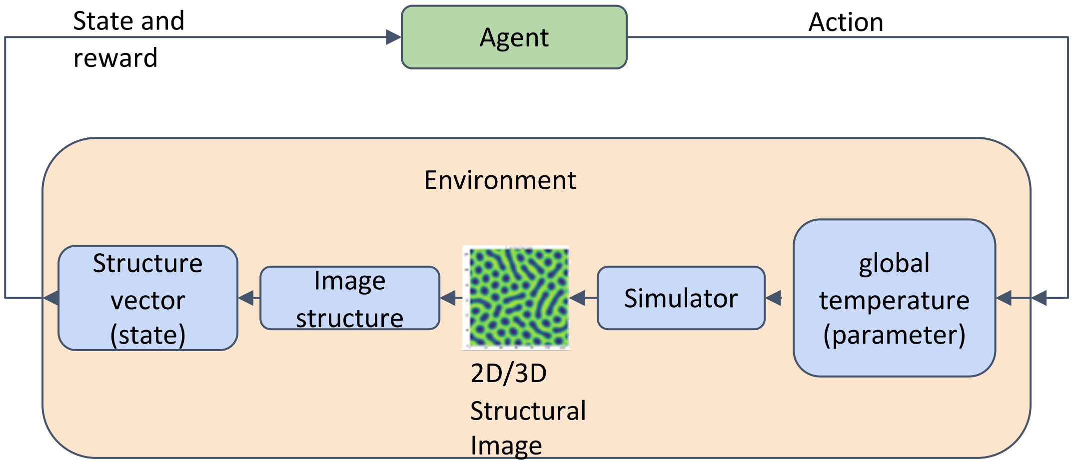


Image from Nanoscale, 2018, 10, 416. Choo, Majewski, Fukuto, Osuji and Yager.

Reinforcement Learning for BCP Self-Annealing



Reinforcement Learning (RL) system:

- Simulate BCP self-assembly (via PDE or MD sim) and produce a real-space image.
- Create training data (vector) based on morphology of BCP, uses conversion to fourier space
- Train RL system with new data point
- Query RL policy developed so far on best parameters to try next.
- Repeat until reach target morphology.

- Mapping to Finite Markov Decision Process (MDP):
  - Agent – scientist controlling temperature of BCP experiment
  - Environment – BCP simulation
  - Actions – increase temperature, decrease temperature
  - Reward – numerical value comparing morphology to target morphology
  - State – structure vector (characterizes morphology of BCP)

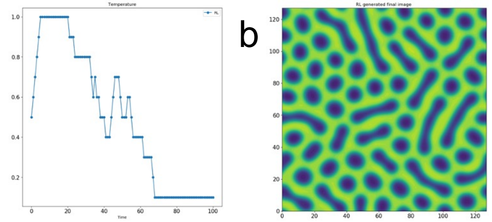
Long-term vision would be to use the learned policy at an experiment and also to ultimately transfer experimental data into the RL system.

BCP Reinforcement Learning Challenges and Results

Challenges:

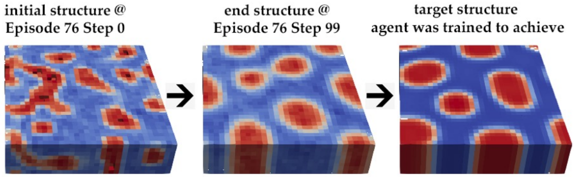
- Training data
- Structure vector to capture characteristics

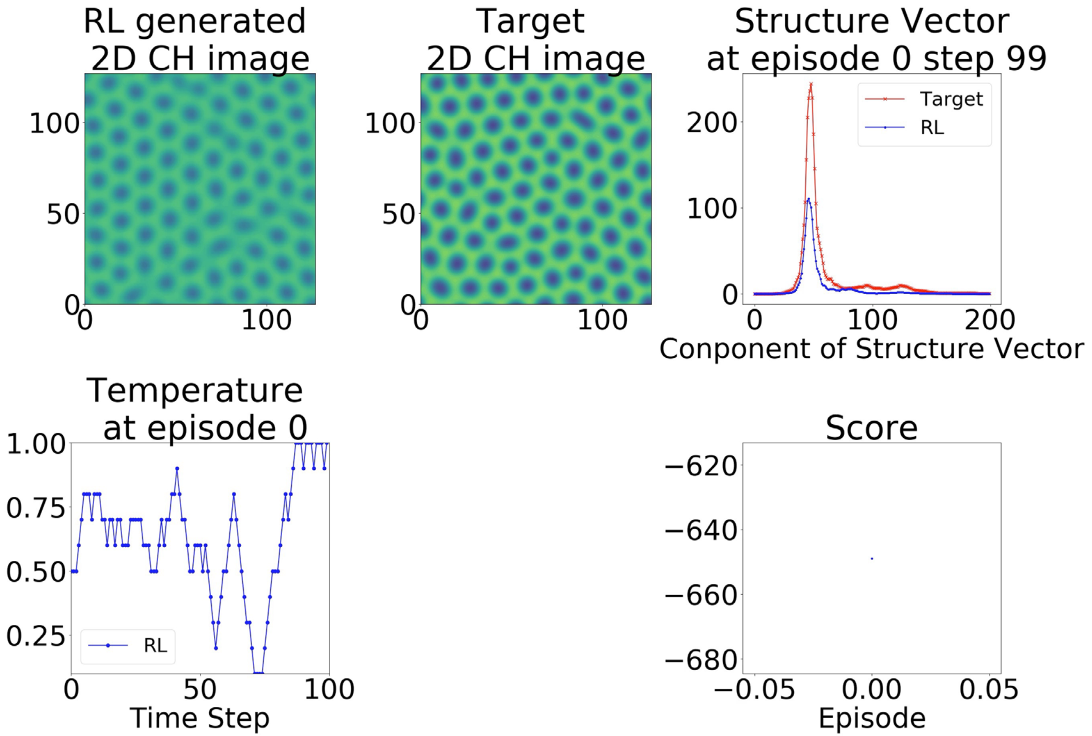
a b



RL algorithm develops policy that helps control temperature during self-annealing (a), which results in BCP morphology (b).

3D Block Co-polymer Reinforcement Learning Application



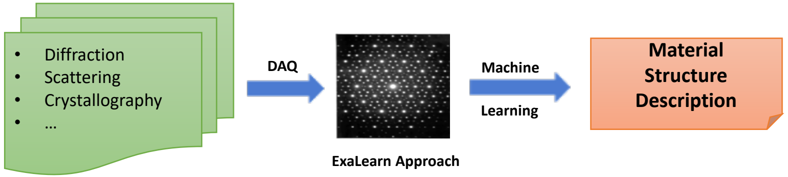


Questions?

Backup Slides

Extreme-scale Machine Learning for Inverse Problems

- **Long-term goal:** develop and deploy ML-driven solutions of large-scale inverse problems that are directly relevant to DOE-related science and technology
- Given a set of observations, **inverse problems** seek to determine the parameters that produced those observations.
- Inverse problems arise in numerous DOE-related scientific application domains, e.g.,
  - Fusion physics: given plasma equilibrium profiles in tokamaks/stellarators; determine device diagnostics.
  - Microscopy: various kinds of microscopy–electron, scanning tunneling, transmission electron, and others; given a microscopy image, determine the material properties that produced the observed image.
  - X-ray crystallography: determine structure of target from diffraction patterns produced by it upon bombardment by incident X-ray beam.
  - Additive manufacturing: determining thermal parameters from target solidification microstructures in powder-bed metal additive manufacturing.
- **Short-term goal:** Develop extreme-scale ML framework to solve the inverse problem of material structure determination from neutron scattering experiments.



ExaLearn Approach

Team Members: Cristina Garcia-Cardona (LANL), Ramakrishnan Kannan (ORNL), Travis Johnston (ORNL), Thomas Proffen (ORNL), Daniel Olds (BNL), Katherine Page (ORNL/UTK). Team Lead: Sudip K. Seal (ORNL)

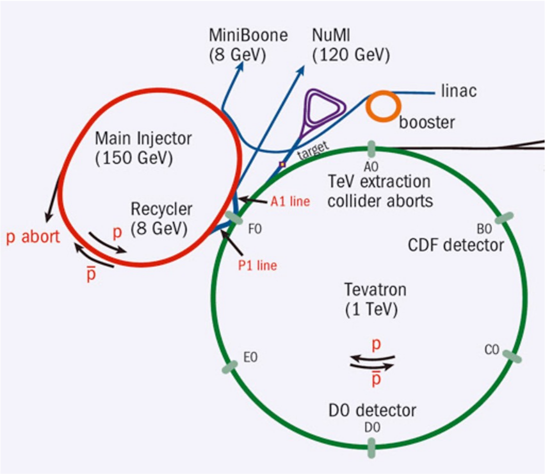
Learning to Predict Material Structure from Neutron Scattering Data, Workshop on Big Data, Tools and Methods (BTSD), IEEE Big Data 2019, Los Angeles, Dec 9-12, 2019.

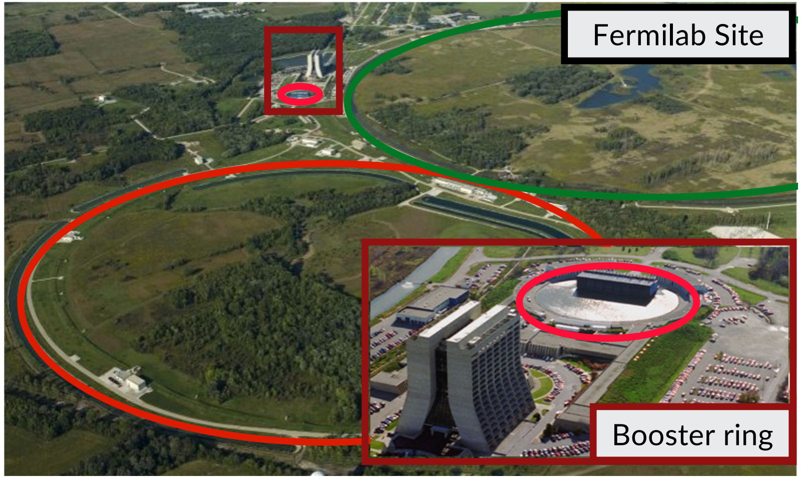
Structure Prediction from Neutron Scattering Profiles: A Data Sciences Approach *IEEE Big Data 2020, Dec 10-13, 2020.*

ExaBooster -- FermiLab (FNAL) Booster

Problem definition:
Reduce beam losses in the FNAL Booster by developing a Machine Learning (ML) model that provides optimal set of actions for accelerator controls

FNAL Accelerator Complex:

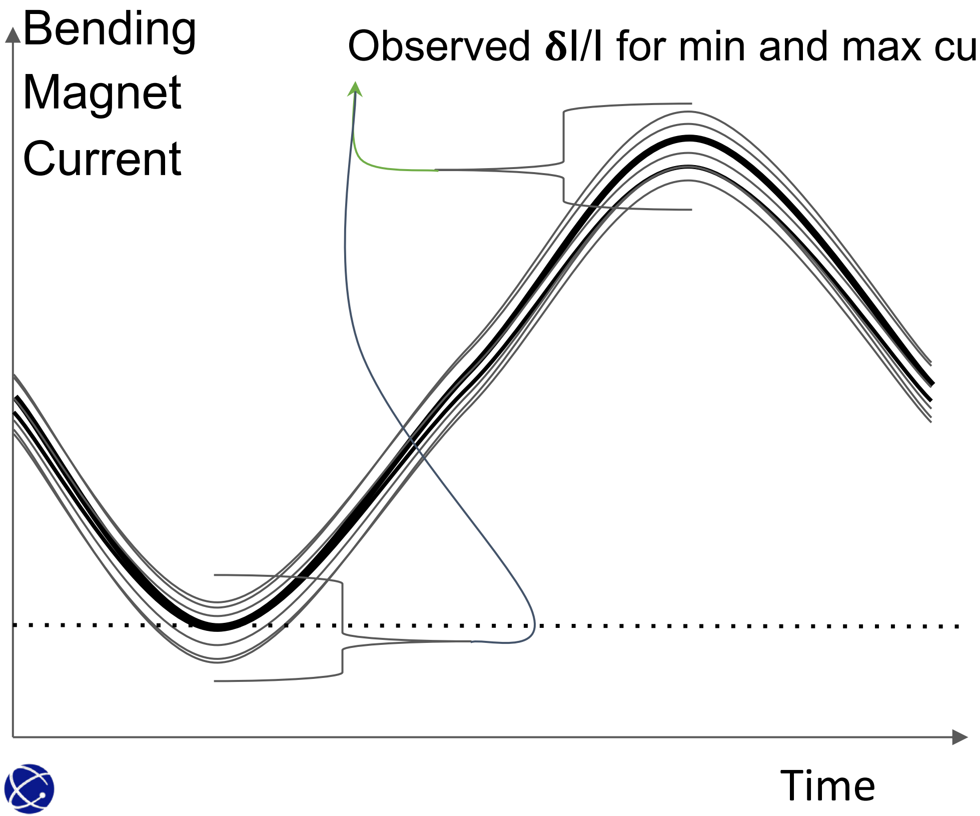




Courtesy: Christian Herwig

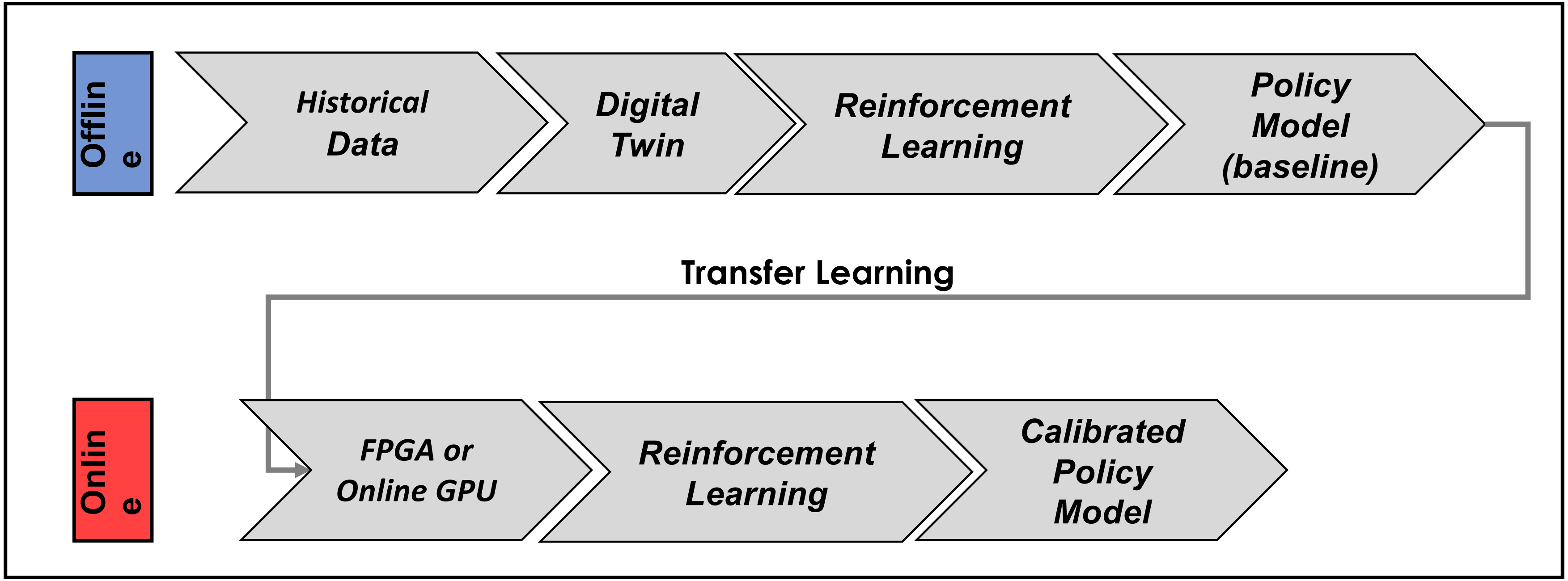
Data is available at zenoto
Original work developed by PNNL, FNAL, University of California San Diego, Columbia University

ExaBooster: The Need for Improving Regulation



- Spread in B-field degrades beam quality and contributes to losses
- Focusing on min for now:
  B_VIMIN = Setting to achieve
  B:VIMIN = Prescribed remedy from PID regulator circuit
  B:IMINER = Error discrepancy
- Policy model is focused on controlling the regulator to reduce the error

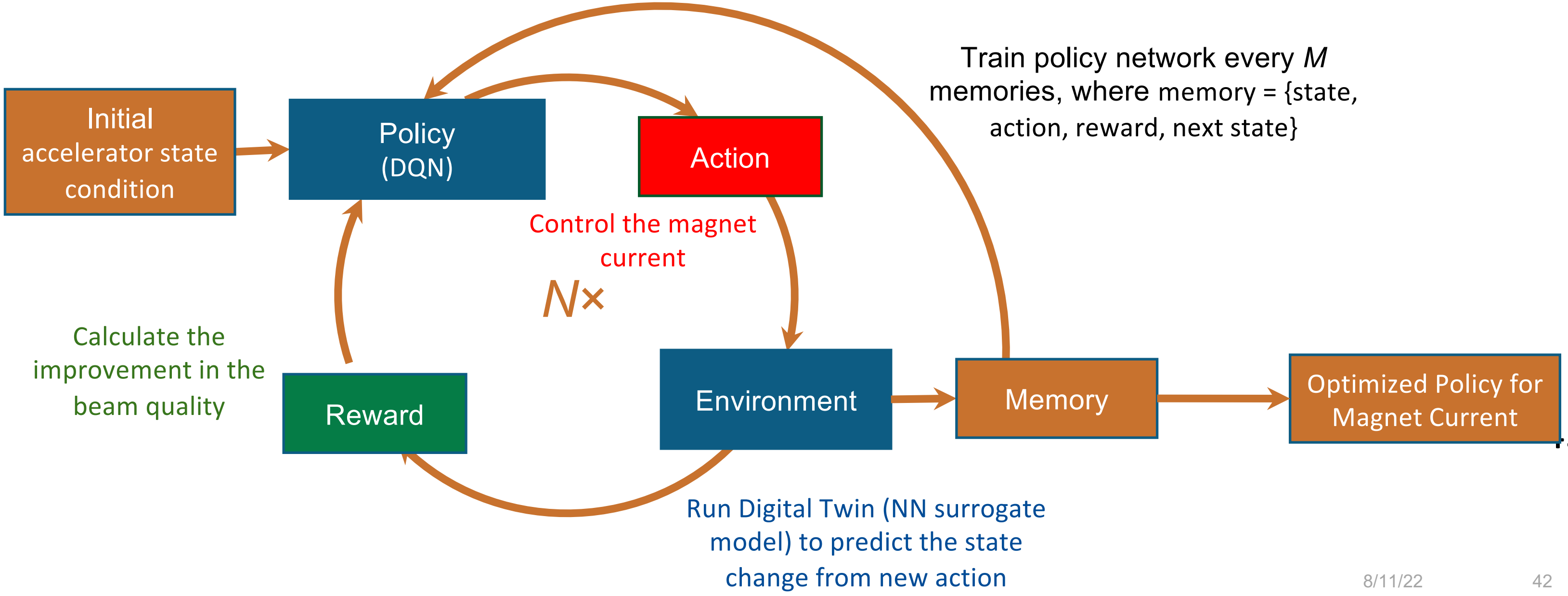
Proof of Concept Workflow for ExaBooster



- Digital twin provides accurate predictions of future time for key variables to be used by the reinforcement learning framework
- Historical temporal information from key variables was available based on subject matter expert input

Reinforcement Learning Framework for ExaBooster

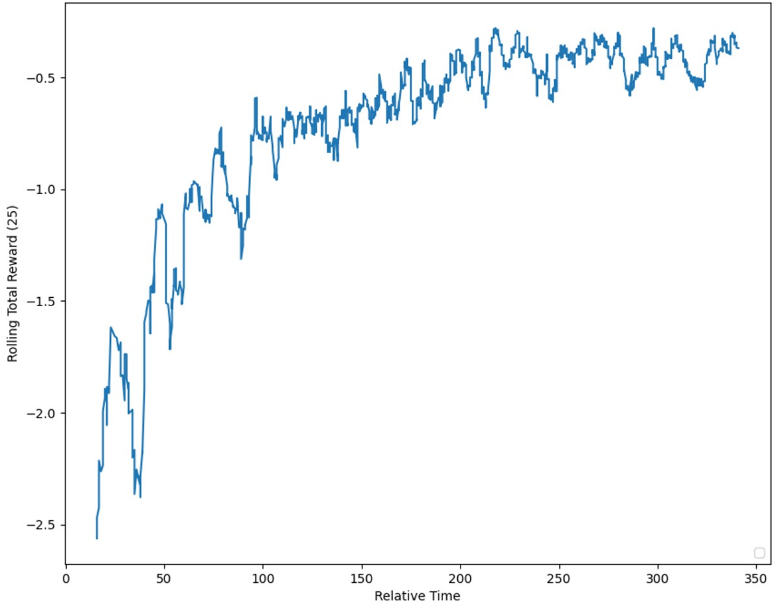
Starting with an initial accelerator state, can we train a RL policy to improve beam quality through the magnet current ?

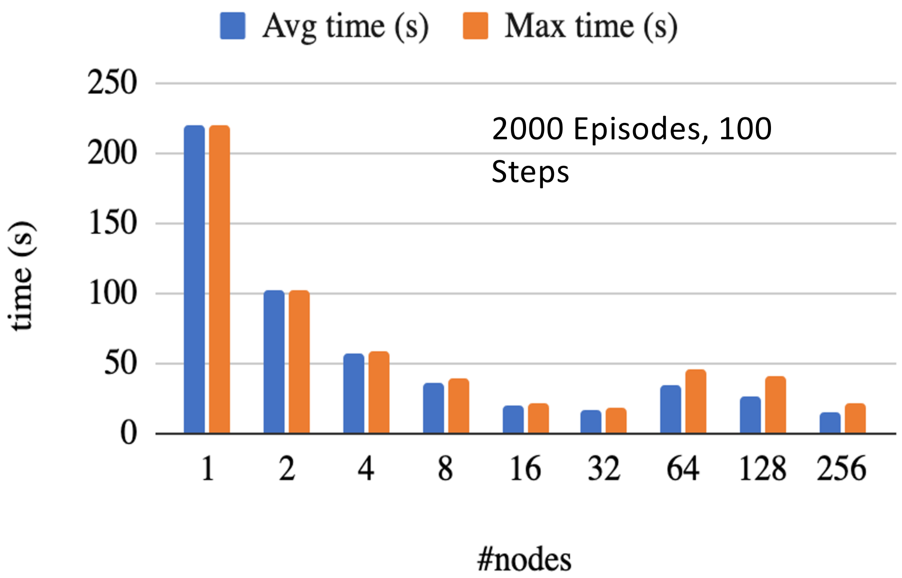


ExaBooster Performance Results

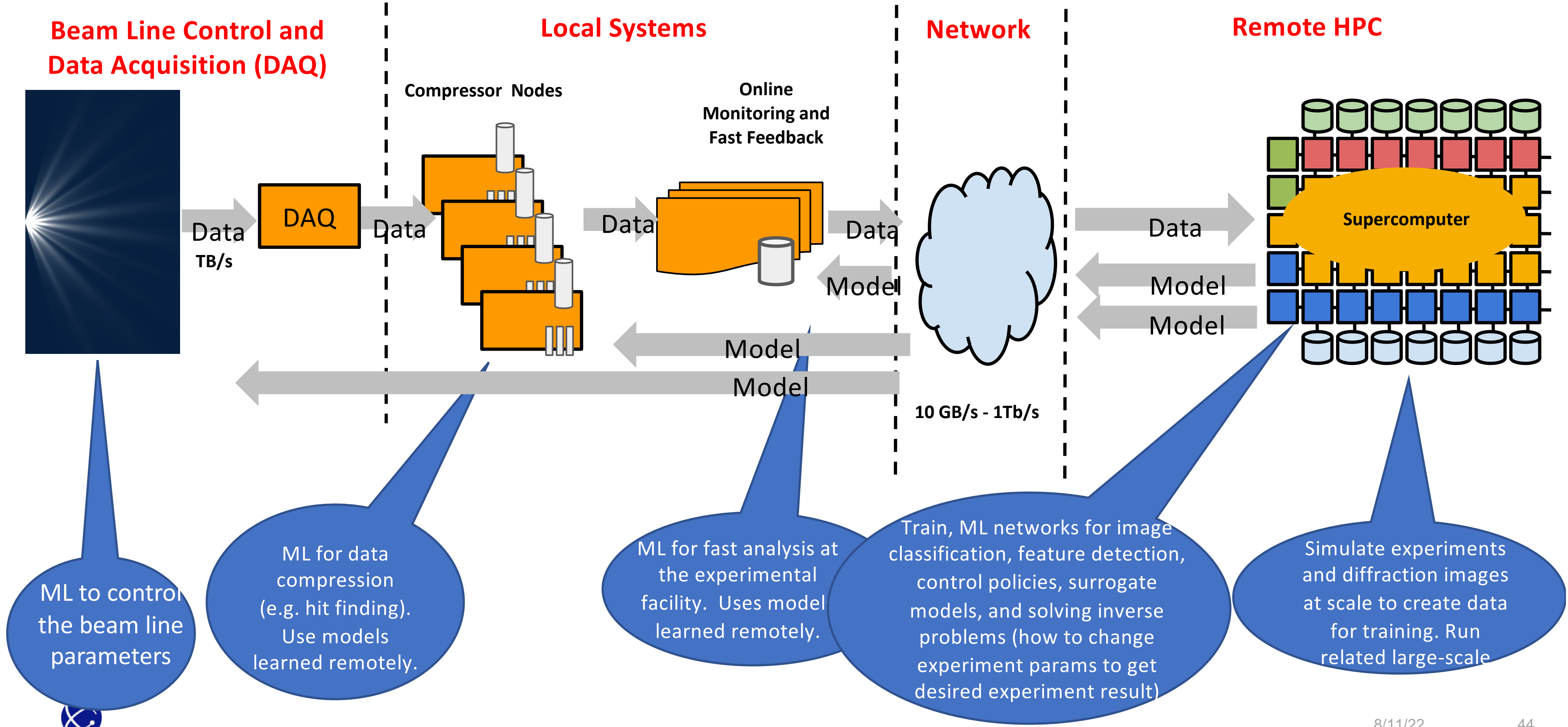
The optimization was formulated as an episodic problem:

- An episode is composed of 100 sequential steps
- After each episode the environment was reset to the same initial state
- A batch size of 32 experiences were randomly sampled to train the active policy model
- A ϵ-greedy method was used to control the level of exploration/exploitation





Machine Learning in Experimental Workflows



Interfacility workflow icons from ExaFEL project workflow