LA-UR-20-27469

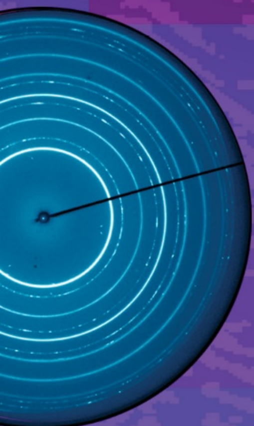
Approved for public release; distribution is unlimited.

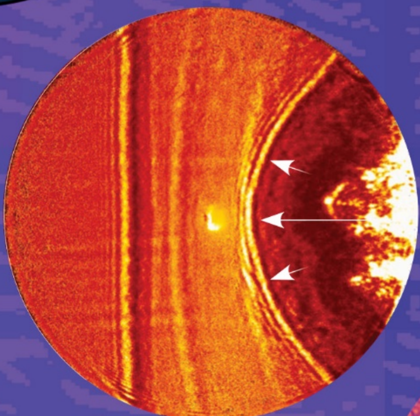
Title:	Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities Workshop Report
Author(s):	Sweeney, Christine Marie; Sturtevant, Blake; Biwer, Christopher Michael; Bolme, Cynthia Anne; Huber, Rachel Colleen; Huston, Larissa Qian; McBride, Emma; Miyagi, Lowell; Prescher, Clemens; Ramos, Kyle James; Smith, Jesse; Turton, Terece
Intended for:	Report
Issued:	2021-03-04 (rev.2)

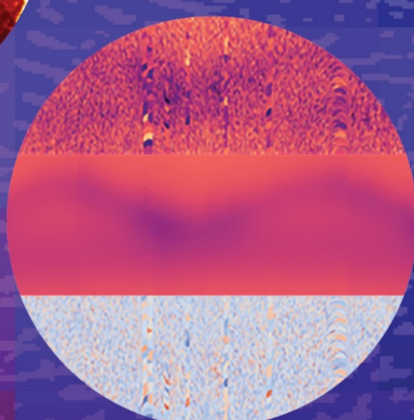
Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities

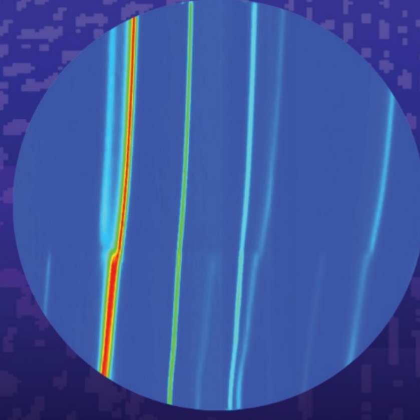
September 8-11, 2020

Workshop Report









Los Alamos
NATIONAL LABORATORY

Cover: Don Montoya, LANL

Images: Background taken from Figure 7. Spheres (left to right): X-ray diffraction pattern from Blake Sturtevant (LANL), Figure 12, Figure 3, contour plot from Blake Sturtevant (LANL), and the automatic colorization using ColorMapND for a multi-channel fluorescence image of a mixed ionic-electronic conductor sample [from S. Cheng, W. Xu, K. Mueller, “ColorMapND: A Data-Driven Approach for Mapping Multivariate Data to Color,” IEEE Trans. on Visualization and Computer Graphics (TVCG), 2018].

LA-UR-20-27469

Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities

Workshop Report

September 8–11, 2020

Virtual Venue

Christine Sweeney, Blake Sturtevant, Christopher Biwer, Cynthia Bolme, Rachel Huber, Larissa Huston, Emma McBride, Lowell Miyagi, Clemens Prescher, Kyle Ramos, Jesse Smith, Terece Turton

Chair:	Christine Sweeney, Los Alamos National Laboratory
Steering Committee:	Blake Sturtevant, Los Alamos National Laboratory Christopher Biwer, Los Alamos National Laboratory Cynthia Bolme, Los Alamos National Laboratory Rachel Huber, Los Alamos National Laboratory Larissa Huston, Los Alamos National Laboratory Emma McBride, SLAC National Accelerator Laboratory Lowell Miyagi, University of Utah Clemens Prescher, Deutschen Elektronen-Synchrotron Kyle Ramos, Los Alamos National Laboratory Jesse Smith, Advanced Photon Source, Argonne National Laboratory
Keynote Speaker:	Nenad Velisavljevic Staff Scientist, Physics Division, Lawrence Livermore National Laboratory Director of High Pressure Collaborative Access Team (HPCAT) at the Advanced Photon Source, Argonne National Laboratory
Plenary Speakers:	Matthew Horton, Lawrence Berkeley National Laboratory Harinarayan Krishnan, Lawrence Berkeley National Laboratory Zhehui (Jeph) Wang, Los Alamos National Laboratory
Session Speakers:	Richard Briggs, Lawrence Livermore National Laboratory Erik Watkins, Los Alamos National Laboratory Trevor Willey, Lawrence Livermore National Laboratory Mungo Frost, SLAC National Accelerator Laboratory Simon Hunt, Department of Materials, University of Manchester, Manchester, UK Rachel Husband, Deutsches Elektronen-Synchrotron David Montgomery, Los Alamos National Laboratory, P-24 Bob Nagler, SLAC National Accelerator Laboratory Robert Von Dreele, Advanced Photon Source, Argonne National Laboratory Devin Francom, Los Alamos National Laboratory Daniel Orban, University of Minnesota Wei Xu, Brookhaven National Laboratory Thomas Caswell, Brookhaven National Laboratory Valerio Mariani, SLAC National Accelerator Laboratory Lars Ehm, Stony Brook University Hauke Marquardt, University of Oxford Michael Armstrong, Lawrence Livermore National Laboratory Sven Vogel, Los Alamos National Laboratory
Panelists:	Debbie Bard, Lawrence Berkeley National Laboratory Jana Thayer, SLAC National Accelerator Laboratory Thomas Michelat, European XFEL Kevin Jorissen, Amazon Web Services
Session and Panel Leads:	Rachel Huber, Los Alamos National Laboratory Jesse Smith, HPCAT, Argonne National Laboratory Emma McBride, SLAC National Accelerator Laboratory Brian Jensen, Los Alamos National Laboratory Terece Turton, Los Alamos National Laboratory

	Christopher Biwer, Los Alamos National Laboratory
Institutional Host:	Richard Sheffield, Los Alamos National Laboratory
Website:	Else Tennessen, Los Alamos National Laboratory
Workshop Sponsor:	Los Alamos National Laboratory
Workshop Venue:	Virtual
Communications:	Don Montoya and Emily Rybarcyk, Los Alamos National Laboratory

This report is available on the web at www.lanl.gov/conferences/compress.

Executive Summary

The application of high pressure to materials has enabled discoveries in scientific fields such as planetary science, materials science, and materials synthesis. Recent advances in X-ray user light sources and other facilities, co-location and integration of user facilities with high-pressure drivers, availability of high-performance computing (HPC) platforms, and the development of new data science techniques have created opportunities for, and challenges in, advancing data analytics for rapid and dynamic compression experiments.

To address these challenges, harness the emerging technology now available, and expedite scientific discovery, Los Alamos National Laboratory (LANL) hosted a virtual workshop entitled "Data Science and Computation for Rapid and Dynamic Compression Workflows at Experimental Facilities" from September 8 to 11, 2020. The workshop included 95 registered scientists and analytics experts from 15 universities, 9 United States (US) national laboratories, 5 US and European X-ray light sources, neutron sources such as the Los Alamos Neutron Science Center (LANSCE), other big science facilities such as the National Ignition Facility (NIF), and an industry representative. The workshop included 31 invited talks and 4 lightning talks by students and post-docs.

The workshop began with a keynote presentation by Dr. Nenad Velisavljevic, Director of the High Pressure Collaborative Access Team (HPCAT) at the Advanced Photon Source (APS). Plenary talks focused on the Basic Energy Sciences Data Project light source software stack, the Materials Project community data sharing platform, and detector innovations. Breakout sessions and panels covered data analytics for static and dynamic compression experiments; X-ray diffraction (XRD) and scattering data types; time-dependent measurements using status apparatus, image processing, machine learning (ML), visualization, real-time, and post hoc analytic workflows; and facility computational resources.

The workshop identified the following priority gaps: new large data volumes and high data rates from upgraded facilities cannot be handled; data analytics are not performed during experiments, which results in inefficient use of valuable experimental run time; data characteristics such as noise, phase wrap, data sparseness, and artifacts impede analyses; few labeled datasets and metadata are available for advanced data science; and experiments cannot be conducted remotely, nor can resources be harnessed from multiple facilities.

The workshop identified the following priority areas for research and investment: analysis algorithm development for image processing, machine learning, and signal processing; universally applicable and accessible software tools with common interfaces and the ability to be customized across a wide range of experimental apparati; integration of multiple streams of data, multiple analysis techniques, and data types; metadata creation and management and common data pedigree despite different origins; science and computing user facilities to invest in user engagement and encourage interdisciplinary teams; algorithms and data storage infrastructure for data processing and reduction of increased data volumes and velocity; forward model development and combination with data analytic techniques; and portable interfacility and remote workflows to encourage collaboration.

Workshop on Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities

Session Topics

Data Science:
Image Processing, Advanced Analysis, Modeling, Machine Learning, Statistics, Visualization.

Compression Data and Analysis:
X-ray Diffraction, Images, VISAR, Phase Fractions and Transitions, Features

Technology Advances:
High-energy Sources, Improved Detectors, Increased Repetition Rates, Supercomputers

Computational Workflows:
Real-time Analytics, Computing Facility Support, Resources, Tools

Priority Gaps

Cannot handle new large data volumes and velocities from upgraded facilities.

Data analytics are not performed during experiments, thwarts experimental design.

Data characteristics such as noise, phase wrap, data sparseness and artifacts impede analyses.

Few labeled datasets and metadata for advanced data science.

Cannot conduct experiments remotely or harness resources from multiple facilities.

Priority Investments

Analysis algorithm development for image processing, machine learning, and signal processing.

Universally applicable and accessible software tools with common interfaces and ability to be customized.

Integration of multiple streams of data, multiple analysis techniques and data types.

Metadata creation and management, common data pedigree despite different origins.

Science and computing user facilities invest in user engagement and encouraging interdisciplinary teams.

Algorithms and data storage infrastructure for data processing and reduction of increased data volumes and velocity.

Forward model development and combination with data analytic techniques.

Portable interfacility and remote workflows to encourage collaboration.

Table of Contents

List of Acronyms, Abbreviations, and Initialisms

2D	Two-Dimensional
ALS	Advanced Light Source
API	Application Programming Interface
APS	Advanced Photon Source
AWS	Amazon Web Services
BES	Basic Energy Sciences
BNL	Brookhaven National Laboratory
CBF	Crystallographic Binary File
CCD	Charge-Coupled Device
CFN	Center for Functional Nanomaterials
CMS	Complex Materials Scattering
CNN	Convolutional Neural Network
CTF	Contrast Transfer Function
DAC	Diamond Anvil Cell
DCS	Dynamic Compression Sector
dDAC	Dynamic Diamond Anvil Cell
DESY	Deutsches Elektronen-Synchrotron
DiPOLE	Diode Pumped Optical Laser for Experiments
ECB	Extreme Conditions Beamline
FEL	Free Electron Laser
GB	Gigabyte
GPU	Graphics Processing Unit
GSAS	General Structure Analysis System
HDF	Heterogeneous Data Format
HPC	High-Performance Computing
HPCAT	High Pressure Collaborative Access Team
LANL	Los Alamos National Laboratory
LANSCE	Los Alamos Neutron Science Center
LCLS	Linac Coherent Light Source
Linac	Linear Accelerator
LLNL	Lawrence Livermore National Laboratory
MAUD	Material Analysis Using Diffraction
MEC	Materials in Extreme Conditions
MDS	Multidimensional Scaling
ML	Machine Learning
NERSC	National Energy Research Scientific Computing Center
NIF	National Ignition Facility

NSLS-II	National Synchrotron Light Source–II
PCI	Phase Contrast Imaging
PETRA III	Synchrotron light source at DESY
SAXS	Small-Angle X-Ray Scattering
SLAC	Stanford Linear Accelerator Center
TIFF	Tagged Image Format File
US	United States
VISAR	Velocity Interferometer System for Any Reflector
XANES	X-Ray Absorption Near Edge Structure
XFEL	X-Ray Free Electron Laser
XPCI	X-Ray Phase Contrast Imaging
XPCS	X-Ray Photon Correlation Spectroscopy
XRD	X-Ray Diffraction

1.0 WORKSHOP OBJECTIVES AND BACKGROUND

1.1 Introduction

The application of high pressure to materials has enabled discoveries in a diverse range of scientific fields. It has helped scientists understand the interior of the earth, measure the properties of materials, synthesize new materials, and understand the behavior of materials at extreme conditions. Quasi-static and dynamic compression experiments are performed in third- and fourth-generation light sources around the world. These techniques and drivers include diamond anvil cell (DAC),[1] dynamic DAC (dDAC),[2] laser,[3] gas gun,[4] and high-explosive drivers.[5] Compression experiments study the high-pressure phases of numerous metals and soft materials and the mixing of liquids and gases.

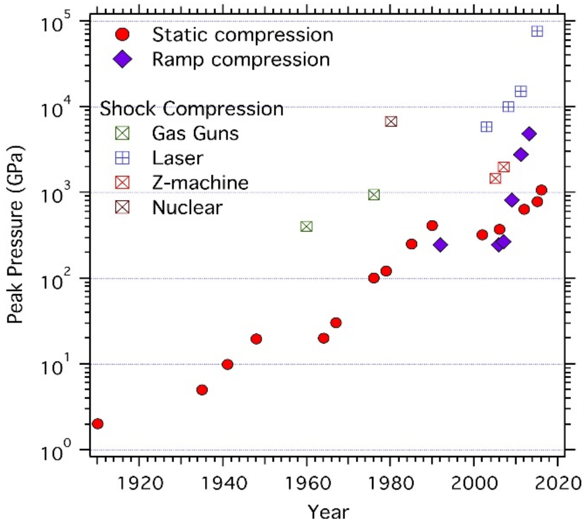


Figure 1. Progression of pressures achieved in the laboratory as a function of year for shock (open squares), static (red circles), and ramp-compression techniques (purple diamonds).[6]

To meet the current data analysis challenges for compression science, harness the emerging technology available, and expedite scientific discovery, Los Alamos National Laboratory (LANL) hosted a virtual workshop entitled "Data Science and Computation for Rapid and Dynamic Compression Workflows at Experimental Facilities" from September 8 to 11, 2020. The workshop included 95 registered scientists and analytics experts from 15 universities, nine United States (US) national laboratories, five US and European X-ray light sources, neutron sources such as the Los Alamos Neutron Science Center (LANSCE), other facilities such as the National Ignition Facility (NIF), and one industry representative. The workshop included 31 invited talks and four lightning talks by students and postdocs. Video conferencing was enabled by WebEx, talks were live, and questions were asked in real time using the chat feature. A more traditional Q&A session occurred after each talk.

After a formal welcome from Antoinette Taylor, Associate Lab Director for Physical Sciences at LANL, the workshop began with a keynote presentation from Dr. Nenad Velisavljevic, Director of the High Pressure Collaborative Access Team (HPCAT) at the Advanced Photon Source (APS). Plenary talks focused on the Basic Energy Sciences Data Project light source software stack, the Materials Project community data sharing platform, and detector innovations. Breakout sessions and panels covered data analytics for static and dynamic compression experiments; X-ray diffraction (XRD) and scattering data types; time-dependent measurements using status apparatus, image processing, ML, visualization, real-time, and post-hoc analytic workflows; and facility computational resources. The workshop aimed to find consistent themes among various compression science experiment sub-communities so that crosscutting needs and priority investment areas using data science could be discerned.

The contributions and level of engagement by workshop attendees were remarkable despite the virtual workshop format and the relatively specialized experimental domain; this is encouraging and bodes well for future activity in this community. The issues raised in this workshop are important to having a technological edge in an ever-more competitive national and international environment and are important for accelerating scientific discovery supporting academia, industry, and the national mission of laboratories represented at the workshop.

This report presents a brief background on recent developments in compression science that relate to data analysis challenges; summarizes the major themes and topics covered in the workshop, including current status, needs, and areas of investment for each topic; and provides a final list of crosscutting priority needs and priority areas of investment. The report ends with suggested activities and appendices that provide further details on the workshop.

1.2 Background

Experimental measurements on materials under high pressures have always been challenging to obtain because of the small volumes of compressed materials (such as in DACs) and because of the transient nature of the compressed states, such as those states reached through ramp or shock compression. These challenges are being continually addressed using the evolving precision and resolution of diagnostics. These diagnostics have long been invaluable tools for understanding properties such as the phase, structure, magnetic ordering, and electronic properties of materials under extreme pressures and temperatures.

Recent advances in light source technologies have produced high-energy pulsed X-ray beams with intensity adequate to make high-precision materials measurements using a single X-ray pulse. Upgraded facilities provide X-rays with higher brilliance, increased energy, increased pulse repetition rate, improved coherence, and tighter focusing. The short durations of these pulses are typically ~100 ps at X-ray synchrotrons and <100 fs at X-ray free electron lasers (XFELs). Short pulses enable time-resolved measurements that present transformational opportunities for compression science. These technological advances have enabled a rapid expansion in the number of compression experiments and in the number of diverse diagnostic techniques. Increased brightness has allowed for more imaging techniques, for example.

Improvements in detector technology (e.g., improved efficiency for hard X-rays, improved repetition rate, and higher resolution) augment the upgraded sources and further enable time-resolved diffraction and imaging in dynamic and rapid compression experiments. Research in detector technology involves new materials, innovations in the structure of detectors, an increased field of view, and novel data science to enhance detector hardware performance.[7]

Increased X-ray pulse repetition rates and faster detectors have significantly advanced the pace of high-pressure experiments. The commercialization and widespread use of hybrid pixel-array detectors[8] facilitates large-format, two-dimensional (2D) X-ray imaging at frequencies on the order of kilohertz to megahertz. As a result of significantly more data of higher resolution conveyed by detectors, the data volumes and the velocity at which they are produced have become a new challenge. For example, for dDAC experiments, in a typical 48–72 hour beamtime, this can result in >10,000 images or 400–900 GB of data. As shock studies move to 10 Hz sources and Diode Pumped Optical Laser for Experiments (DiPOLE) lasers come online,[9] datasets will become much larger. To fully use the new data, collaboration with data scientists is essential in the development of data analysis, visualization, and computational workflows. New approaches to data storage, sharing, and management are also needed to support data analytics.

In addition to the revolution in X-ray light sources, this workshop was motivated by the co-location and integration of the light sources with high-pressure drivers, including dDAC, large volume hydraulic presses (e.g., multi-anvil or Paris-Edinburgh style), gas guns, high-energy lasers, and high explosives. Ramp and rapid compression experiments at light sources have been enabled by the development of and remote, programmable operation of mechanical, pneumatic, and piezoelectric pressure generation devices.[10] These newly enabled technical capabilities further support scientific possibilities[11] for research areas, including fast equation-of-state, materials synthesis and metastability, and generating high strain rates.

While light source science is growing and encountering new challenges, computing is also evolving. Large-scale and heterogeneous hardware resources and data science (Figure 4) can be leveraged to help meet light source data challenges. High-performance computing (HPC) and cloud computing support processing large amounts of data where potentially multiple analyses can be performed in parallel. They also support computationally intensive simulations that can be used in conjunction with data analytics. However, experimentalists do not typically run large simulations or access large computing facilities remotely, so they require workflow-enabling abstractions and mechanisms to harness this computing power.

2.0 DATA ANALYSIS OBJECTIVES AND TOOLS

2.1 Data Types and Processing for Compression Experiments

2.1.1 Background and State of the Field

Compression experiments often rely on XRD images for their primary analysis. Analysis of XRD provides information on pressure reached during an experiment, typically through the use of a well-characterized standard material that is phase-stable over the pressures of interest (e.g., copper, gold, or platinum). XRDs for dDAC experiments on powder or polycrystalline samples, for example, consist of concentric rings because of the random orientations of the numerous crystallites. Shock experiments in which a single crystal sample is used yield XRDs that are in a speckled pattern instead of rings. Further peak fitting and advanced analysis of integrated XRD can yield information on frame number and pressure of a phase change, phase fractions during the course of a phase transition, and texture information. Recently, small-angle X-ray scattering (SAXS) has also been used to study compression and dynamic processes and provides information about the ordering of very large features.

An imaging microscope can be positioned between two diffraction detectors in a dynamic compression platform to provide companion images that can show, for example, nucleation and growth and melting at different pressure and/or temperature conditions and that can help track changes in particle size as a function of frame number and compression rate. For shock compression experiments, Velocity Interferometer System for Any Reflector (VISAR) data[12] are used to measure the velocity of the material during the shock and are often coordinated with XRD collected simultaneously. VISAR can be difficult to analyze in real-time but is useful in planning what laser pulse shape and what pressure are needed for the subsequent experiment. These compression experiment data types are shown in Figure 2.

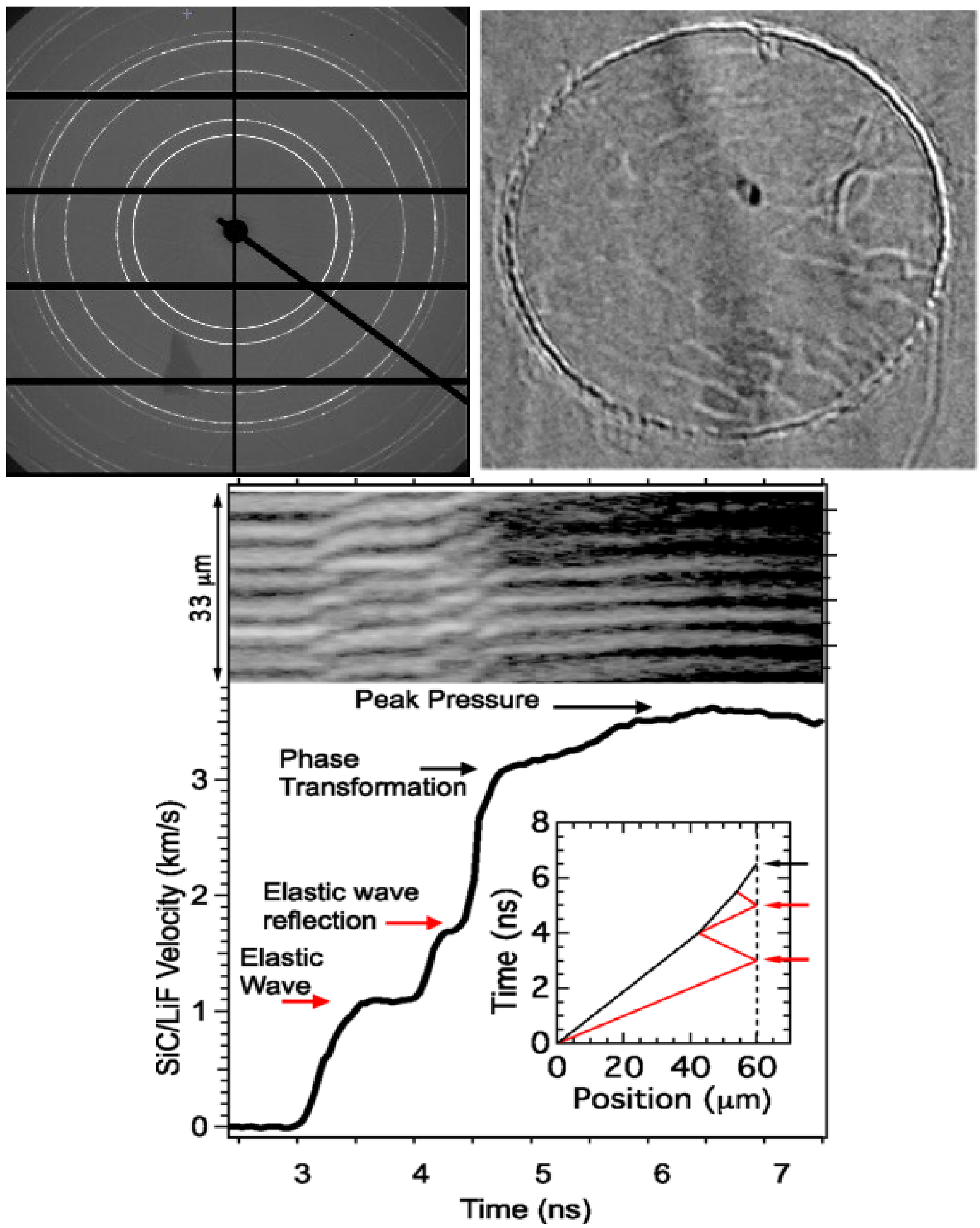


Figure 2. Typical data types for compression experiments. Top left: (XRD) Powder diffraction pattern collected from a DAC at APS. Top right: (image) Crystallization of Ga from Pioneers in Cameras and Optoelectronics (PCO) high-speed camera; image processing as part of a DESY/LLNL collaboration. Bottom: (plots) VISAR wave for shock entering SiC and schematic wave diagram with elastic wave (red) and second elastic reverberation (black).[13]

2.1.2 Needs Identified

Real-time data analysis to steer experiments is needed in both the shock and dDAC communities and would be helpful even in “slower” traditional static DAC experiments. In dDAC experiments, real-time determination of pressure is available, but analysis of the uniaxial stress state and the identification of phase transitions with quantified phase fractions during the course of the transition is still needed. For shock compression experiments, processing VISAR in real time to inform the next shot is advantageous but is not easy to do.

Compression experiments can result in large data sets on the order of tens of thousands of images and nearly a terabyte of data. Large data can be another impediment to real-time data analysis. XFEL campaigns, in which time between shots can be mere minutes, produce larger data volumes.

Many users currently use independently developed software to analyze data, and each user may have his or her own software and/or code. A number of community tools for data analysis exist such as Rietveld refinement, General Structure Analysis System II (GSAS-II)[14] and Material Analysis Using Diffraction (MAUD)[15]. However, current tools are time consuming, not optimized for batch processing, or batch processing is not feasible when sample complexities are present (e.g., texture, phase transitions, and deviatoric stresses).

Experimental campaigns require long stretches of experimentalist time, frequently more than 12 hours; therefore, user-friendly analytical techniques are preferred as they allow for easier concentration over these long periods. Most of these experimental regimes would benefit from universal software that could provide initial integration, background subtraction, and efficient masking of any artifacts produced by drivers or pressure trackers.

Issues in specific beamlines include how the Dynamic Compression Sector (DCS) at APS uses a pink beam to provide more photons to their users; however, this pink beam results in broad and asymmetric diffraction peaks and incorrect sample-to-detector calibrations. A means to quickly and accurately correct for this artifact of the beamline would be greatly beneficial to all users. DCS also uses four charge-coupled device (CCD) cameras to collect radiographic images of a sample using four sequential X-ray bunches, separated by ~150 ns in time. The correction of these four cameras cannot be easily performed by the user. Ghosting, where subsequent images of a phosphor are collected so close in time that the fluorescence from the previous image has not fully decayed, is also an issue.

Finally, SAXS data presents challenges in data analysis because it relies on form factor scattering instead of the traditional diffraction peaks. This results in datasets that are extremely difficult to reduce at the beam time.

2.1.3 Recommended Investment Areas

It was suggested that a portable analysis package that could be implemented at multiple beamlines to provide on-the-fly, easy-to-use XRD, SAXS, pressure calibration, and velocimetry analysis between experiments would be extremely valuable. It would be beneficial if the experimentalist would be willing to have a transfer program such as Globus to directly import the data to the analysis package instead of relying on the use of portable memory (e.g., USB drives) to transfer data between computers.

Dynamic compression experiments, particularly shock compression, often have complex detector geometries with limited coverage of reciprocal space. This can result in a non-unique interpretation of the data. Forward modeling underlying physical processes to directly fit data (e.g., modeling diffraction patterns for parent and daughter phases during a phase transformation using orientation relations between the phases to match a series of diffraction images collected during a phase transformation) can provide a way forward to interpret diffraction data originating from detectors with limited coverage.

There is a need for widely usable software, coordinated composable software components, and/or community-defined interfaces, particularly for real-time data analytics. In the synchrotron DAC community, this exists to an extent. Most beamlines use Dioptas[16] for rapid pressure and phase identification. However, the shock compression community uses many different codes to process VISAR data.

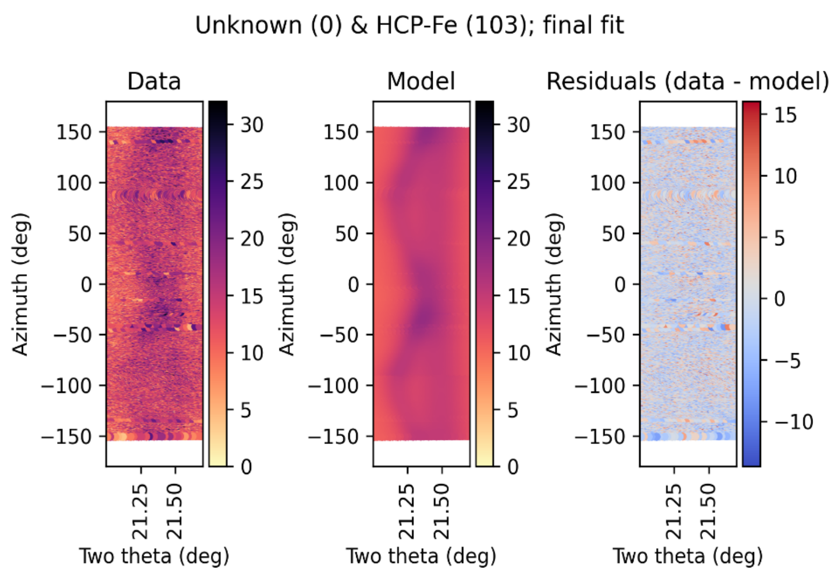


Figure 3. The interaction of X-rays with crystals by a process of diffraction is used to prove the structure of matter at the atomic scale. The diffraction of X-rays at specific angles produces peaks like those in the data (left). Fitting these peaks with improved models (model fit, center; residuals, right) advances our understanding of how matter behaves. The model we proposed here uses Fourier series to fit the data in a single step rather than using the multistep process. Images courtesy of Simon Hunt and Danielle Fenech.

Online analysis tools for time-dependent data should support rapid assessment of data quality and sample evolution with increasing pressure. For example, real-time optical microscopy combined with XRD would facilitate a researcher's ability to evaluate sample evolution. Automated image analysis—of both the optical and diffraction images—could help isolate and identify unique pressure-dependent features in very large datasets. Online tools like these could be modular and part of a larger tool set that could be customized for various time-dependent measurement needs. Additional effort should be directed at batch processing XRD patterns using full-pattern techniques, visualization, profile matching, or even Rietveld refinement. Even modest advances in evaluation and analysis software for use both during and after time-dependent high-pressure experiments could greatly advance the field.

2.2 X-Ray Diffraction Analysis Tools

2.2.1 Background and State of the Field

Software packages available in the scientific community are used for analyzing data from XRD experiments. However, most software packages are not optimized for speed or usability but rather for technical capabilities. They often emerged in other communities, where good data quality can usually be ensured through careful experimental design. Unfortunately, XRD data from dynamic shock and DAC experiments often cannot provide data of comparable quality because of the short exposures in time-resolved measurements, which leads to an inevitable failure of the optimization routines in many of those packages.

A number of complete analysis packages for powder XRD exist, such as Material Analysis Using Diffraction (MAUD)[17] and General Structure Analysis System–II (GSAS-II).[18] These packages perform image integration and simple single pattern LeBail fits or Rietveld refinement of microstrains and textures as well as batch processing of multiple files. GSAS-II can process images on the fly as they are collected. The rate is limited to 1–2 seconds per image, but the user can view the evolution of the integrated powder patterns with changes in experimental conditions. The sequential analysis tools in GSAS-II allow fitting of sequential results (e.g., volumes) to simple analytical functions to extract things of interest (e.g., equation of state coefficients). Optimization tools using HPC (such as Spotlight,[19] which uses Mystic[20]) explore ranges of lattice parameters and weight fractions for phases provided by the user and then feed them into a GSAS language script for full Rietveld.

While these packages are powerful, they lack the speed necessary for fast visualization during the beamtime. Many users at different light sources therefore prefer other packages for online analysis, such as Dioptas,[21] PDIndexer,[22] or XDI.[23] They provide fast feedback with respect to phase content, pressure, and temperature of the sample and provide information valuable in deciding the next experimental steps. Dioptas provides multi-panel/multi-detector calibration and fitting. However, these tools lack the detailed publication-ready analysis. Post-beamtime data analysis is typically done through a combination of the software packages listed above or in combination with more specialized Rietveld programs (e.g., TOPAS,[24] JANA,[25] and FullProf[26]). Analysis of mixed phases through HEXRD[27] is also available. New features are being added regularly, such as the ability to do texture, Rietveld, Laue, unusual detector geometries, and more.

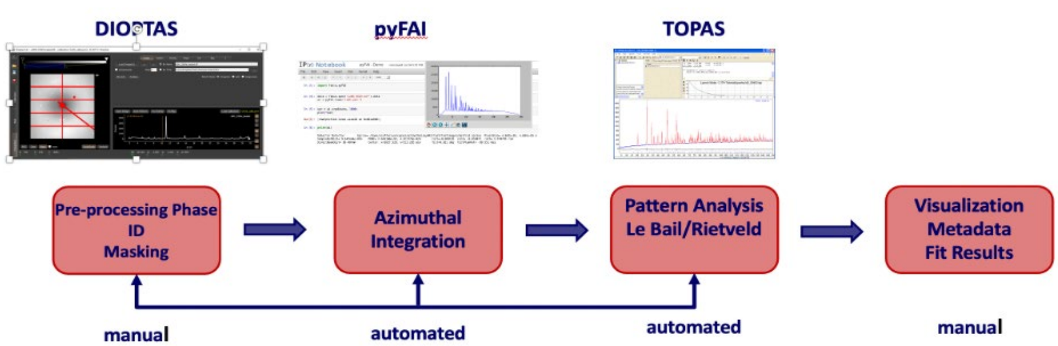


***Figure* 4. Data analysis approach for reaction of feldspars. Diagram courtesy of Lars Ehm.**

2.2.2 Needs Identified

There is currently no single software package for high-pressure powder XRD available that serves the needs for all use cases. Switching between different software for either online or offline, single file or batch processing, etc., is necessary. This can be cumbersome because of transferring data between software and exporting and importing images and data. The question is whether a single software package is actually necessary. A better interplay between software packages would reduce the amount of new code that needs to be written and also reduce the number of possible bugs and errors. In the conference chat, one workshop participant asked, "Why is it that beamlines develop code separately and not together more?" Open-source software was also praised several times for its ability to connect the community and make progress.

A major issue will be the data analysis and visualization of the thousands of diffraction images collected. A large amount of data collected requires an automated data processing pipeline for post hoc analysis. This is challenging, however, because algorithms are often too slow or not optimized for imperfect data; manual input is needed, which makes the processing tedious, especially when re-analysis is needed. One participant seriously questioned whether it is possible to "automate experiments to make effective use of these data pipelines, i.e. for diffraction and imaging and not only one sample at a time but multiple samples one after the other." It was also mentioned that without metadata (e.g., temperature, load, and pressure) associated with each dataset, it is impossible to produce sequential results.

A lack of general solutions was also mentioned, with one participant commenting that "solving a specific problem is always easier than solving the general problem … and communication is hard/takes time. In the short term, [creating] your own is faster, long term (I think) it is much slower."

2.2.3 Recommended Investment Areas

- ML as a pre-conditioner for current software input to reduce the manual input
- Improvement of automatic routines in terms of robustness, user friendliness, and adaptability
- Protocols for software interoperability as well as capabilities for being able to use existing software in an HPC environment (e.g., parallel analysis of the many images). The U.S. Department of Energy's Basic Energy Sciences (BES) is funding an effort to create common software tools called the Data Pilot Project (see Figure 5) for light source data analytic workflows in ptychography and tomography, however rapid and dynamic compression analyses are not included.

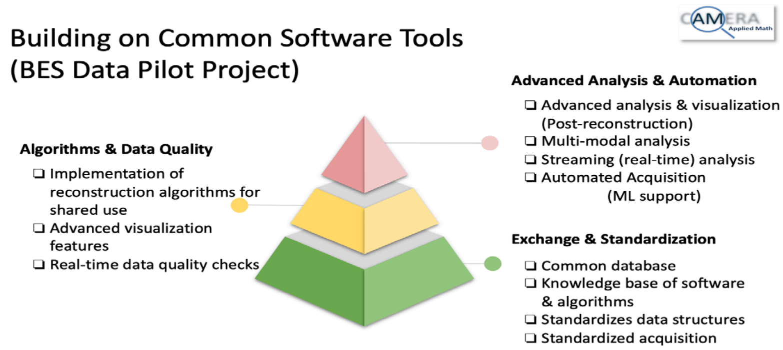


Figure 5. BES Data Pilot Project diagram. Image courtesy of Hari Krishnan (Lawrence Berkeley Laboratory).

3.0 DATA SCIENCE TECHNIQUES

3.1 Image Processing

3.1.1 Background and State of the Field

X-ray imaging allows for the observation of mesoscale phenomena that are difficult to investigate with other techniques. For example, imaging can reveal void collapse, complex wave interactions in shock compressed matter, phase transition, crack propagation, and nucleation. Direct imaging (i.e., absorption imaging) and phase contrast imaging (PCI) are both powerful tools. Phase retrieval analysis on PCI data allows for the extraction of properties such as density. The APS has dedicated graphics processing unit (GPU) clusters running accelerated phase retrieval codes at various coherent diffraction beamlines that keep up with data acquisition. Experimentalists use a continuum of techniques that ranges from absorption radiography to coherent diffractive imaging. The enhanced brightness and spatial coherence of recent and future light sources are making measurements with these techniques feasible during dynamic high-pressure events (i.e., single event imaging) that use drivers such as gas guns, laser compression, and dDAC.[2]

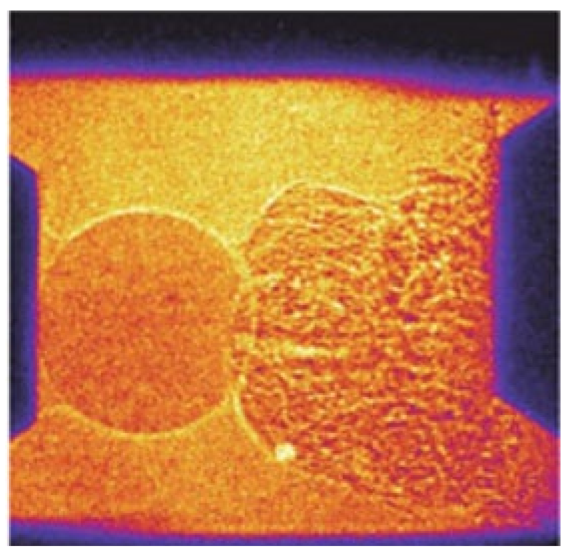


Figure 6. Dynamic multi-frame X-ray phase contrast imaging (XPCI) consisting of a steel projectile impacting three BS spheres at a velocity of 0.237 km/s. The times relative to impact are shown in the false color, where black represents complete absorption of the X-rays. Experiment at APS.[28]

3.1.2 Needs Identified

Challenges for imaging algorithms include phase wraps through a given sample, noise in the data (whether from the source or detector), and spatial resolution from the experiment geometry (i.e., source size, divergence, artifacts from X-ray beamline windows, and sample-to-detector distances). Information to allow on-the-fly adjustments to mitigate such problems is vital.

Some analysis tools are being used, though many are not readily available to the community at large and mostly exist as locally developed codes. An example is contrast transfer function (CTF) phase retrieval with advanced regularization techniques to account for noise looks. The technology looks promising but is not available.

Forward calculation capabilities are not available, but they have the potential to aid experiment design when used with imaging. Forward models need full diffraction calculation, energy-dependent complex indices, X-ray energy spectral response, and the ability to correct for blur, lens features, and more.

A new larger volume of data is anticipated as a result of the high repetition rate of pulsed lasers, and this will require a better workflow for handling these data and data science methods for accurate phase retrieval that incorporate experimental challenges.

3.1.3 Recommended Investment Areas

Workshop participants suggested a prioritized investment in imaging algorithms (in terms of impact and necessary effort) towards different levels of analysis. These items included

1. digital image analysis to track and quantify discrete features or changes in texture in the digital images as they propagate (e.g., during shock loading),
2. phase retrieval to recover a complex index of refraction and thus density,
3. recovery of size distributions from analysis of the scatter represented in contrast texture in images, and
4. a physics model–based interpretation of sparse datasets (e.g., single-event imaging).

Research is needed for real-time coherent diffractive imaging inversion (or similar) and ptychography for use during compression experiments. These lensless reconstruction techniques can be applied through iterative phase retrieval, which can be computationally expensive, and through new and potentially faster ML techniques[29] that use deep generative networks.

Algorithm acceleration efforts are also necessary to enable imaging analysis to scale up to 10 Hz repetition rates enabled by DiPOLE at the European XFEL or the Materials in Extreme Conditions (MEC) upgrade at the Linear Accelerator (Linac) Coherent Light Source (LCLS) or even for dDAC experiments at these facilities. Hardware acceleration through GPUs is a promising technique for imaging algorithms, and HPC clusters could be used more widely.

Combining PCI measurements with simultaneous or complementary XRD measurements will also provide a powerful tool and allow more simplistic phase extraction. Examples include simultaneous imaging and diffraction at free electron laser (FEL) facilities[30] and imaging and diffraction in separate experiments.[31]

3.2 Advanced Analysis, Modeling, and Machine Learning for Compression Workflows

3.2.1 Background

A number of advanced techniques for analysis are available, and new techniques are currently being researched. Analysis and simulation of XRD have been done with GSAS and GSAS-II[14,32,33] for a number of years, while new features, such as analysis of crystal structure refinements for materials at very high pressures via pink beam, are being added. Analysis of mixed phases via HEXRD[27] is also available, and new features are being added regularly. Advanced analytics workflows that use a combination of forward models, such as Flag[34] and HYADES,[35] emulators, and uncertainty quantification have shown promise[36] in comparing to experimental data and helping find model parameters (see Figure 7).

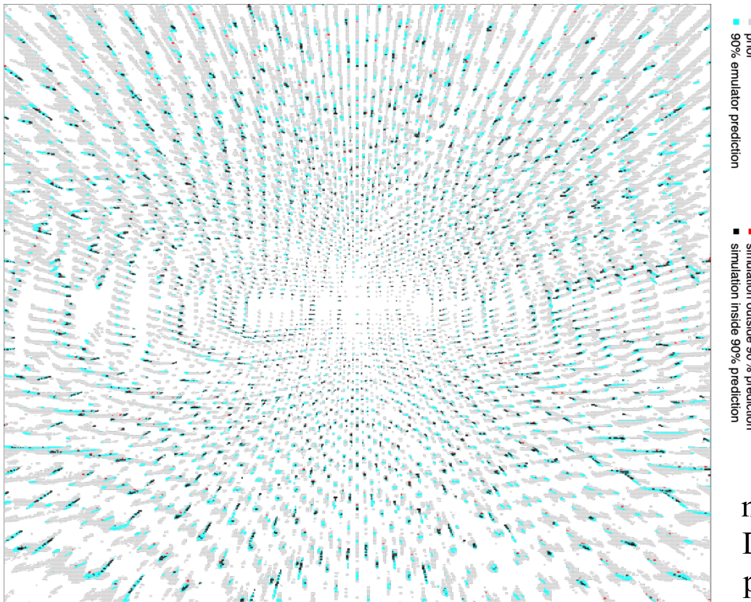


Figure 7. Diffraction image holdout prediction shows an accurate prediction by Flag with the DiscoFlux hydrocode emulator and BarberShop XRD simulator. Gray is prior, and black is the simulation inside 90% prediction. Red is the simulation outside 90% prediction. Small blue regions are 90% emulator prediction.[36]

Modern ML [via convolutional neural networks (CNNs)][37] and previous multidimensional scaling (MDS)[38] methods are being used to classify crystal structures for neutron diffraction. A feature detection algorithm based on image processing and unsupervised ML (clustering) techniques can identify individual grains and their size distribution;[39] training data are generated with molecular dynamics simulations. Deep generative networks for ptychographic imaging[29] are also being researched and could potentially be used in the domain of dynamic compression to quickly provide reconstructed images of materials of interest.

3.2.2 Needs Identified

Some needs exist in the areas of obtaining data, going beyond the limitations of current models, and handling more complex analyses. Obtaining data for advanced analytics and ML can be difficult because few models are available in the rapid/dynamic compression domain. For example, models for laser-driven shock are not available. Also, some of the models may be quite computationally expensive and time-consuming to run, especially at the molecular dynamics level. Another shortcoming is that some models can provide more than one prediction and are therefore not unique. Lastly, scientists would often like to go beyond forward models and analyze and interpret texture and orientation data. They want fast prediction of internal parameters of crystal structures. In some cases, they are limited by experimental data and cannot analyze complex structures. For example, the small width of the pink beam energy distribution limits analysis to very simple structures. Figure 8 shows some efforts toward analysis of pink beam diffraction.

3.2.3 *Recommended Investment Areas*

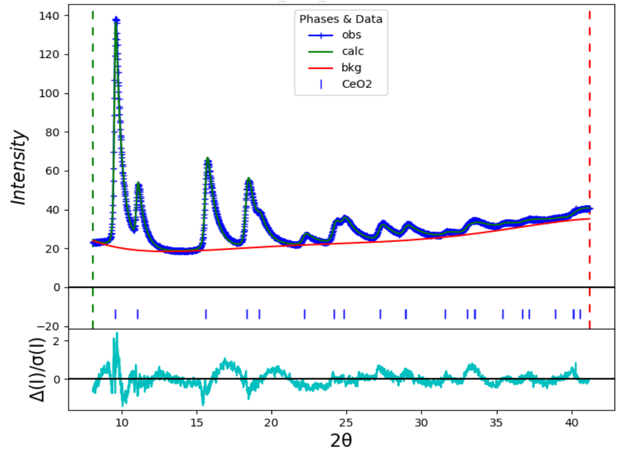


Figure 8. A profile function for pink beam diffraction used in single-peak fits and in a Rietveld refinement for CeO_2. This feature has been added to the GSAS-II package. Image courtesy of R. Von Dreele.

A suggested research area includes the development of data science techniques. One way to tackle the data problem is by solving inverse problems using experimental data (rather than simulations) and performing sensitivity analysis. Emulator research also could produce more accurate uncertainty and handle functional input (using the deformation gradient as input). Multi-modal data analysis techniques could be used by, for example, including combining simulation with experiment data and the experimental timeline of X-ray probe time, translation offset, and edge wave timing.

Continued development of forward models is another area of potentially great effect, especially when combined with data science and ML techniques. It was suggested that if a user can include the spectral width and divergence of the X-ray source in a forward model of diffraction, we might be able to better resolve features from a given diffraction pattern; however, this requires a sophisticated inverse problem solver. The problem of determining mixed phases could also be helped with forward models. The simpler problem of indexing (identifying the correct lattice and unit cell parameters) from a sharp powder diffraction pattern is essentially impossible for samples consisting of more than a single phase. Even for one-phase samples, indexing can sometimes be surprisingly difficult.

ML research areas include expanding current capabilities and quickly predicting internal parameters of crystal structure (e.g., lattice lengths, lattice angles, and thermal parameters) and using these quantities in more advanced analyses such as texture analysis. As previously mentioned, ML is also a promising area for imaging. Forward modeling, ML, and/or advanced analysis is needed to include mixed phases (possibly in combination with deviatoric stress) or mixed crystallographic states and to determine how much of the probed region is in one phase versus the other. Changes to the beam characteristics, such as sharper pink beam distribution, could allow for analysis of more complex materials if enough sample data are available.

3.3 Visualization

3.3.1 *Background and State of the Field (with citations to publications/code)*

Interdisciplinary efforts between scientists and visualization researchers have produced workflows to support real-time data visualization needs and post hoc data exploration. The resultant tools address the heterogenous, high-dimensional data typical of dynamic compression experiments. Image-based techniques such as computer vision and ML have also proven useful. Several tools have been based on Cinema, a database approach to the exploration of high-dimensional data.[40] Cinema:Bandit[41] seen in Figure 9 is a specialized Cinema tool designed for a continuous beamline workflow. It brings together VISAR images, velocimetry profiles, XRD images, and XRD patterns linked by a parallel coordinates plot that allows the user to subselect the data. Cinema:Bandit was used in a runtime shock physics workflow at LCLS. Cinema:Debye-Scherrer[42] similarly links run parameters to individual Rietveld plots, allowing the user to quickly find outliers and patterns. Cinema:Quest[43] allows the user to explore dimensionally reduced input and output parameter space through an interactive drag-and-track mechanism. Researchers at Brookhaven National Laboratory (BNL) have developed several visualization tools to support beamline workflows. MultiSciView[44] allows a user-defined layout with views such as image pixels, a parallel coordinates plot, and data selection tools. The various views allow the user to drill into the data to find anomalies and interactively explore a large image dataset. This tool has been integrated into in situ workflows at the Center for Functional Nanomaterials (CFN) and Complex Materials Scattering (CMS) beamlines at the National Synchrotron Light Source–II (NSLS-II). Other work includes tools that incorporate ML (such as VisCNN[45]) to support feature detection and modelling and data-driven color mapping for X-ray images[46] (this also in use at NSLS-II at the Hard X-ray Nanoprobe beamline).

3.3.2 Needs Identified

As mentioned in Section 3.1, real-time analysis tools are needed to affect runtime decisions. Cinema:Bandit and MultiSciView have each been successfully used at beamlines. However, useful applications have not always migrated to common use at beamlines. Although these modular component–based approaches can potentially support different experimental workflows at multiple facilities, they are often viewed as one-off applications. The ability to bring beamline metadata into the data output stream and into real-time visualization tools to support run decision-making is also a significant issue. Awareness of the available tools and integration into the diverse beamline software stacks are both needed to make these tools more widely available.

3.3.3 Recommended Investment Areas

The development of the aforementioned tools relied heavily on interdisciplinary teams with visualization researchers embedded into experimental teams. Increased funding will be needed to support this type of interdisciplinary approach. Research is needed to extend these efforts to develop additional components for a wider range of experimental workflows. The continued strategic development of prototype tools that support quantitative visual comparison between models and experimental data is another area that could significantly affect scientific insights. Investment in software engineering is needed to incorporate these tools into the beamline software stack.

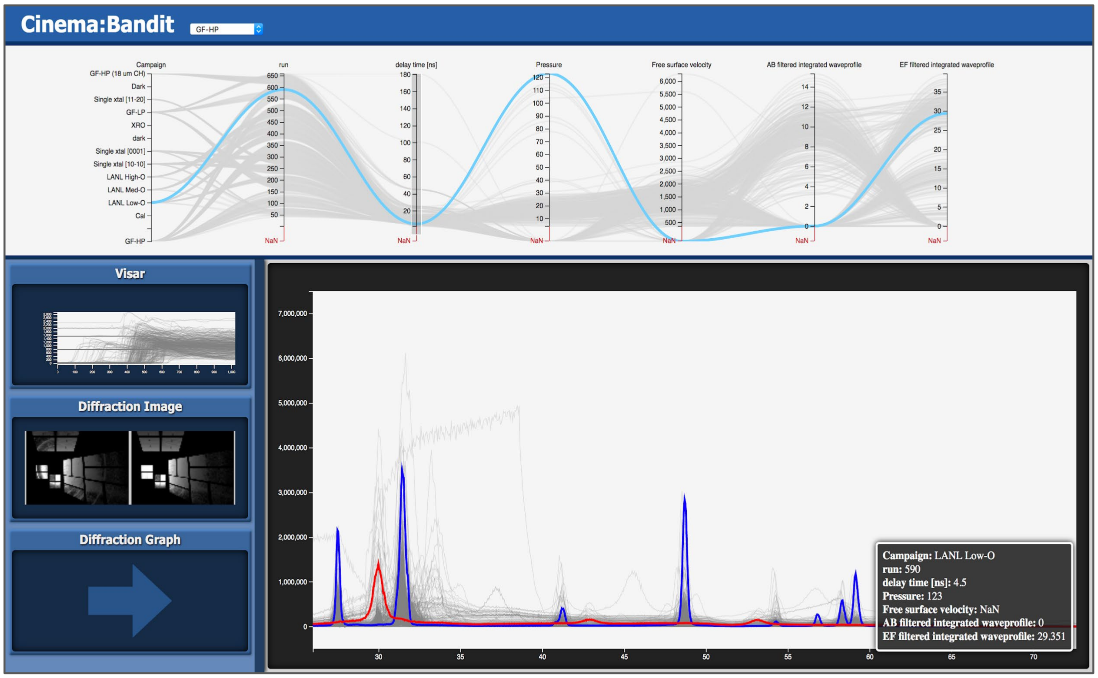


Figure 9. The Cinema:Bandit visualization tool has a parallel coordinates plot (top) that links the data across the various view panels. The main view screen (bottom right) can view any of the three sets of data (left column of views): VISAR, diffraction images, or diffraction graph.[41]

4.0 COMPUTATIONAL WORKFLOWS AND FACILITY SUPPORT

4.1 Real-Time Analytics Workflows and Computation

4.1.1 Background

Frameworks have been developed to facilitate real-time data management, processing, and visualization at beamline facilities.[47,48,49,50,51] These frameworks provide libraries for data acquisition and loading because hardware, data formats (e.g., heterogeneous data format [HDF], Nexus, tagged image format file [TIFF], or crystallographic binary file [CBF]), and metadata vary across facilities. After the data are captured, a data broker organizes the data and associated metadata so it is available at collection time. This abstraction decouples the data acquisition and formats from the analysis code, allowing scientists to share analyses between experiments and beamline facilities.

One common application of these frameworks is to monitor data quality, which allows scientists to make immediate decisions. The Online Monitor framework[52] has been deployed in this capacity at LCLS and the synchrotron light source PETRA III at Deutsches Elektronen-Synchrotron (DESY) for responding to experiments where the sample was missed, the injector was clogged, or the injector liquid jet moved.[50] Domain knowledge can be added to these monitors to create automated alarms or controls that alert the scientists or make small corrections. Adaptive logic is an emerging research area, and frameworks such as Adaptive BlueSky[53] shown in Figure 10 have been designed to apply adaptive logic and steer experiments. An early example of Adaptive BlueSky's capabilities was to efficiently sample the profile of Gaussian beams. Discussion between computational and experimental scientists are ongoing to identify use cases for applying adaptive logic during experiments.

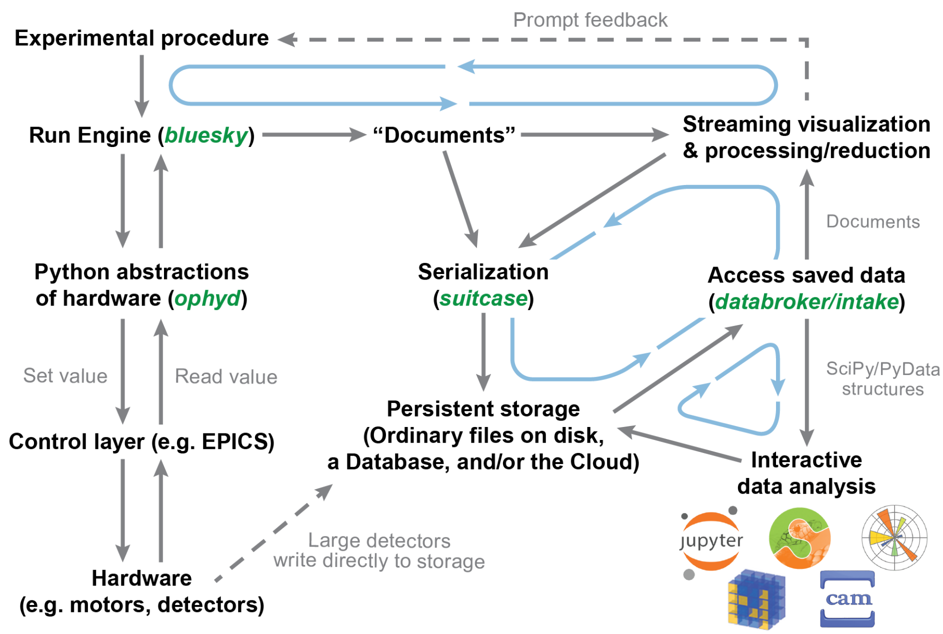


Figure 10. BlueSky infrastructure for experimental data collection and data management.[53]

4.1.2 Needs Identified

These frameworks for real-time analyses are executed at beamline facilities close to the detectors and streamed data sources or local data storage for a quick response time. However, access to these resources may require the user to physically be at the facility. Therefore, efforts are ongoing to develop support for remotely executing and monitoring workflows within these frameworks. Aside from remote connections, there are two additional use cases identified: (1) developing, validating, and reproducing the analysis post hoc; and (2) performing computationally intensive analyses that require data transfer to remote sites that can perform the analysis. Tools such as Globus[54,55] are currently used to transfer data, and there has been some research in remote transfer using BlueSky;[56] however, more support to handle these additional use cases would increase applicability.

An emerging application of this research area is using information to direct the experiment in real-time. To accomplish this, there needs to be an increase in the communication and synergy between computational and experimental scientists. Computational scientists need to understand what information is useful during an experiment and then codify domain-knowledge in logic to create alarms, controls, and/or visualizations that help steer experiments.

4.1.3 Recommended Investment Areas

In practice, it can be time-consuming to implement a backend at a new facility to collect data, as there may be unique and/or specific requirements to access the data acquisition system, computing resources, or hardware. The development entails working with facility personnel to understand the data acquisition methods unique to the facility. This knowledge is compartmentalized across projects and the effort is duplicated for each framework.

Investment in a shared data acquisition or data broker library between frameworks would provide a single, user-friendly interface for querying data and metadata from various facilities that would unify this knowledge. This would allow frameworks to leverage previous work in the community to accelerate the implementation of analyses at facilities. Investment in this technology and increased communication between computational and experimental scientists would foster a collaborative community for real-time analytics.

4.2 Facility Computational Resources

4.2.1 Background

Previous workshops on HPC and exascale computing for experimental workflows have shown a need for development in the area of HPC and experimental data analysis.[57] The National Energy Research Scientific Computing Center (NERSC), a supercomputing facility located geographically near the Advanced Light Source (ALS), is part of the Superfacility Project (funded by the Office of Science) and is crafting tools and interfaces and is performing community outreach to meet a substantial experimental workflow user base (35% of NERSC users), as shown in Figure 12. As a light source user facility with both synchrotron and XFEL sources, the Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory has a solid base of experience with experimental workflows and, in anticipation of the LCLS-II upgrade, has performed extensive research on data reduction and planning for a local science data facility. The EuXFEL is a newer facility with basic computing infrastructure and convenient ties to an established DESY synchrotron experimental analysis computational infrastructure.[58,59] Amazon Web Services (AWS) is well-established in cloud computing for the commercial domain and is working toward a stronger connection with the scientific domain by hosting datasets and notebook-based access to data[60] and by connecting with notebook-based access to data.[61]

Figure 11. Descriptions of the Superfacility Initiative. Image courtesy of Debbie Bard.

The workshop also highlighted other data sharing and software efforts. The Materials Project[62] hosts data with a direct access application programming interface (API) and includes 500,000 K-edge X-ray absorption near edge structure (XANES) spectra for over 50,000 materials. They also have data for XRD in all phases. They promote the use of open-source software for analysis of the data.[63] The Basic Energy Sciences (BES) Data Pilot Project was also highlighted for its work in connecting standardized software stacks for three experiment acquisition and analysis workflows (X-ray photon correlation spectroscopy [XPCS], ptychography, and tomography) with user facilities services such as the Superfacility as well as data storage and archive.

4.2.2 Needs Identified

Computing facility policy, authentication and federation, and human steps in the experimental analytics workflow thwart automated workflows. Data volumes and velocity are increasingly requiring inline data reduction, data management, metadata generation, and fast approximate analysis through new techniques such as ML. Computationally sophisticated real-time analyses are needed in a regime of fast feedback for feature extraction and data quality measurements. There was interest on the part of a number of participants in learning more about GPUs and how they could help with faster analytics. More research on the specifics of workflows and scientists requirements is needed to support many experimental regimes and a diverse set of users. Finally, mobile workflow support is needed so that scientists can run experimental data analysis at each facility without having to redesign the workflow each time.

4.2.3 Recommended Investment Areas

Recommended areas of investment for computing facilities are workflows, data management, and user engagement. Users would like clean protocols and interfaces for analytics and workflow software (e.g., scheduling, reservations, launching workflows, monitoring and optimizing job status and results). In the area of data management, research is needed for tools to handle data volumes for local processing, fast feedback with data reduction, and processing close to the detector (edge). Data migration is an area of investment to move data fast by means of transparent data transfer and efficient network allocation from user facilities to remote processing. Data repositories are another relevant data management area for research to better understand and use ways to deposit and curate data in various locations, including the cloud, along with container toolsets that allow processing of data anywhere. Data provenance and metadata are key to preserving the context in which data were taken and in enabling analysis, especially if data are made available to others through data platforms. Lastly, a major area of investment identified continues to be user engagement. User partnerships, integration between computing and light source facilities, and increased engagement with cloud providers are key to making better use of computing resources. Initiatives like the Materials Genome Initiative (MGI) also support data science research and development (R&D) infrastructure for materials discovery.[64]

5.0 PRIORITY GAPS

The workshop identified several priority gaps that crosscut many, if not all, of the sessions.

- Although upgraded facilities provide more opportunities for scientific discovery than ever before, current software and data analytic workflows for compression science cannot handle the new large data volumes and velocities. It is clear that scientists are thrilled to have the additional time points in their data series, a larger field of view, and a higher image resolution. However, they currently do not have the software stacks and algorithms for data management and cannot fully leverage the power of this newfound data.
- Data analysis tools available can perform specific tasks but are often manual and usually too time-intensive to be able to help inform the current experiment. With scarce beam times, scientists are seeing missed opportunities when they post-process data. The small-scale analytics do not go far enough to enable experimental design during a beam time and do not harness the computing power that is available for faster analytics.
- Compression science data characteristics present challenges for analytics. Noise, phase wraps, spatial resolution, data sparseness, and artifacts in data impede analyses such as peak finding and imaging. Scientists are aware that overcoming these challenges could lead to some of the largest scientific breakthroughs.
- For data science to be most effective, dataset availability is key. Compression science, because of its experimental nature, has few labelled datasets. It is therefore difficult to obtain training and simulation data needed for ML, statistical data science, and experimental data analytics supplementation. Models can typically provide this; however, this field lacks a rich set of suitable models and shortcomings exist in some of the models that are currently available.
- As a corollary to dataset availability, metadata generation, management, and availability are also key; however, metadata capabilities are sorely lacking in the area of compression science data. This thwarts data science and data management.
- Scientists would like to be able to conduct experiments remotely more often and to harness resources from multiple facilities; however, these capabilities are just beginning to become available. Much more needs to be done to make the process seamless for scientists and make the analytics portable across facilities.

6.0 PRIORITY INVESTMENTS

- Algorithm development is a top priority for analytics like imaging, ML, and signal processing. Analytics are needed that are integrated real-time into the toolchain to guide decision making, close to the detector in some cases, so that features are detected and data are reduced, and that include visualization so that the user is involved in the analytics.
- Investments need to be made in more universally applicable and accessible software tools. As with the BES Data Pilot project, compression science also needs tools with common interfaces between them that can provide the ability to load user-made or niche "plugins" useful to a specific experiment. Scientists anticipate that this solution will be more agile and responsive to individual needs than monolithic tools.
- Integration of multiple streams of data would lead to a better diagnostic capability for compression science. Scientists want to combine multiple techniques and data types (e.g.,

combine data from different detectors such as diffraction and imaging and combine modeling and analytics).

- Metadata creation and data management are required if we are going to treat compression science data as first class. Data need to have the same pedigree despite coming from different data streams. Mechanisms like data brokers and shared data acquisition are needed to help scientists properly tag data with basic metadata such as sample configuration and positions and to place the data in accessible repositories. Connections need to be made between metadata, data repositories, and real-time tools and visualization.
- Both science and computing user facilities need to invest in user engagement and interdisciplinary teams. In this way, scientists can articulate needs for different regimes and work with software developers and computational scientists to enable the workflows that are needed.
- Algorithms and data storage infrastructure are needed for data processing and reduction to handle increased volumes and velocity of data.
- Forward model development for various compression regimes is needed and should be combined with data analytic techniques. Many current models need further development to match experiment requirements more closely so that they can be used for accurate prediction.
- Interfacility or facility-to-remote laptop workflows that support remote and collaborative experiment decision making, data analytics, and processing need much greater investment. New compression science software stacks and interfaces need to connect to these workflows and do so in a portable way.

7.0 SUMMARY AND SUGGESTED ACTIVITIES

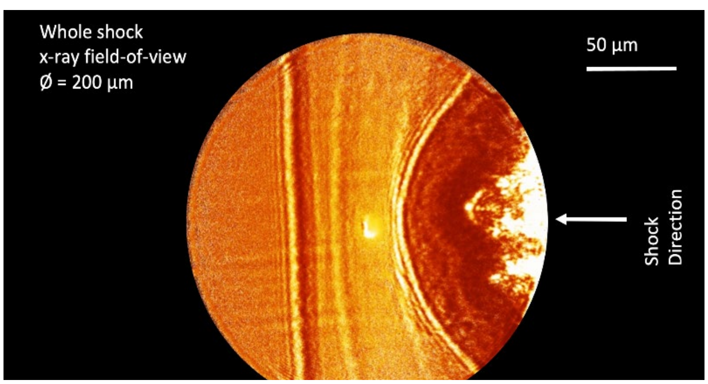


Figure 12. Direct imaging of ultrafast lattice dynamics.[31]

The workshop gathered a widely representative group of compression science experts, data scientists, and computer scientists and included a number of deep-dive discussion sessions on compression science analytics requirements, analytics tools, advanced analytics and workflows, and facility computational support. The workshop was successful in going beyond the discussion of generalities of compression science data analytics and honed in on specific use cases and deficiencies that warrant further attention and investment.

Priority needs were identified for large data processing, handling unique data characteristics, and in a lack of training data, metadata, forward models, real-time tools, and remote experiment tools. Investment areas were identified for algorithm development, software tools, data integration, metadata creation, user engagement, forward models, and interfacility workflows.

The community seems especially engaged and ready to collaborate to tackle some of the identified challenges. Many people indicated that a number of useful tools exist and that some of the challenges seem doable with modifications to tools available, while other challenges will require considerable investment into research on novel solutions or will require community coordination to settle on mutually beneficial abstractions for software interfaces and integration. Suggested activities are to increase awareness of community data analytics needs through collaboration and to continue workshops at user facility meetings, professional society meetings, or other relevant venues.

ACKNOWLEDGMENTS

This workshop was sponsored by Los Alamos National Laboratory through Richard Sheffield, the Institutional Host.

Thank you to the steering committee for helping to bring the workshop idea to fruition and thank you to all who attended, contributed to discussions, gave presentations, and helped write this report.

REFERENCES

1. L. Dubrovinsky, N. Dubrovinskaia, V.B. Prakapenka, and A.M. Abakumov, "Implementation of micro-ball nanodiamond anvils for high-pressure studies above 6 Mbar," *Nat. Commun. 3(1163)*, doi:10.1038/ncomms2160 (2012).
2. Z. Jenei, H.P. Liermann, R. Husband, A.S.J. Méndez, D. Pennicard, H. Marquardt, E.F. O'Bannon, A. Pakhomova, Z. Konopkova, K. Glazyrin, M. Wendt, S. Wenz, E.E. McBride, W. Morgenroth, B. Winkler, A. Rothkirch, M. Hanfland, and W.J. Evans, "New dynamic diamond anvil cells for tera-pascal per second fast compression x-ray diffraction experiments," *Rev. Sci. Instrum. 90(6)*, 065144 (2019).
3. B. Nagler, B. Arnold, G. Bouchard, R.F. Boyce, R.M. Boyce, A. Callen, M. Campbell, R. Curiel, E. Galtier, J. Garofoli, E. Granados, J. Hastings, G. Hays, P. Heimann, R.W. Lee, D. Milathianaki, L. Plummer, A. Schropp, A. Wallace, M. Welch, W. White, Z. Xing, J. Yin, J. Young, U. Zastrau, and H.J. Lee, "The Matter in Extreme Conditions Instrument at the Linac Coherent Light Source," *J. Synchrotron Radiat.* 22, 520–525 (2015).
4. D. Capatina, K. D'Amico, J. Nudell, J. Collins, and O. Schmidt, "DCS—A high flux beamline for time resolved dynamic compression science—Design highlights," *AIP Conf. Proc* 1741, 030036 (2016).
5. E.B. Watkins, K.A. Velizhanin, D.M. Dattelbaum, R.L. Gustavsen, T.D. Aslam, D.W. Podlesak, R.C. Huber, M.A. Firestone, B.S. Ringstrand, T.M. Willey, M. Bagge-Hansen, R. Hodgin, L. Lauderbach, T. van Buuren, N. Sinclair, P.A. Rigg, S. Seifert, and T. Gog, "Evolution of Carbon Clusters in the Detonation Products of the Triaminotrinitrobenzene (TATB)-Based Explosive PBX 9502," *J. Phys. Chem. C 121(41)*, 23129–23140 (2017).
6. T.S. Duffy and R.F. Smith, "Ultra-High Pressure Dynamic Compression Materials," *Front. Earth. Sci. 7(23)*, doi:10.3389/feart.2019.00023 (2019).
7. Z. Wang, M. Demarteau, and J. Kline, "The Special Issue of NIMA for ULITIMA 2018," *Nucl. Instrum. Methods Phys. Res., Sect. A* 955, 163245 (2020).
8. Miceli, "Application of Pixel array detectors at X-ray synchrotrons," *J. Instrum.* 4, doi:10.1088/1748-0221/4/03/P03024 (2009).
9. S. Banerjee, K. Ertel, P.D. Mason, P.J. Phillips, M. De Vido, J.M. Smith, T.J. Butcher, C. Hernandez-Gomez, R. J. S. Greenhalgh, and J.L. Collier, "DiPOLE: a 10 J, 10 Hz cryogenic gas cooled multi-slab nanosecond Yb:YAG laser," *Opt. Express 23(15)*, 19542–19551 (2015).
10. S.V. Sinogeikin, J.S. Smith, E. Rod, C. Lin, C. Kenney-Benson, and G. Shen, "Online remote control systems for static and dynamic compression and decompression using diamond anvil cells," *Rev. Sci. Instrum. 86(7)*, 072209 (2015).
11. J.S. Smith, S.V. Sinogeikin, C. Lin, E. Rod, L. Bai, and G. Shen, "Developments in time-resolved high pressure x-ray diffraction using rapid compression and decompression," *Rev. Sci. Instrum. 86(7)*, 072208 (2015).
12. L.M. Barker and R.E. Hollenbach, "Laser interferometer for measuring high velocities of any reflecting surface," *J. Appl. Phys. 43(11)*, 4669–4675 (1972).

13. S.J. Tracy, R.F. Smith, J.K. Wicks, D.E. Fratanduono, A.E. Gleason, C.A. Bolme, V.B. Prakapenka, S. Speziale, K. Appel, A. Fernandez-Pañella, H.J. Lee, A. MacKinnon, F. Tavella, J.H. Eggert, and T.S. Duffy, "*In situ* observation of a phase transition in silicon carbide under shock compression using pulsed x-ray diffraction," *Phys. Rev. B* 99, 214106 (2019).
14. B.H. Toby and R.B. Von Dreele, "*GSAS-II*: the genesis of a modern open-source all purpose crystallography software package," *J. Appl. Crystallogr.* 46, 544–549 (2013).
15. L. Lutterotti, S. Matthies, and H. Wenk, "MAUD: a friendly Java program for material analysis using diffraction," *IUCr: Newsletter of the CPD* (1999).
16. C. Prescher and V.B. Prakapenka, "DIOPTAS: a program for reduction of two-dimensional X-ray diffraction data and data exploration," *High Pressure Res. 35(3)*, 223–230 (2015).
17. Radiographema, "MAUD: Materials Analysis Using Diffraction," http://maud.radiographema.eu/ (2020).
18. Trac, "GSAS-2," https://subversion.xray.aps.anl.gov/trac/pyGSAS (2020).
19. GitHub, "Spotlight," https://github.com/lanl/spotlight (2020).
20. GitHub, "Mystic," https://github.com/uqfoundation/mystic (2020).
21. Clemens Prescher, "Dioptas," http://www.clemensprescher.com/programs/dioptas (2020).
22. GitHub, "PDIndexer," https://github.com/seto77/PDIndexer (2020).
23. R. Hrubiak, J.S. Smith, and G. Shen, "Multimode scanning X-ray diffraction microscopy for diamond anvil cells," *Rev. Sci. Instrum. 90(2)*, 025109 (2019).
24. Bruker, "TOPAS Software," https://www.bruker.com/products/x-ray-diffraction-and-elemental-analysis/x-ray-diffraction/xrd-software/topas.html (2020).
25. Institute of Physics, "JANA," http://jana.fzu.cz/ (2020).
26. Institut Laue-Langevin, "FullProf Suite," https://www.ill.eu/sites/fullprof/ (2006).
27. HEX2020 [27] GitHub, "heXRD: Highly Extensible X-Ray Diffraction Toolkit," https://github.com/hexrd (2020).
28. B.J. Jensen, D.S. Montgomery, A.J. Iverson, C.A. Carlson, B. Clements, M. Short, and D.A. Fredenburg, "X-Ray Phase Contrast Imaging of Granular Systems," in *Shock Phenomena in Granular and Porous Materials*, T.J. Vogler and D.A. Fredenburg, Eds. (Springer Nature Switzerland, 2019).
29. M.J. Cherukara, T. Zhou, Y. Nashed, P. Enfedaque, A. Hexemer, R.J. Harder, and M.V. Holt, "AI-enabled high-resolution scanning coherent diffraction imaging," *Appl. Phys. Lett. 117(4)*, 044103 (2020).
30. F. Seiboth, L.B. Fletcher, D. McGonegle, S. Anzellini, L.E. Dresselhaus-Cooper, M. Frost, E. Galtier, S. Goede, M. Harmand, H.J. Lee, A.L. Levitan, K. Miyanishi, B. Nagler, I. Nam, N. Ozaki, M. Rödel, A. Schropp, C. Spindloe, P. Sun, J.S. Wark, J. Hastings, S.H. Glenzer, and E.E. McBride, "Simultaneous 8.2 keV phase-contrast imaging and 24.6 keV X-ray diffraction from shock-compressed matter at the LCLS," *Appl. Phys. Lett.* 112, 221907 (2018).
31. S. Brennan Brown, A.E. Gleason, E. Galtier, A. Higginbotham, B. Arnold, A. Fry, E. Granados, A. Hashim, C.G. Schroer, A. Schropp, F. Seiboth, F. Tavella, Z. Xing, W. Mao,

H.J. Lee, and B. Nagler, "Direct imaging of ultrafast lattice dynamics," *Sci. Adv. 5(3)*, eaau8044 (2019).
32. R.B. Von Dreele, J.D. Jorgensen, and C.G. Windsor, "Rietveld refinement with spallation neutron powder diffraction data," *J. Appl. Crystallogr.* 15, 581–589 (1982).
33. H. Xu, P.J. Heaney, D.M. Yates, R.B. Von Dreele, and M.A. Bourke, "Structural mechanisms underlying near-zero thermal expansion in β-eucryptite: A combined synchrotron x-ray and neutron Rietveld analysis," *J. Mater. Res. 14(7)*, 3138–3151 (1999).
34. J.L. Hill, "(U) User's Manual for FLAG version 3.6.0," Los Alamos National Laboratory document LA-CP-17-20057 (February 24, 2017).
35. Cascade Applied Sciences Inc., "HYADES," https://casinc.com/hyades.html (2015).
36. D. Francom, D.J. Walters, J.L. Barber, D.J. Luscher, E. Lawrence, A. Biswas, C.M. Biwer, D. Banesh, J. Lazarz, S.C. Vogel, K. Ramos, C. Bolme, R.L. Sandberg, and J. Ahrens, "Simulation and Emulation of X-Ray Diffraction from Dynamic Compression Experiments," *J. Dynamic Behavior Mater.*, doi:10.1007/s4870-020-00254-8 (2020).
37. C. Garcia-Cardona, R. Kannan, T. Johnston, T. Proffen, K. Page, and S.K. Seal, "Learning to Predict Material Structure from Neutron Scattering Data," in *2019 IEEE International Conference on Big Data* (IEEE, Los Angeles, CA, December 9–12, 2019) pp. 4490–4497.
38. A. Mead, "Review of the Development of Multidimensional Scaling Methods," *Journal of the Royal Statistical Society, Series D (The Statistician) 41(1)*, 27–39 (1992).
39. H. Chan, M.J. Cherukara, B. Narayanan, T.D. Loeffler, C. Benmore, S.K. Gray, and S.K.R.S. Sankaranarayana, "Machine learning coarse grained models for water," *Nat. Commun. 10(379)*, doi:10.1038/s41467-018-08222-6 (2019).
40. J. Ahrens, S. Jourdain, P. O'Leary, J. Patchett, D.H. Rogers, and M. Petersen, "An Image-Based Approach to Extreme Sacle *In Situ* Visualization and Analysis," in *SC' 14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE, New Orleans, LA, November 16–21, 2014), pp. 424–434.
41. D. Obran, D. Banesh, C. Tauxe, C.M. Biwer, A. Biswas, R. Saavedra, C. Sweeney, R.L. Sandberg, C.A. Bolme, J. Ahrens, and D. Rogers, "Cinema:Bandit: a visualization application for beamline science demonstrated on SFEL shock physics experiments," *J. Synchrotron Radiat.* 27, 1–10 (2020).
42. S.C. Vogel, C.M. Biwer, D.H. Rogers, J.P. Ahrens, R.E. Hackenberg, D. Onken, and J. Zhang, "Interactive visualization of multi-data-set Rietveld analyses using *Cinema:Debye-Scherrer*," *J. Appl. Crystallogr. 51(3)*, doi:10.1107/S1600576718003989 (2018).
43. D. Orban, D.F. Keefe, A. Biswas, J. Ahrens, and D. Rogers, "Drag and Track: A Direct Manipulation Interface for Contextualizing Data Instances within a Continuous Parameter Space," *IEEE Transactions on Visualization & Computer Graphics 25(1)*, 256–266 (2019).
44. W. Zhong, W. Xu, K.G. Yager, G.S. Doerk, J. Zhao, Y. Tian, S. Ha, C. Xie, Y. Zhong, K. Mueller, and K.K. Van Dam, "MultiSciView: Multivariate Scientific X-ray Image Visual Exploration with Cross-Data Space Views," *Visual Informatics* 2, 14–25 (2018).

45. X. Huang, S. Jamonnak, Y. Zhao, B. Wang, M, Hoai, K. Yager, and W. Xu, "Interactive Visual Study of Multiple Attributes Learning Model of X-Ray Scattering Images," *IEEE Transactions on Visualization & Computer Graphics 2020*, arXiv:2009.02256 (2020).
46. S. Cheng, W. Xu, and K. Mueller, "ColorMapND: A Data-Driven Approach and Tool for Mapping Multivariate Data to Color," *IEEE Transactions on Visualization & Computer Graphics 2018* 24, doi:10.1109/TVCG.2018.2808489 (2018).
47. A. Arkilic, D.B. Allan, T. Caswell, L. Li, K. Lauer, and S. Abeykoon, "Towards Integrated Facility-Wide Data Acquisition and Analysis at NSLS-II," *Synchrotron Radiation News 30(2)*, 44–45 (2017).
48. A. Barty, R.A. Kirian, F.R.N.C. Maia, M. Hantke, C.H. Yoon, T.A. White, and H. Chapman, "Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data," *J. Appl. Crystallogr. 47(3)*, 1118–1131 (2014).
49. B.J. Daurer, M.F. Hantke, C. Nettelblad, and F.R.N.C. Maia, "Hummingbird: monitoring and analyzing flash X-ray imaging experiments in real time," *J. Appl. Crystallogr. 49(3)*, 1042–1047 (2016).
50. V. Mariani, A. Morgan, C.H. Yoon, T.J. Lane, T.A. White, C. O'Grady, M. Kuhn, S. Aplin, J. Koglin, A. Barty, and H.N. Chapman, "OnDA: online data analysis and feedback for serial X-ray imaging," *J. Appl. Crystallogr. 49(3)*, 1073–1080 (2016).
51. D. Damiani, M. Dubrovin, I. Gaponenko, W. Kroeger, T. Lane, A. Mitra, C. O'Grady, A. Salnikov, A. Sanchez-Gonzalez, D. Schneider, and C.H. Yoon, "Linac Coherent Light Source Data Analysis Using Psana," *J. Appl. Crystallogr. 49(2)*, 672–679 (2016).
52. SLAC National Accelerator Laboratory, "OM: Online Data Analysis for X-Ray Imaging," https://www.ondamonitor.com./html/index.html (2020).
53. Bluesky, "bluesky-adaptive documentation," https://blueskyproject.io/bluesky-adaptive/ (2020).
54. I. Foster, "Globus Online: Accelerating and Democratizing Science through Cloud-Based Services," *IEEE Internet Computing 15(3)*, 70–73 (2011).
55. B. Allen, J. Bresnahan, L. Childers, I. Foster, G. Kandaswamy, R. Kettimuthu, J. Kordas, M. Link, S. Martin, K. Pickett, and S. Tuecke, "Software as a Service for Data Scientists," *Commun. ACM 55(2)*, 81–88 (2012).
56. J. Deslippe, A. Essiari, S.J. Patton, T. Samak, C.E. Tull, A. Hexemer, D. Kumar, D. Parkinson, and P. Stewart, "Workflow Management for Real-time Analysis of Lightsource Experiments," in *2014 9th Workshop on Workflows in Support of Large-Scale Science* (IEEE, New Orleans, LA, November 16, 2014), pp. 31–40.
57. *Basic Energy Sciences Exascale Requirements Review: An Office of Science review sponsored jointly by Advanced Scientific Computing Research and Basic Energy Sciences* (U.S. Department of Energy, Rockville, MD, November 3–5, 2015).
58. GitHub, "European XFEL," https://github.com/European-XFEL (2020).
59. European XFEL, "Data analysis at European XFEL," https://in.xfel.eu/readthedocs/docs/data-analysis-user-documentation/en/latest/index.html (2020).

60. Amazon Web Services, “Registry of Open Data on AWS,” https://registry.opendata.aws/ (2020).
61. PANGEO, “Technical Architecture,” https://pangeo.io/architecture.html (2020).
62. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K.A. Persson, “Commentary: The Materials Project: A materials genome approach to accelerating materials innovation,” *APL Mater. 1(1)*, 011002 (2013).
63. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, and G. Ceder, “Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis,” *Comput. Mater. Sci.* 68, 314–319 (2013).
64. “Materials Genome Initiative,” https://mgi.gov (2011).

APPENDIX A. WORKSHOP WEBSITE

The website for this workshop can be found at http://www.lanl.gov/conferences/compress.

Los Alamos National Laboratory

search site

Data Science & Computation for Compression Experiments

HOME | AGENDA | KEYNOTE | ABSTRACTS | PARTICIPANTS

REGISTER

Contact

Workshop Chair
Christine Sweeney
(505) 551-2367

About the workshop

Rapid and dynamic compression experiments are used to further our understanding of materials properties as well as Earth and planetary science. Rapid compression techniques such as dynamic diamond anvil cell (dDAC) provide a controlled load/pressure profile to materials. Dynamic or shock compression experiments such as laser or gas gun shock experiments at light sources provide information on materials under very high pressure. Both techniques have many commonalities in terms of experimental datatypes, analyses needed *during* an experiment, analyses *post*-experiment and goals of studying pressure-induced phase transitions.

This workshop will focus on computational needs of rapid and dynamic compression experiments at user light sources. It will also make connections to related experimental workflows and data from other experimental facilities such as neutron facilities and laser facilities. As facilities are upgraded, new detectors become available, and experimental compression rates increase, the need is greater for real-time analysis during experiments, as well as post-experiment data analysis that enables scientific discovery. The workshop will highlight all aspects of the experimental workflows from theory and simulation through experimental design and execution, detector development and reading, high performance computing, data reduction, statistical inference, statistical emulation, and machine learning.

As a tangible outcome, the workshop will identify current and future computational requirements for rapid and dynamic compression, the set of data science and supporting components that are needed to fill those requirements, and experiment workflow patterns that will utilize the components. Based on these findings, the workshop will identify future research directions to support these computational requirements and summarize them in a report.

APPENDIX B. LIST OF PARTICIPANTS

First Name	Last Name	Affiliation
Sohan	Ahmed	Washington State University
Minta	Akin	Lawrence Livermore National Laboratory
Michael	Armstrong	Lawrence Livermore National Laboratory
Jason	Baker	Los Alamos National Laboratory
Deborah	Bard	NERSC/Lawrence Berkeley National Laboratory
Anton	Barty	Deutsches Elektronen-Synchrotron
Cindy	Bolme	Los Alamos National Laboratory
Arun	Bommannavar	HPCAT-Argonne National Laboratory
Thomas	Caswell	Brookhaven National Laboratory
Chris	Coffelt	Georgia Institute of Technology
John	Copley	Princeton University
Samantha	Couper	University of Utah
Adrien	Descamps	SLAC National Accelerator Laboratory
Daniel	Dolan	Sandia National Laboratories
Thomas	Duffy	Princeton University
Sakun	Duwal	Sandia National Laboratories
Lars	Ehm	Stony Brook University
Devin	Francom	Los Alamos National Laboratory
Mungo	Frost	SLAC National Accelerator Laboratory
Eric	Galtier	SLAC National Accelerator Laboratory
Sirus	Han	Princeton University
Nicholas	Hartley	SLAC
Matthew	Horton	Lawrence Berkeley National Laboratory
Alex	Howard	Washington State University
Jiangeng	Huang	University of California, Santa Cruz
Rachel	Huber	Los Alamos National Laboratory
Simon	Hunt	University of Manchester
Rachel	Husband	Deutsches Elektronen-Synchrotron
Larissa	Huston	Los Alamos National Laboratory
Zsolt	Jenei	Lawrence Livermore National Laboratory
Jianjun	Jiang	Princeton University
Kevin	Jorissen	AWS
Ebad	Kamil	European XFEL, Germany
Chantelle	Kiessner	University of Utah
DONGHOON	KIM	Princeton University
Pawel	Kozlowski	Los Alamos National Laboratory

Harinarayan	Krishnan	Lawrence Berkeley National Laboratory
Roopali	Kukreja	University of California at Davis
Anmol	Lamichhane	University of Illinois at Chicago
Earl	Lawrence	Los Alamos National Laboratory
Hanns-Peter	Liermann	Deutsches Elektronen-Synchrotron
Rachel	Lim	Carnegie Mellon University
Mingda	Lv	Michigan State University
Valerio	Mariani	SLAC National Accelerator Laboratory
Hauke	Marquardt	University of Oxford
Chris	McGuire	University of Edinburgh
Thomas	Michelat	European XFEL
Lowell	Miyagi	University of Utah
David	Montgomery	Los Alamos National Laboratory
Eric	Moss	Los Alamos National Laboratory
Bob	Nagler	SLAC National Accelerator Laboratory
Earl	O'Bannon	Lawrence Livermore National Laboratory
Ian	Ocampo	Princeton University
Claudia Camila	Parisuana Barranca	Stanford University
Nilesh	Salke	University of Illinois at Chicago
Daniel	Savage	Los Alamos National Laboratory
Hannah	Shelton	Lawrence Livermore National Laboratory
Saransh	Singh	Lawrence Livermore National Laboratory
Jesse	Smith	HPCAT—Argonne National Laboratory
Blake	Sturtevant	Los Alamos National Laboratory
Christine	Sweeney	Los Alamos National Laboratory
Toni	Taylor	Los Alamos National Laboratory
Naresh	Thadhani	Georgia Institute of Technology
Jana	Thayer	SLAC National Accelerator Laboratory
Sally	Tracy	Carnegie Institution for Science
Terry	Turton	Los Alamos National Laboratory
Nenad	Velisavljevic	Physics Division, Lawrence Livermore National Laboratory and High Pressure Collaborative Access Team, Advanced Photon Source
Sven C.	Vogel	Los Alamos National Laboratory
Robert	Von Dreele	Advanced Photon Source, Argonne National Laboratory
Zhehui (Jeph)	Wang	Los Alamos National Laboratory
Matt	Whitaker	Stony Brook University / NSLS-II
Wei	Xu	Brookhaven National Laboratory
Sohan	Ahmed	Washington State University

APPENDIX C. WORKSHOP AGENDA

LOS ALAMOS NATIONAL LABORATORY **Agenda**

Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities

September 8-11, 2020

Tuesday, September 8, 2020 (Times in MDT)

9:00 am – 9:05 am	Welcome, Committee Introductions	Christine Sweeney (LANL)
9:05 am – 9:15 am	Formal Welcome, Opening Conference Remarks	Antoinette (Toni) Taylor (LANL), Associate Laboratory Director fo Physical Sciences
9:15 – 10:15 am	Keynote Address: Data Science and Application in High Pressure Research	Nenad Velisavljevic, Staff Scientist, LLNL, HPCAT Director, APS
10:15 – 10:30	Charge to Participants. Workshop Approach	Christine Sweeney (LANL)
10:30 -- 10:45	Break	
10:45 -- 11:15	Plenary: Building a Community Platform for Data Science with the Materials Project	Matthew Horton (LBL)
11:15 - 11:45	Plenary: Towards a BES Light Source Wide Sustainable Software Stack	Hari Krishnan (LBL)
11:45 – 12:15	Lunch break	
12:15 – 1:30	Panel: Commonalities of Data and Analytics for Static and Dynamic Compression	Host: Lowell Miyagi (U. Utah) Earl O'Bannon (LLNL) Eric Galtier (SLAC) Rachel Husband (DESY) Sally Tracy(Carnegie Inst)
1:30 -- 1:45	Break	
1:45 -- 2:45	Session (Speakers/Discussion): XRD and Scattering: Data Processing Tools	Host: Rachel Huber (LANL) Trevor Willey (LLNL) Erik Watkins (LANL) Richard Briggs (LLNL)
2:45 - 3:00	Preview of Day 2	Christine Sweeney (LANL)

Institutional Host: Richard Sheffield Agenda Revised: 9/3/20
Technical Host: Christine Sweeney, CCS-7, 505-551-2367

LOS ALAMOS NATIONAL LABORATORY Agenda

Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities
September 8-11, 2020

Wednesday, September 9, 2020 (Times in MDT)

9:00 am – 9:10am	Welcome to Day 2, Announcements, Charge for Day 2	Christine Sweeney (LANL)
9:10 am - 9:15am	Day 1 Session Report back	Session Leads
9:15 – 10:15am	Session (Speakers/Discussion): Time-dependent Measurements Using Static Apparatus	Host: Jesse Smith (ANL) Mungo Frost (SLAC) Simon Hunt (U. Manchester) Rachel Husband (DESY)
10:15-10:30	Break	
10:30 - 11:30	Session (Speakers/Discussion): Image Processing	Hosts: Brian Jensen (LANL) and Emma McBride (SLAC) David Montgomery (LANL) Bob Nagler (SLAC)
11:30 - 12:00	Lightning Talks (students, post docs)	Blake Sturtevant (LANL)
12:00 – 12:30	Lunch	
12:30 - 1:30	Session (Speakers/Discussion): Machine Learning for Compression Workflows	Host: Mathew Cherukara (ANL) Devin Francom (LANL) Sudip Seal (ORNL) Robert Von Dreele (ANL)
1:30 - 1:40	Break	
1:40 - 2:40	Session (Speakers/Discussion): Visualization	Host: Terry Turton (LANL) Dan Orban (U. Minnesota) Wei Xu (BNL)
2:40 - 2:55	Day 2 Session Report Backs	Session Leads
2:55 - 3:00	Preview of Day 3	Christine Sweeney (LANL)

LOS ALAMOS NATIONAL LABORATORY **Agenda**

Data Science and Computation for Rapid and Dynamic Compression Experiment Workflows at Experimental Facilities
September 8-11, 2020

Thursday, September 10, 2020 (Times in MDT)

9:00 am – 9:10 am	Welcome to Day 3, Report-backs	Christine Sweeney, Session Hosts
9:10 - 10:10	Session (Speakers/Discussion): Real-time Analytics Workflows and Computation	Host: Chris Biwer (LANL) Valerio Mariani (SLAC) Thomas Caswell (BNL)
10:10 - 10:25	Break	
10:25-11:40	Session (Speakers/Discussion): Post Hoc Computational Tools	Host: Clemens Prescher (DESY) Lars Ehm (Stony Brook) Hauke Marquardt (U. of Oxford) Michael Armstrong (LLNL) Sven Vogel (LANL)
11:40 - 12:20	Lunch	
12:20 - 12:50	Plenary: 6H detector Frontiers and Recent Progress	Jeph Wang (LANL)
12:50 - 2:05	Panel and discussion: Facility Computational Resources in Support of Rapid and Dynamic Compression Experimental Data Analysis Workflows	Host: Christine Sweeney (LANL) Debbie Bard (LBL) Jana Thayer (SLAC) Thomas Michelat (EuXFEL) Kevin Jorissen (Amazon)
2:05 - 2:20	Break	
2:20-2:30	Report backs	
2:30 - 3:20	Discussion of gaps, priorities and path forward	
3:20 - 3:30	Workshop Concluding Remarks	Christine Sweeney (LANL)

Los Alamos National Laboratory Agenda

Friday, September 11, 2020 (Times in MDT)

9:00 am – 12 noon	Report Writing	Steering Committee, and optional for everyone else

APPENDIX D. KEYNOTE, PLENARY, AND LIGHTNING TALK ABSTRACTS AND PANELS

Keynote: Data Science and Application in High Pressure Research

Nenad Velisavljevic

Staff Scientist, Physics Division, Lawrence Livermore National Laboratory (LLNL)

Director for High Pressure Collaborative Access Team (HPCAT) at the Advanced Photon Source (APS), Argonne National Laboratory

Data science has far reaching implications. As with any major innovation or technological advancement there are significant pros and cons. In this talk we'll take a broad (and hopefully fun) perspective look at data science and how this field is naturally merging with the evolution of high-pressure research field. I will provide quick history and current status of high-pressure research and conclude with the look at the implications of data science use and more systematic implementation for future research.

Plenary: Building a community platform for data science with the Materials Project

Matthew Horton, Lawrence Berkeley National Laboratory

The Materials Project is a public-facing resource of computed and experimental data for over a hundred thousand crystalline materials. This talk will discuss the lessons learned from building and curating a data platform like the Materials Project, how to make this data most useful to a community of users from government, academia, industry and education, and how the data can be used for applications including materials discovery, experimental characterization and machine learning.

Plenary: Towards a BES Light Source Wide Sustainable Software Stack

Harinarayan Krishnan, Lawrence Berkeley National Laboratory

This talk will highlight development efforts as part of the work that is being undertaken for DOE light sources by the BES Data Pilot project and the Center for Advanced Mathematics for Energy Research Applications (CAMERA). Additionally, the talk will delve into the intricacies of deploying a real time event-driven nano-tomography processing pipeline with the aim to better understand the data pipeline and processing requirements needed for production environments. Tomography is a workhorse technique used at many scientific user facilities worldwide, and makes an ideal candidate in capturing the complexities of executing a multi-system distributed workflow in realtime.

Panel: Commonalities of Data and Analytics for Static and Dynamic Compression

Panel Lead: Lowell Miyagi, University of Utah

This panel discussion seeks to identify data collection practices, data types and analytic tools that are utilized for a range of static and dynamic compression experiments ranging from traditional quasi static diamond anvil cell, to piezo driven dynamic diamond anvil cell to shock compression experiments. The goal of this panel will be to identify current commonalities between these techniques as well as identify current and future challenges to enable more efficient and targeted data collection and analysis.

Session: XRD and Scattering: Data Processing Tools

Session Lead: Rachel Huber, Los Alamos National Laboratory

X-ray diffraction and scattering at both synchrotrons and X-ray free electron laser facilities produce large amounts of data that prove a tremendous task to reduce and analyze after the beamtime. However, even more importantly in some cases, on-the-fly analysis at the beamline would be highly beneficial to determine if the current beamtime plan is the most prudent course of action. This session will explore static and dynamic experiments and how these experiments may benefit from improved data analysis methods, including the benefits of on-the-fly analysis at the beamline.

Session Talk: Scientists under rapid compression: data processing of dynamic compression experiments at LCLS

Richard Briggs, Lawrence Livermore National Laboratory

Vast improvements have been made to online (and offline) data analysis tools employed during dynamic compression experiments at the MEC end-station of LCLS. In this talk, I'll present some of my thoughts on the history, progress, and lessons learned from the tools used to process XRD data collected during shock compression experiments at MEC since 2013.

Session Talk: SAXS in DACs and other extremes

Erik Watkins, Los Alamos National Laboratory

Small angle x-ray scattering (SAXS) probes ordering in materials over length scales ranging from around a nanometer to several hundred nanometers. Here, I will discuss recent applications of SAXS to study silica nanospheres and amorphous carbon at high pressures (up to 10s of GPa) in diamond anvil cells (DACs) as well as in-situ SAXS studies of demixing in high-Z alloys.

Session Talk: Data Processing Tools for Detonation Science at the Advanced Photon Source

Trevor Willey, Lawrence Livermore National Laboratory

We have been performing small-scale detonation experiments at the Advanced Photon Source for several years, and outline successes and challenges in processing scattering and diffraction data from these dynamic events.

Session: Time-dependent measurements using static apparatus

Session Lead: Jesse Smith, HPCAT, Argonne National Laboratory

Complementary advances in light sources and high pressure apparatus make it possible to carry out time-dependent extreme-conditions measurements using static high pressure apparatus. The types and volume of data generated by these novel techniques can easily overwhelm the tools traditionally used for analyzing static high pressure data. While there have been a few notable software tools developed in recent years, there remains an urgent need for the continued development of image and pattern analysis software to facilitate the reduction and analysis of data collected during rapid/ramp compression and time-dependent imaging. In this session we explore a few relevant experimental techniques and hope to articulate some of the software tools which could benefit future work.

Session Talk: Outrunning Reactions with Rapid Compression

Mungo Frost, SLAC National Accelerator Laboratory

Lithium is among a number of substances which undergoes reactions in diamond anvil cells making its study challenging. By compressing rapidly this can be overcome allowing structural data to be collected at high P-T conditions prior to anvil failure. The methods used to analyze the very weak scattering data will be discussed, as well as future challenges for this kind of experiment.

Session Talk: Improved strain resolution in time and exposure constrained diffraction patterns using Fourier series

Simon Hunt, Department of Materials, University of Manchester, Manchester, UK

The determination of differential strain from noisy 2D X-ray diffraction is limited by data processing and reduction techniques (i.e. diffraction pattern caking and azimuthally independent peak fitting). Here we present a new method that directly fits 2D diffraction data by using Fourier series to describe the azimuthal properties of diffraction peaks. Testing on a sample data set shows significantly more precise differential strain values that enables the extraction of previously unresolvable time dependent properties.

Session Talk: Time-resolved X-ray diffraction and Imaging of dynamically compressed samples in the diamond anvil cell

Rachel Husband, Deutsches Elektronen-Synchrotron

The fast diffraction set-up at the Extreme Conditions Beamline (ECB) at PETRA III offers the possibility of performing simultaneous time-resolved X-ray diffraction and imaging of samples as they are dynamically compressed using a piezo-driven diamond anvil cell. Recent improvements of detector technology mean that it is now possible to continuously collect diffraction data at kHz rates, and a single compression cycle can generate between ~100-10000 images. The large volume of data generated in these experiments has the potential to revolutionize the landscape of typical synchrotron experiments, and the development of suitable batch processing tools is required to maximize the information that can be extracted from such large datasets.

Session: Image Processing

Session Leads: Emma McBride, SLAC National Accelerator Laboratory and Brian Jensen, Los Alamos National Laboratory

Session Talk: Quantitative Phase Retrieval for Dynamic Compression Experiments

David Montgomery, Los Alamos National Laboratory, P-24

I will discuss the issues surrounding quantitative phase retrieval at synchrotrons in regimes where diffraction must be included, and in the presence of noise and blur. Single step methods do not generally apply to these problems, and the traditional Gerchberg-Saxton/Fienup methods don't do well in the presence of noise and cannot include real blur effects. Phase retrieval for these problems is ill-posed and often nonlinear. I will discuss progress in this area and future needs.

Session Talk: Phase Contrast and Direct Imaging Experiments of Shock Compressed Matter at MEC

Bob Nagler, SLAC National Accelerator Laboratory

I will present an overview of imaging experiments that have been performed at at the Matter in Extreme Conditions instrument at LCLS. Past scientific results will be shown, and we will show why there is a need for (improved) phase retrieval algorithms.

Session: Machine Learning

Session Lead: Mathew Cherukara, Argonne National Laboratory

Session Talk: The pink beam powder diffraction profile

Robert Von Dreele, Advanced Photon Source, Argonne National Laboratory

The powder diffraction profile obtained with a pink beam synchrotron x-ray source is described as the convolution of a back-to-back pair of exponentials convoluted with the Gaussian and Lorentzian components of a pseudo-Voigt. This new function is employed for the first time in a Rietveld refinement using data collected from a single 100ps synchrotron X-ray micropulse.

Session Talk: Simulation and Emulation of X-ray Diffraction from Dynamic Compression Experiments

Devin Francom, Los Alamos National Laboratory

Description of simulation and emulation techniques for X-ray diffraction data, especially for dynamic compression experiments. The focus will be on the emulation (machine learning) method.

Session: Visualization

Session Lead: Terry Turton, Los Alamos National Laboratory

Visualization and visual analytics are a key component for experimental workflows to support both real-time data analysis that can immediately impact run-time decisions and post hoc analysis. In this session, we will explore current visualization approaches and discuss user needs to drive future research directions.

Session Talk: Visually Exploring Parameter Spaces through Intuitive User Interaction

Daniel Orban, University of Minnesota

This talk presents a visualization technique that allows experts to quickly and interactively annotate parameterized simulation instances within input and output contexts. Users can then ask dynamic "what if?" questions by directly manipulating the input variables or output curves. Our approach enables an intuitive user-defined understanding of a continuous parameter space.

Session Talk: Visual Analytics for Multi-Attribute Scientific Data Analysis and Model Explanation

Wei Xu, Brookhaven National Laboratory

In this talk, I will introduce three ongoing visual analytics projects for NSLSII at Brookhaven National Laboratory. These projects aim at providing interactive data analysis and machine learning model understanding for scientific experiments and workflows.

Session: Real-time Analytics

Session Lead: Christopher Biwer, Los Alamos National Laboratory

Performing real-time data analysis requires automation to manage data, execute analysis tasks, and visualize results. In this session, the speakers will present examples of tools for building real-time analyses. We will discuss how these tools can enable feedback used to steer experiments and meet the requirements analyses have to be performed in real-time.

Session Talk: Adaptive BlueSky

Thomas Caswell, Brookhaven National Laboratory

Prompt / near real-time analysis is essential to make the best use of the available beamtime. Traditionally this is done by having a human-in-the-loop evaluating preliminary results and drive subsequent experiments. Using Bluesky, a data acquisition and management suite developed at NSLS-II and in use at all 5 light sources, we demonstrate the use AI/ML methods to steer more efficient data acquisition at NSLS-II.

Session Talk: OM: quasi-real time feedback for serial X-ray imaging experiments

Valerio Mariani, SLAC National Accelerator Laboratory

We introduce the OM software package. OM can monitor experimental conditions in quasi real-time, providing researches with essential information required for quick decision making. Quick feedback allows experiments to be stopped or steered in a different direction when unfavorable conditions are met, preventing the collection of unfavorable data and preserving valuable resources.

Session: Post hoc Data Analytics

Session Lead: Clemens Prescher, Deutsches Elektronen-Synchrotron, Hamburg

In this session examples of detailed analysis of large sets of X-ray diffraction data collected during dynamic compression experiments will be presented, specifically focussing on publication ready analysis after the experiment (post hoc).

Session Talk: From 2D-Diffraction images to quantitative structural data

Lars Ehm, Stony Brook University

Time-resolved diffraction measurements of material undergoing rapid compression create large amounts of data that allow an in-depth view into the complex process in the material down to the atomic scale. We will present some recent successes and failures in the extraction of quantities structural data from the large set of 2D diffraction data.

Session Talk: Processing of dDAC XRD data

Hauke Marquardt, University of Oxford

The combination of the dynamic diamond-anvil cell (dDAC) with time-resolved XRD imaging allows for tracking the response of materials to changes in pressure with a practically continuous pressure resolution. The resulting data sets often contain several hundreds of diffraction images, requiring an automated fitting routine to effectively analyze the experimental results. Here, I will summarize our experience dealing with such data sets and further discuss the novel information that can be extracted.

Session Talk: Analysis of diffraction data from ultrafast compression of Zr

Michael R. Armstrong, Lawrence Livermore National Laboratory

I will talk about using hydrodynamics simulations to simulate diffraction patterns from samples under nonuniform laser-driven dynamic compression. Under nonuniform compression, diffraction lines will be inhomogeneously broadened, but simulations can obtain quantitatively similar patterns with some simple assumptions.

Session Talk: Microstructure Evolution during dynamic compression in titanium characterized with the XFEL at LCLS-2

Sven C. Vogel, Los Alamos National Laboratory

The transient time of a shock wave traveling at a few 1000 m/s through ten micrometers of materials is a few ns. Post mortem analysis provides only limited insight into the microstructural evolution during such a shock. Advent of XFELs and high-powered optical lasers to induce shocks into materials have enabled in situ characterization of the phase transitions, strains, and texture evolution occurring during dynamic compression materials. We report the first successful application of Rietveld analysis to obtain strains, volume fractions, and textures of the phases occurring during shock of pure titanium metal, thus providing for the first time a fairly complete picture of the microstructure evolution during shock.

Detector Research Plenary: 6H detector frontiers and recent progress

Zhehui (Jeph) Wang, Los Alamos National Laboratory

We summarize the `6H' detector frontiers for high-speed X-ray imaging, utilizing the state-of-the-art high-energy X-ray sources especially the Argonne Advanced Photon Source. The interdisciplinary approach that integrates experiments, instrument development, and data science have delivered encouraging results.

Panel: Facility Computational Resources in Support of Rapid and Dynamic Compression Experimental Data Analysis Workflows

Panel Lead: Christine Sweeney, Los Alamos National Laboratory

In this panel, we will hear perspectives of experts from both experimental facilities and computing facilities. We will discuss the current and future capabiliites of these facilities and how they are adapting to meet the analysis needs of rapid and dynamic compression scientists and the challenges of data management.

Lightning Talks (Wednesday September 9, 2020)

Chris McGuire (Univeristy of Edinburgh), Thermodynamic assessment of the Fe-C binary phase diagram at extreme conditions

Andres Quan (Los Alamos National Laboratory), Visualizing XRD Data with Cinema:Snap

Seunghee Oh (Carnegie Mellon University), In situ Characterization of Phase Evolution of Ni Alloy 718 During Laser Processing with High Energy Synchrotron X-ray Diffraction

Samantha Couper (University of Utah), Double sided laser heating in the radial DAC and a cubic to monoclinic phase transition in FeO

APPENDIX E. WORKSHOP QUESTIONNAIRE

Participants filled out a registration questionnaire on their background and interests; we present a summary of the results here.

The majority of attendees were in one of four categories: mid/late career research staff (32.6%), early career (21.1%), graduate student (20%), or postdoc (13.7%), with the remaining attendees being support/technical staff, industry experts, university professors, managers, and early/mid-career research staff.

Research interests were predominately in dynamic compression (56.8%), static compression (48.4%), imaging (36.8%), computer science (29.5%), statistics (14.7%), and one each in a number of other areas: rotational shear DAC, diffraction/crystallography/microstructure, materials science and engineering, micromechanical modelling, detonation physics, computational and analytic support for light sources, scientific software, crystallography, synchrotrons, X-ray scattering, XFELs, and in-situ X-ray diffraction.

Facilities currently being used or are anticipated to be used included HPCAT (46.3%), LCLS (46.3%), DCS (35.8%), EuXFEL (34.7%), PETRA III (26.3%), and NIF/OMEGA (21.1%). Three participants use GSECARS, two use ALS, and one participant each uses the following facilities: LANSCE, APS 1-1D, NSLS-II, NERSC, APS other beamlines, BL10-XU (SPring-8), OLCF/ALCF, NSLS-II, FLASH, TPW, JLF, APS 34ID-C Compression/Tensile, SSRL, CHESS, ANL, and LANL. Seven participants indicated using no facilities.

Experimental datatypes currently used were primarily XRD images (82.1%), VISAR (34.7%), oscilloscope traces (24.2%), pyrometry (11.6%), and one or two each from XRD spectral data and hyperspectral data, neutron time-of-flight, image sequences and diffraction sequences, X-ray phase contrast images, X-ray imaging data, X-radiographs, WAXS/SAXS, XAS/XANES/NEXAFS, crystallography, data from NSLS-II, direct imaging, neutron scattering images, satellite images, XRTS, radiography, THz, SAXS, NR/XRR, MPDV, Raman, EXAFS, inelastic X-ray scattering, spectroscopy, and VIS images.

Nearly 12 percent of participants felt that all their data and data analysis needs were met; however, 83 percent felt that only some of their needs were met and 5.9 percent felt that their needs were not met at all. Furthermore, 40 percent anticipated changes in data requirements and analyses in the near future. An additional 38.8 percent anticipated that one of the existing techniques would be upgraded and would change their data needs. Finally, 21.2 percent did not predict any changes to data or analytics in the near future.

Data analytics used by experimentalists were predominantly curve fitting (71.4%), Rietveld refinement (51.2%), uncertainty analysis (51.2%), real-time data visualization (42.9%), experimental design (40.5%), principal component analysis (13.1%), neural networks (11.9%), and Gaussian processes (8.3%). Other techniques mentioned were OpenCV, SAXS analyses, independent codes, X-ray CT, conventional analysis image processing (wavelets), iterative phase retrieval (ptychography), diffraction imagine/phase contrast, single crystal refinement, and DIC.

Participants indicated that they would like to start using data analysis techniques of real-time visualization (49.3%), modeling (e.g., finite element models) (42.3%), principal component analysis (39.4%), neural networks (32.4%), uncertainty analysis (29.6%), experimental design (29.6%), Gaussian processes (15.5%), and curve fitting (14.1%). Several people indicated machine learning and SAXS imaging, and some were not sure.

Experimentalists currently use the following computational resources: laptops (89.8%), workstations (59.1%), high performance computing cluster (40.9%), GPUs (19.3%), and the public cloud (1.1%).

In terms of expected outcomes, participants sought a greater understanding of the datatypes and requirements in rapid and/or dynamic compression (72.6%), an improvement in understanding of the science and experimental aspect of rapid and/or dynamic compression (66.3%), building a network and collaborations (58.9%), gaining a better understanding of computer science and how it can be used in data analysis (57.9%), learning about how forthcoming technology will benefit rapid and/or dynamic compression and the computational challenges associated with these technologies (56.8%), identification of reusable analytics, computational techniques and/or software components that can be shared by others doing similar experiments (55.8%), and providing input on important research directions in data analytics and computation that will further science in the rapid and/or dynamic compression areas (36.8%).

Los Alamos
NATIONAL LABORATORY