Image Analysis and Feature Classification in Atomically Resolved Images: Genomic Libraries through Smart Data

Rick Archibald[b1], Alex Belianinov[b4], Albina Borisevich[b3], Miaofang Chi[b4], Eirik Endeve[b1], John Freeland[a5], Doug Fuller[b1,3], Anubhav Jain[c2], Stephen Jesse[b4], Sergei Kalinin[b4], Guruprasad Kora[b3], Nouamane Laanait[b4], Kristin Persson[c2], Amanda Petford-Long[a5], Andrew Lupini[b4], Oleg Ovchinnikov[b4], Galen Shipman[b1,3], David Skinner[c2], Chris Symons[b1], Ray Unocic[b1].

Laboratories

Argonne National Laboratory [a]

Oak Ridge National Laboratory [b]

Berkeley Lab — Lawrence Berkeley National Laboratory [c]

DOE Facilities

OLCF Oak Ridge Leadership Computing Facility [1]

NERSC [2]

CADES Compute & Data Environment for Science [3]

Computational

Oak Ridge National Laboratory — Center for Nanophase Materials Sciences [4]

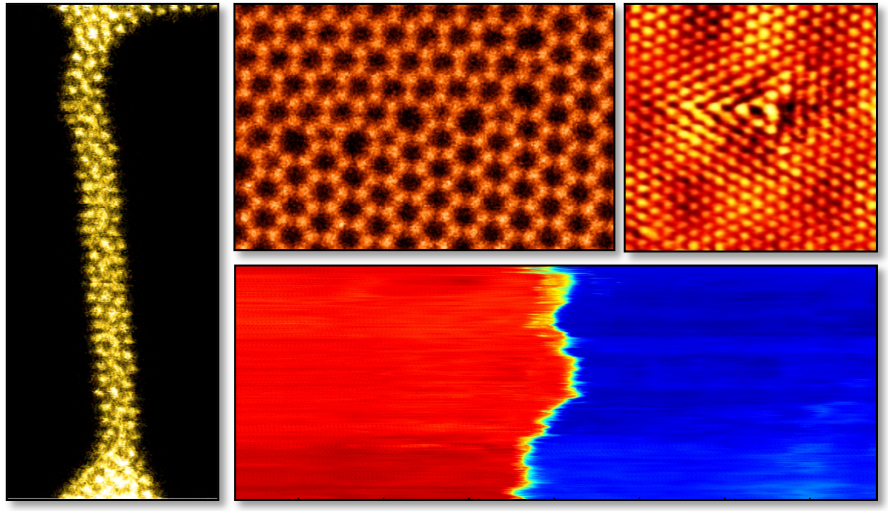
Argonne National Laboratory — Center for Nanoscale Materials [5]

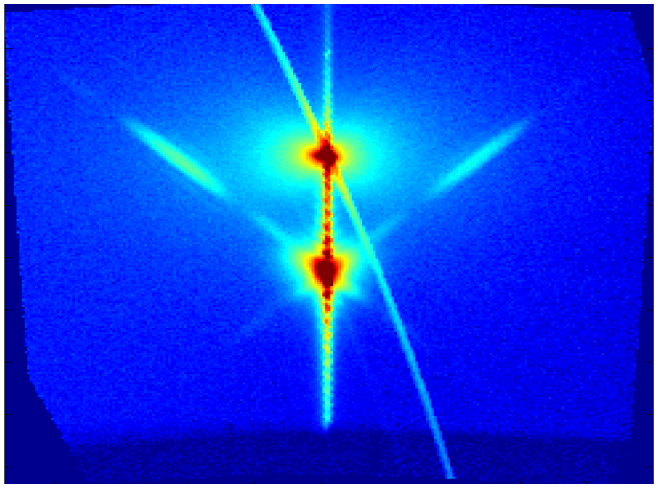
Experimental

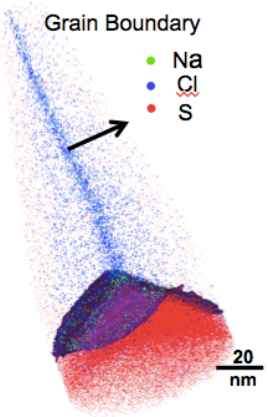
Outstanding imaging capabilities

Providing excellent capabilities for imaging and spectroscopy of a wide range of nanomaterials

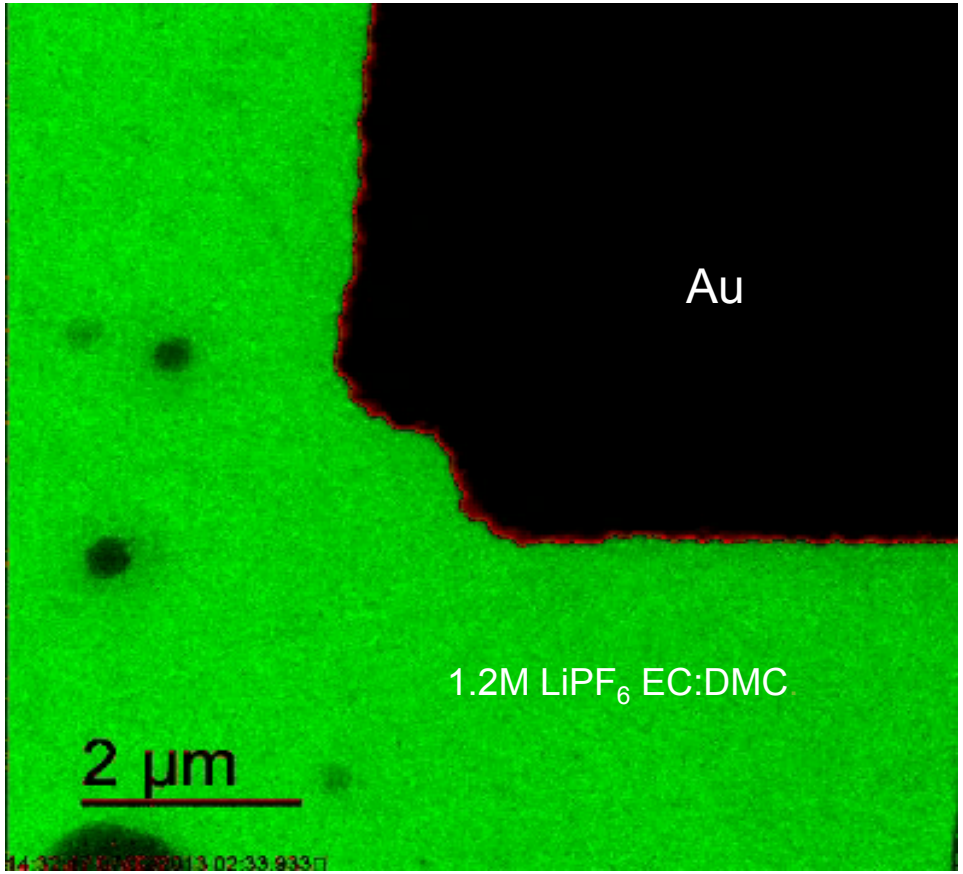
- Atomically resolved TEM, STEM imaging and EELS to correlate atomic structure, bonding, and excitations with material properties (**2MB – 100GB**)
- Advanced modalities of scanning probes to understand and electronic, ionic, and molecular behaviors at the nanoscale (**32MB – 128GB**)
- Innovative capabilities to dynamically study the behavior of materials in operando with X-rays (e.g., electrochemistry, chemical transformations, nucleation, assembly) **(Data rate 0.1-50 Mb/sec) (Dataset 10-100 Gb/System)**
- Atom probe tomography to study 3D nature of individual defects at the single-atom level (1.5 million ions)



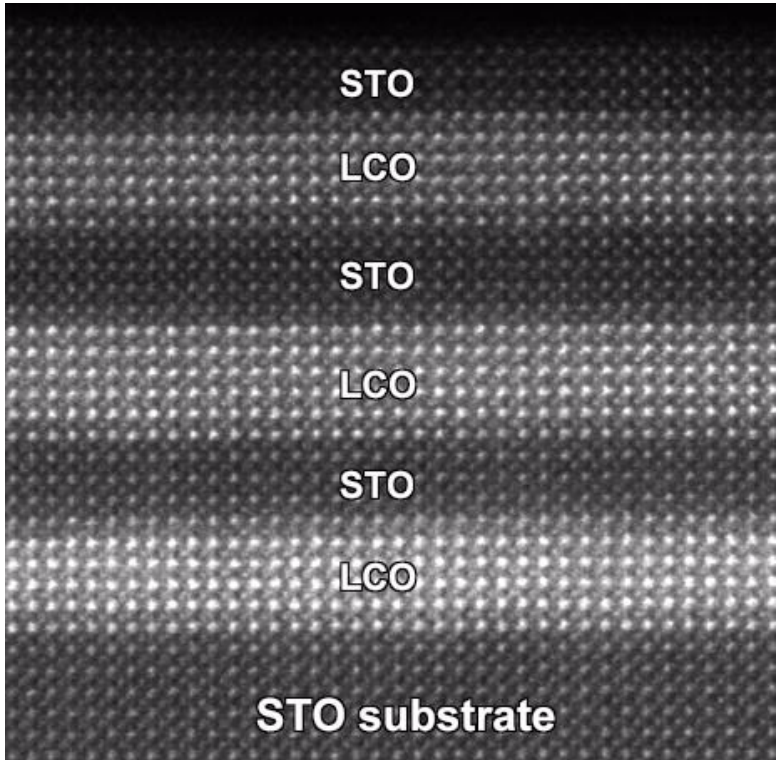




In-situ and dynamic S/TEM



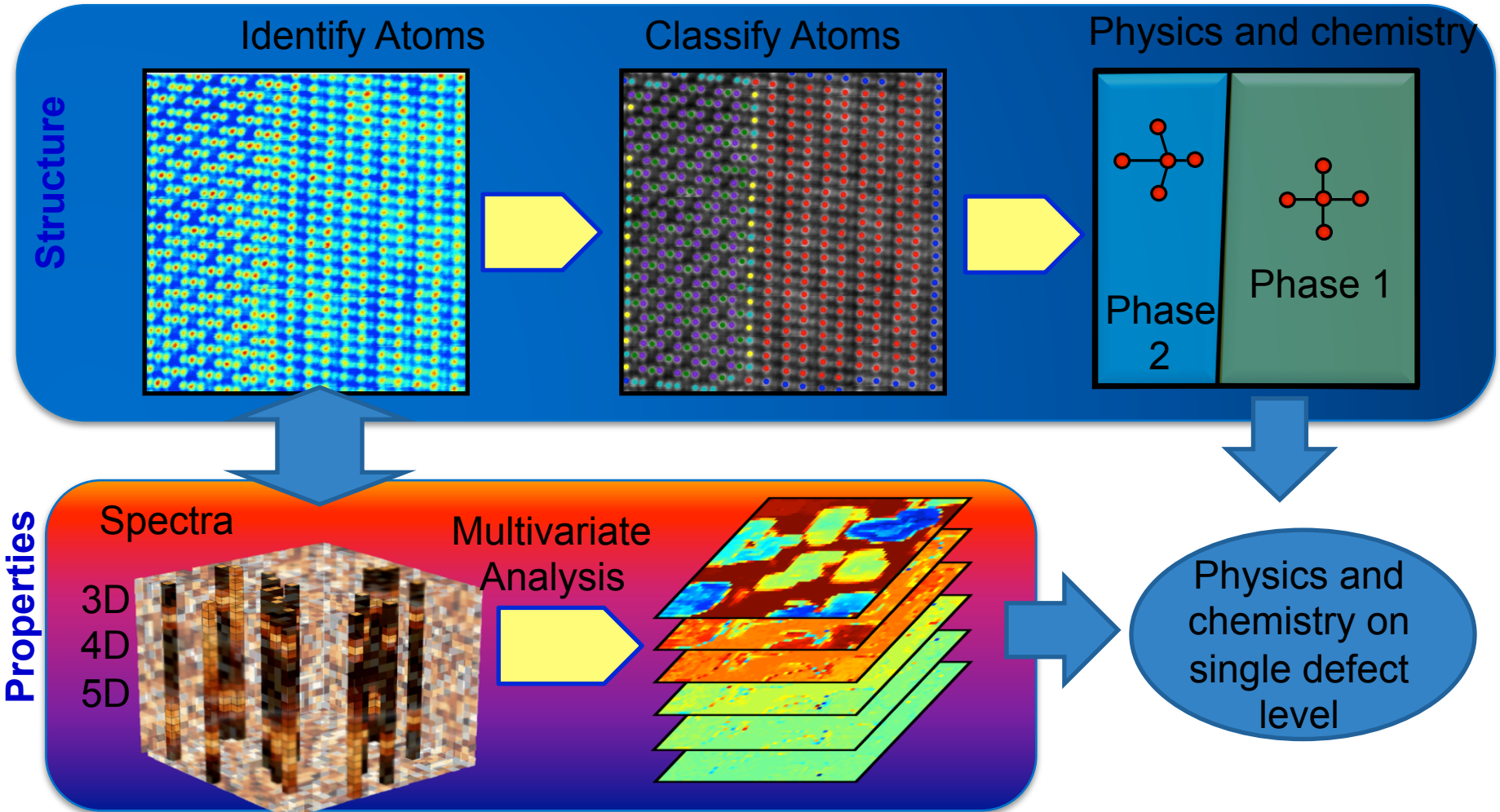
Lithium dendrite nucleation and growth



Vacancy ordering (11-18 nm confinement layer)

Capability: Environmental 1000 images/sec; 10 min pulse; 11-13 TB/day
Dynamic 1,000,000 images/sec; 10 sec pulse; **20 TB /burst**

Atomic-Scale Structure and Functionality



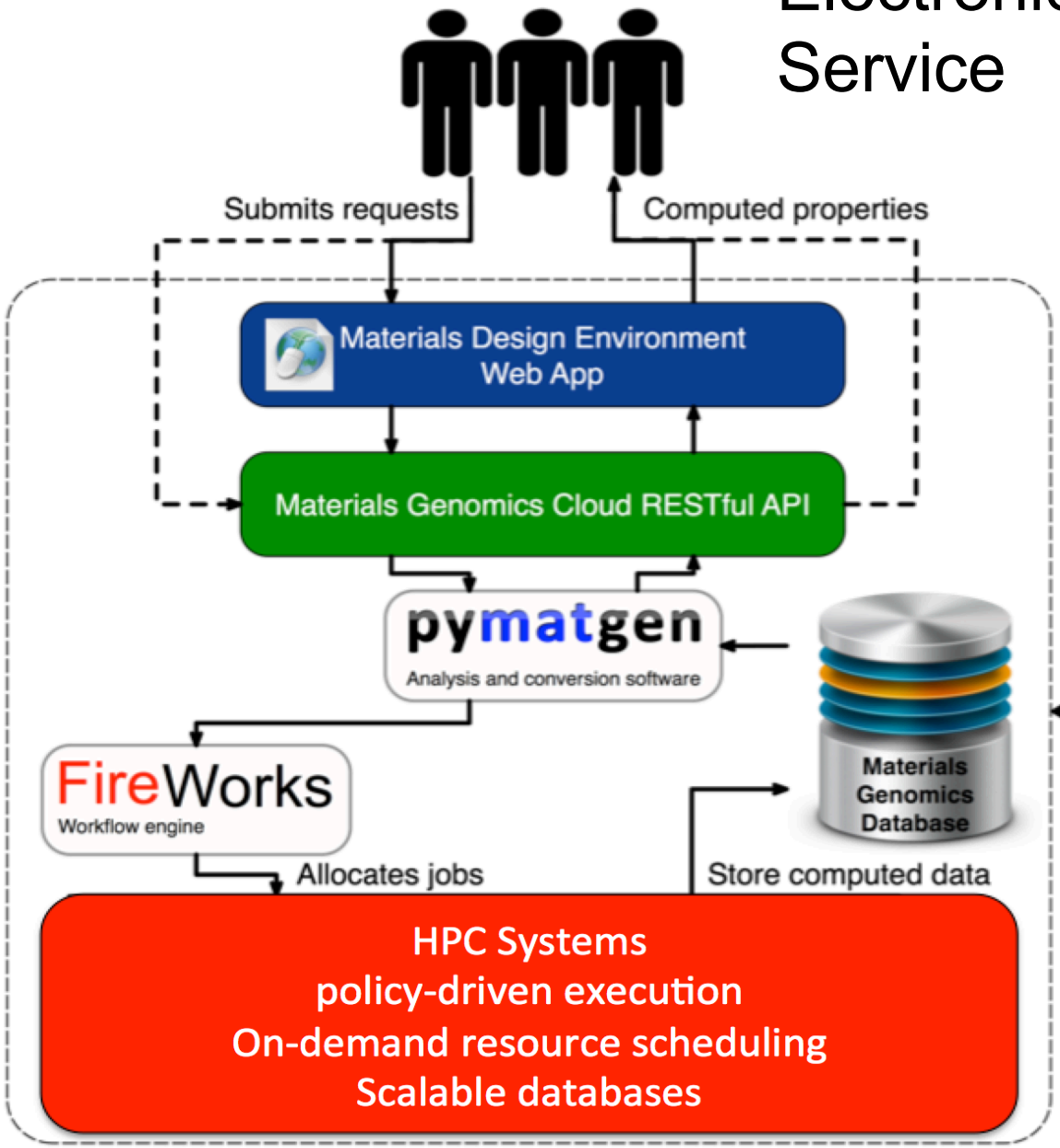
Need new language:

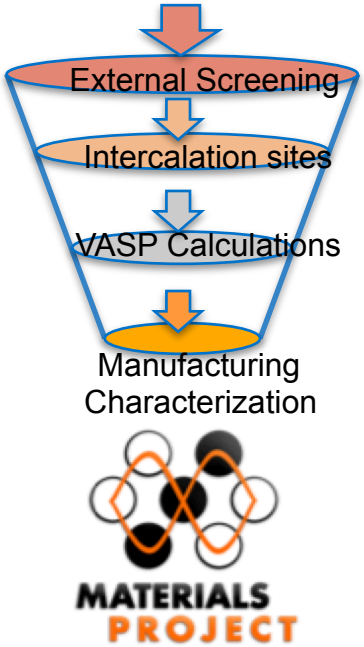
1. What are structural descriptors?
2. How do we define local symmetry, phases and ferroic variants?
3. How do we introduce and quantify translational symmetry?

What do we learn:

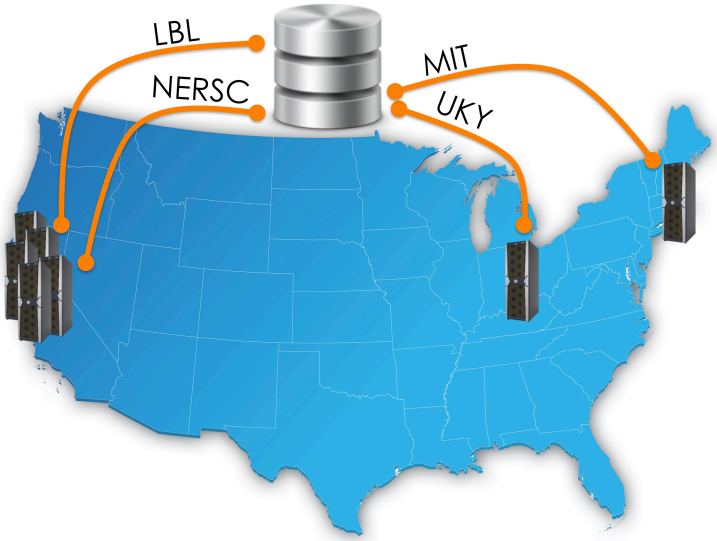
1. Structure-property relationship on single atom, molecule, and defect level
2. Feedback to theory through microscopic degrees of freedom
3. Input to Materials Genome

Electronic Structure as a Service





Advanced Materials Search	Computed material properties	Plot phase diagrams
Battery Explorer	Compute reaction energies	Predict chemical structures



An advanced workflow system leverages the computational power from several locations.

Scientific capabilities

Advanced Materials Search	Compute d material properties	Plot phase diagrams	Battery Explorer	Compute reaction energies	Predict chemical structures	Crystal Toolkit

Web interface and services

Django	Javascript

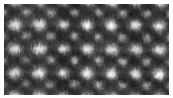
Core Python libraries

Numpy	Scipy	Pymatgen	pymongo

High Level Demonstration Workflow: Dynamic Matter (Perovskite Superlattice)

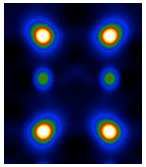
ORNL: STEM/EELS

Data de-noising and regularization



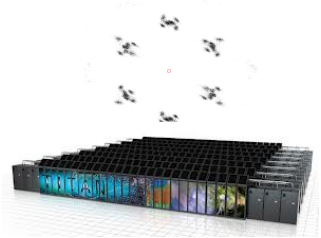
ANL: APS Focused X-Ray

Approximate structural models



Acumen

Image based Substrate and unit cell identification extraction of atomic structure and trajectories



High throughput calculation of electronic structure of these configurations

Knowledge base of experiments and simulations

Data sharing and catalog of results

High Level Demonstration Scientific Goals

- Structural workflow enables structural description of atomically resolved images, identifies phases and ferroic variants, and topological and structural defects.
- A set of datasets including high resolution STEM images of complex catalysts and ferroelectric perovskites are generated.
- Identification and refinement of atomic positions
- Exploration of local structure
- Extraction of structure features such as repeating structural elements, and store them as structure database for cross-image correlation.

Data Driven Scientific Discovery – Mathematical Developments

- Develop feature detection algorithms to identify structures of interest in atomically resolved data
- Develop image registration algorithms to correlate this data with local spectroscopies
- Build scientific knowledge into a database for computation discovery methods
  - Generate column shapes for given structural distortions.
  - General spectral signatures for given local neighborhoods
  - Cost effective approach to enhance experimental databases
- Computational methods can be generated to identify material properties and provide scientist analysis tools and access to massive database of experimental and simulation knowledge.

Supporting near real-time analysis – Computational Developments

ORNL: STEM/EELS
ANL: Focused X-ray

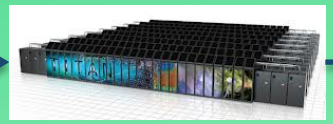
- High throughput image capture
- Multi-modal:
  -High-angle annular dark field (HAADF) detector
  - Electron energy loss spectroscopy
  - Focused X-ray
- O(1000) of images per experiment

Minutes

Transfer files to HPC storage and convert data to HDF5 format

- Data motion via BBCP/GridFTP
- Data conversion from DM3 stacks to HDF5 slices
- Standard data format (HDF5)
- Data layout conducive to HPC algorithms

Seconds



Massively parallel Image processing/ feature detection

- Denoise Image
- Identify atoms
- Identify lattice
- Thousands of images/sec
- Built on MPI +Fortran
- Near linear scaling – some limitations as file counts get extremely large

Seconds

Massively parallel electronic structure calculation

- Massively parallel study of hundreds to thousands of individual configurations
- Calculation of electronic structure
- Based on DFT
- 0(1000) of configurations

Minutes

Scalable Scientific Image Analysis – Demonstration Problem

Problem description:

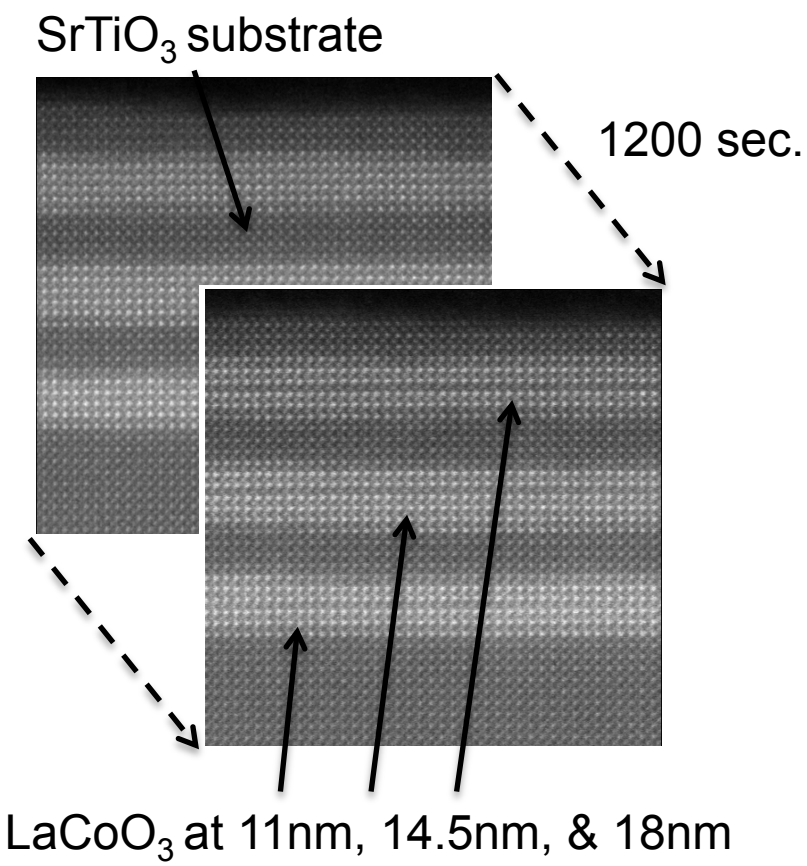


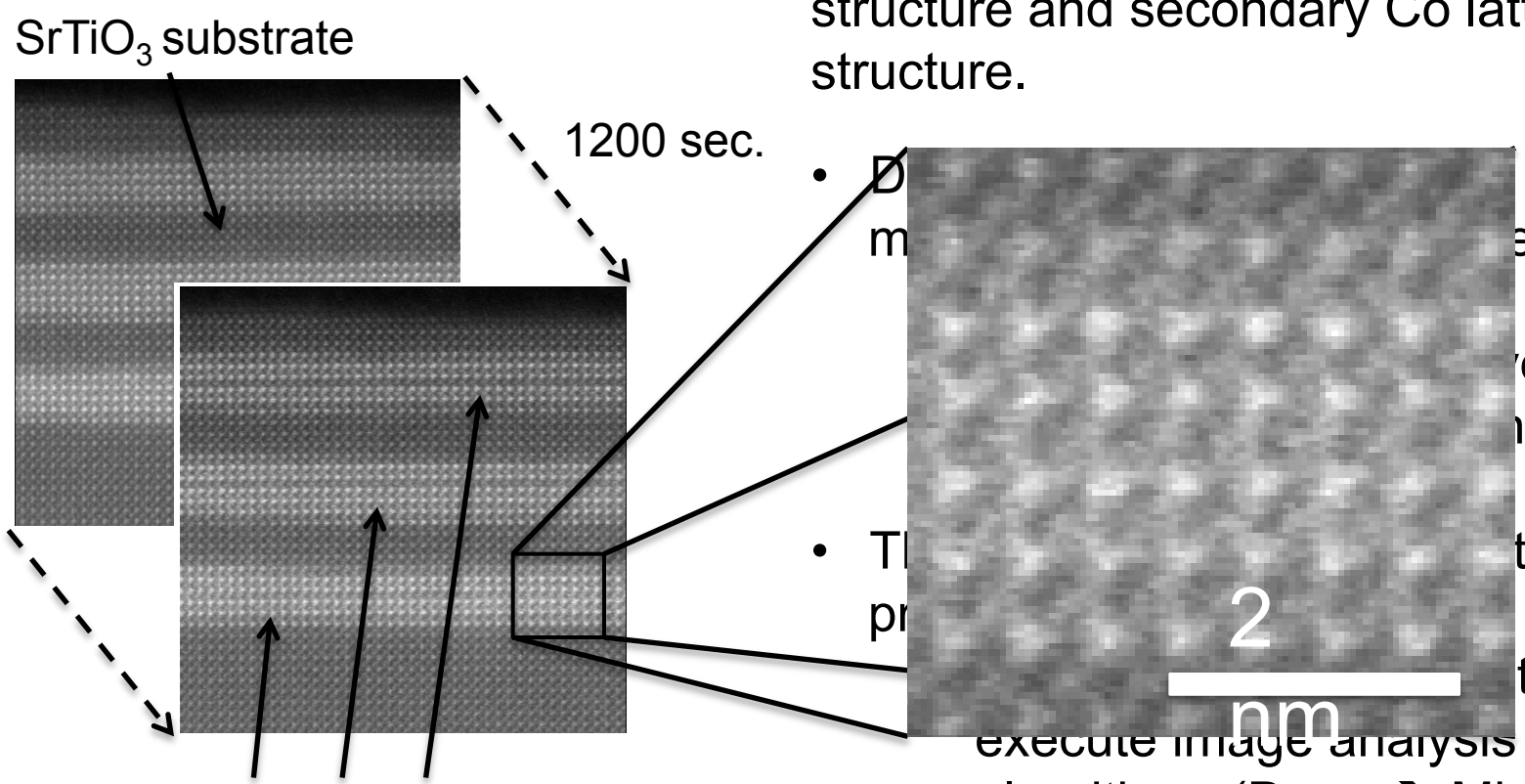
Image analysis goal: For the $LaCoO_3$ layers detect primary La lattice structure and secondary Co lattice structure.

- Drift and sometimes sharp movement in between images
  - Mathematical methods development to preserve physical properties of the system
- Thousands of images need to be processed
  - Produce scalable code to execute image analysis algorithms (Days ➔ Minutes).

Scalable Scientific Image Analysis – Demonstration Problem

Problem description:

Image analysis goal: For the $LaCoO_3$ layers detect primary La lattice structure and secondary Co lattice structure.



$LaCoO_3$ at 11nm, 14.5nm, & 18nm

- D
  m es
  e
  ne
- T to be
  pr
  to
  execute image analysis algorithms (Days ➔ Minutes).

Scalable Scientific Image Analysis – Research landscape

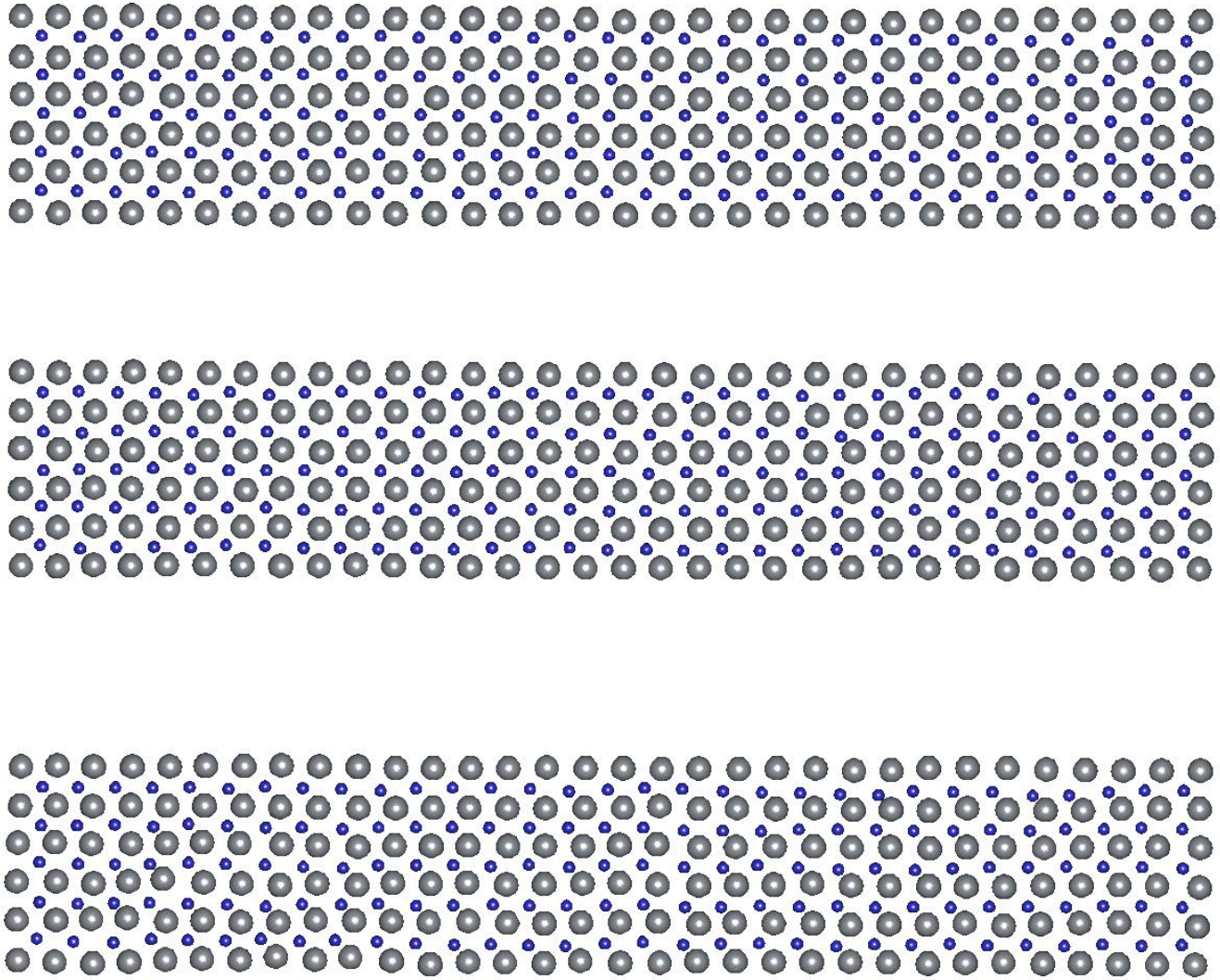
Image analysis is active field of research and should be leveraged

- Generalized method development is the norm
  - ***Limitation***: Methods not developed to take advantage of physical properties of studied systems.
  - ***Advantage***: New methods constantly developed.
- Algorithms valuable and often licensed and not developed at scale
  - ***Limitation***: Cost of running licensed software.
  - ***Opportunity***: Development of science specific, architecture aware code.

Computational & Mathematical – Capability Developed

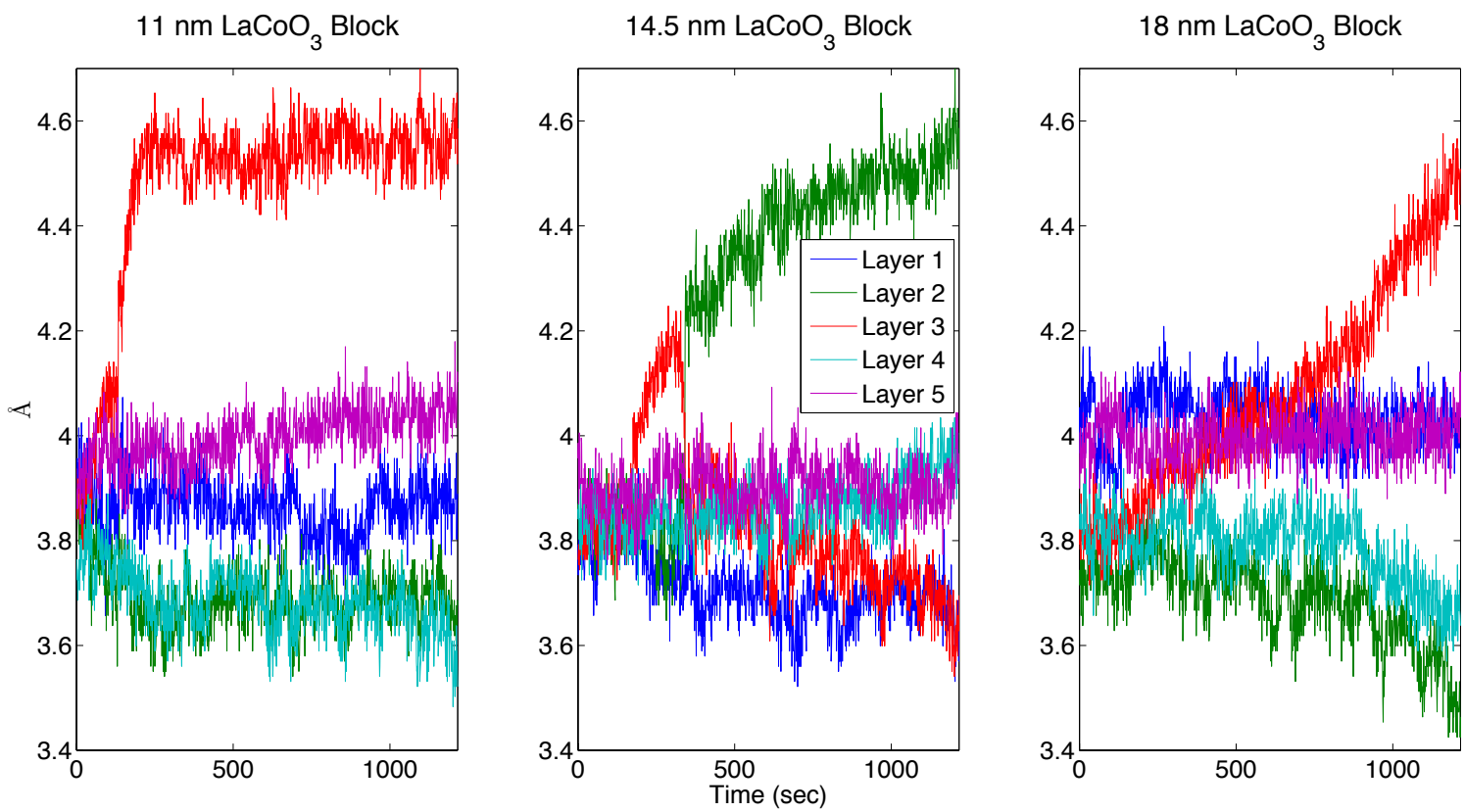
- Developing robust algorithms to detect atoms and place these atoms in lattices
  - Non-local filtering used to maintain structure and reduce image noise
  - Statistical modeling to determine most probable primary and secondary lattice
  - Minimize algorithmic propagation of errors by local and independent design
- Preserve physical properties of the system
  - Rigid transformations used to maintain lattice integrity
- Thousands of images need to be processed
  - Scalable algorithmic design
  - Identify levels of parallelism in workflow

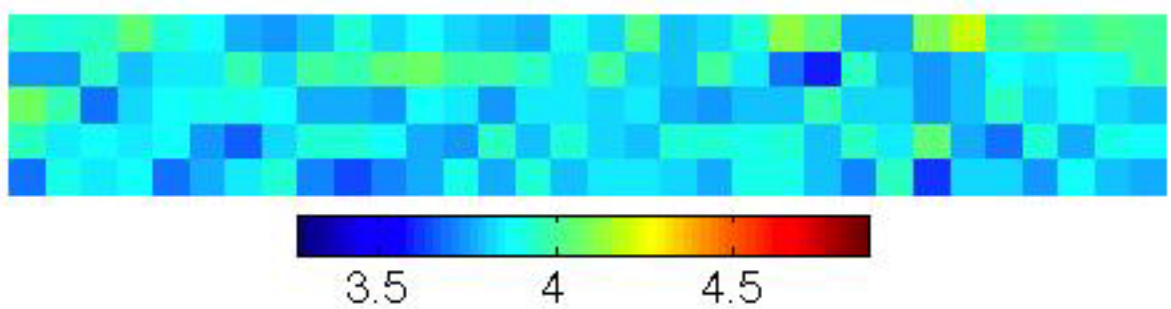
Scalable Scientific Image Analysis



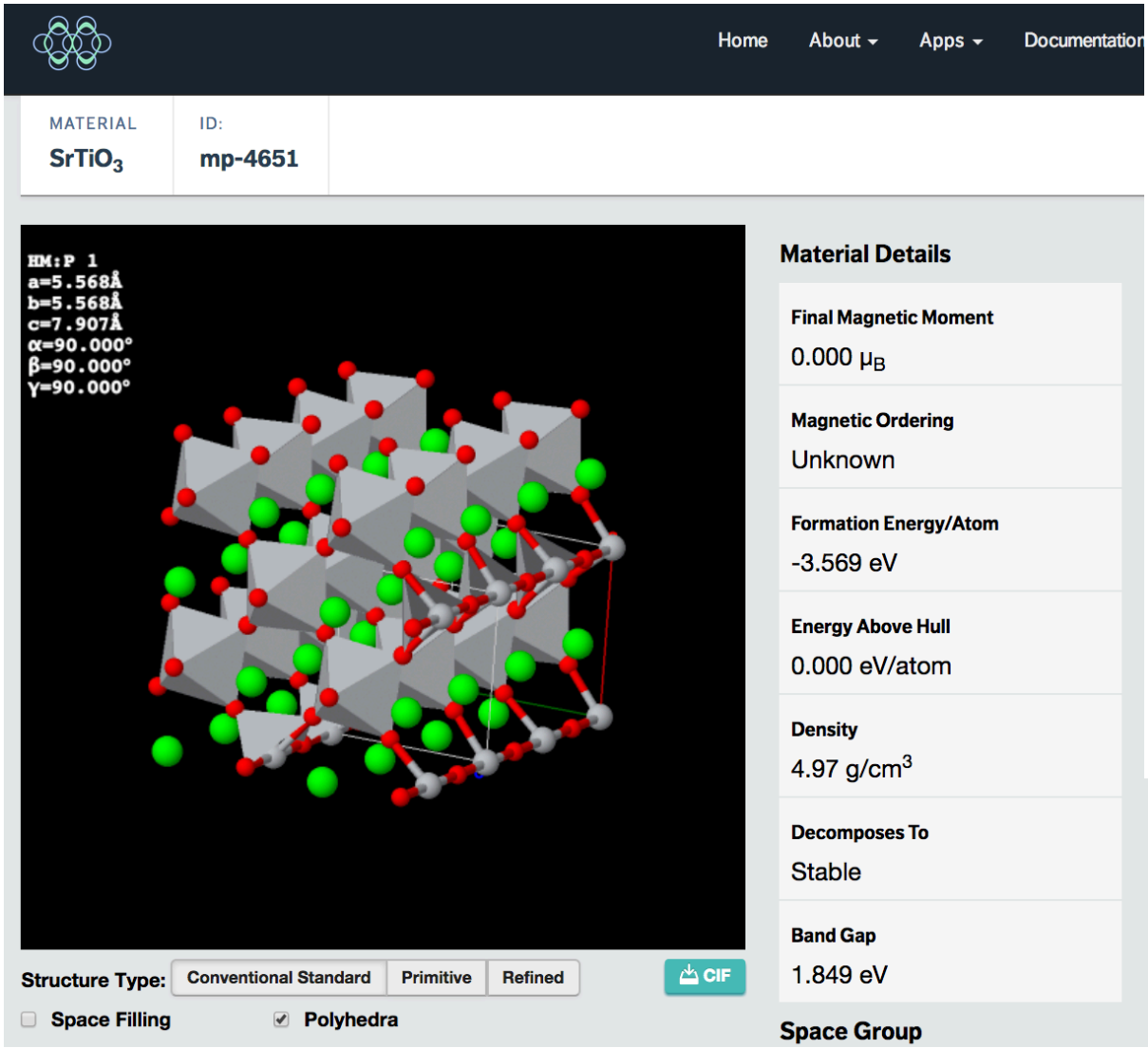
Primary and Secondary Lattice detection. Scaling analysis that would take days into minutes

Scalable Scientific Image Analysis





Augmenting Observation with Simulation



Bring calculated properties to assays

Functional electronic properties "painted on" image w/real-time DFT

Materials Project

<u>Search</u>: match known materials in library

<u>Discover</u>: real-time electronic structure via HPC

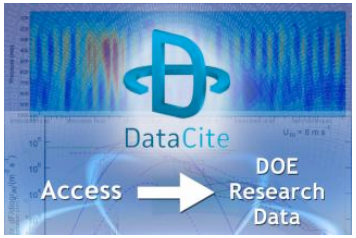
a	b	c
0.2291	0.7291	0
0.25	0.25	0.5
0.2709	0.2291	0
0.7291	0.7709	0

Supporting Digital Data Re-Use

STEM/Focused X-Ray

Feature Extraction

- Raw data
- Extracted features



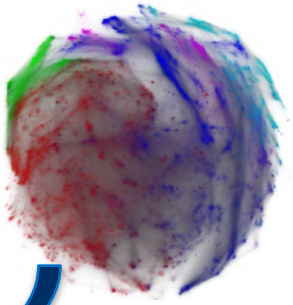
Related work
- Publications

Data Catalog

High-performance Data Infrastructure

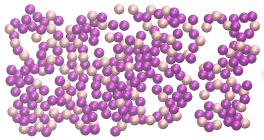
Access, Citation & Data reuse

osti.gov



Citation Graph
- Tobias Kuhn, ETH Zurich

Simulation

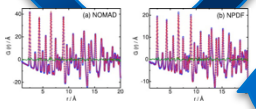


Molecular Dynamics

$$i\hbar \frac{d\Psi(\{r_i\};t)}{dt} = \hat{H}\,\Psi(\{r_i\};t)$$

$$= \sum_{i=1}^{N_e} \nabla_i^2 + \sum_{i=1}^{N_e} V_{nuclear}(r_i) + \sum_{i=1}^{N_e} V_{effective}(r_i)$$

Density Functional Theory



Re-analysis & Derived data Products

CADES
Compute & Data Environment for Science

http://doi.ornl.gov/data
https://www.materialsproject.org/materials/mp-4651/

ORNL CADES DOI Service

doi.ornl.gov/login

CADES DOI Service

HOME Register Data Search

CADES Digital Object Identifier (DOI) System

Dataset DOI Registration

CADES DOI Registration Form

Title	Dataset Name
Description	Description
Product Numbers	Available Product Numbers (Identifier)
Publication Date	Date of the Resource Publication
Creator Emails	Registrar's Email
Contact Name	Primary Contact for the Resource
Contact Organization	Primary Contact's Organization
Contact Numbers	Primary Contact's Phone Number
Contact Email	Email of the Primary Contact
Originating Organization	Resource Originating Orgamization
Sponsor Organization	Resource Sponsor Organization
Contributor Organization	Contributing Organizations
Language	Resource Language
Country	Country of Origination

ORNL CADES DOI Service

doi.ornl.gov/data

Image Analysis and Feature Classification in Atomically Resolved Images: Genomic Libraries through Smart Data

Rick Archibald, Alex Belianinov, Albina Borisevich, Miaofang Chi, Eirik Endeve, John Freeland, Doug Fuller, Anubhav Jain, Stephen Jesse, Sergei Kalinin, Guruprasad Kora, Nouamane Laanait, Kristin Persson, Amanda Petford-Long, Andrew Lupini, Oleg Ovchinnikov, Galen Shipman, David Skinner, Chris Symons, Ray Unocic.

READ MORE

ORNL CADES DOI Service

doi.ornl.gov/data/doi/10.1037.rmh0000009

File: 0.h5

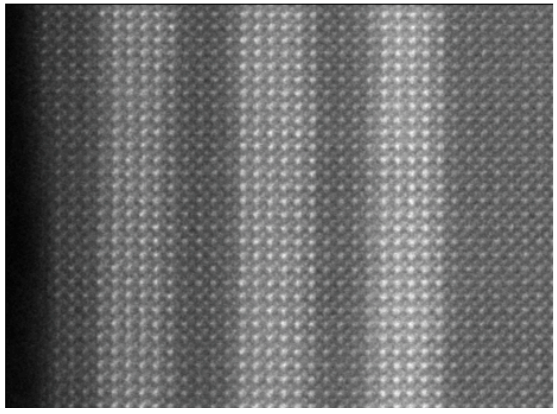
HAADF Image of LaCoO3 from Nion UltraSTEM 100 – HDF5 format

Published November 13, 2014 | Download 0.h5

LaCo03 CNMS STEM HAADF Nion HDF5

File: 0.png

HAADF Image of LaCoO3 from Nion UltraSTEM 100 – PNG format

Published November 13, 2014 | Download 0.png

LaCo03 CNMS STEM HAADF Nion PNG