~~Something about neutrons~~

S.J.L. Billinge

Department of Applied Physics and Applied Mathematics

Columbia University,

CMPMS, Brookhaven National Laboratory

HTTP://thebillingegroup.com

Three more or less unconnected things

S.J.L. Billinge

Department of Applied Physics and Applied Mathematics

Columbia University,

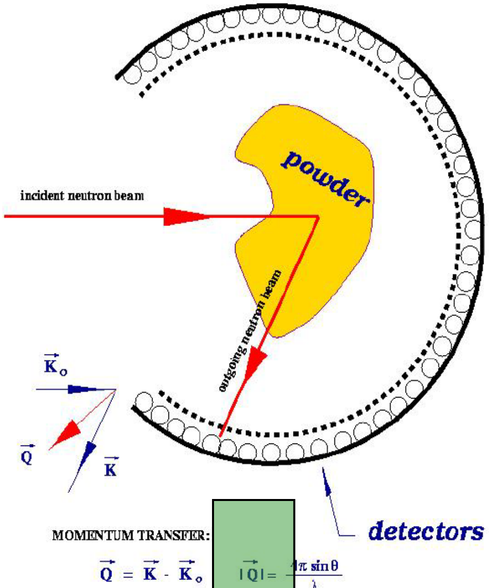
CMPMS, Brookhaven National Laboratory

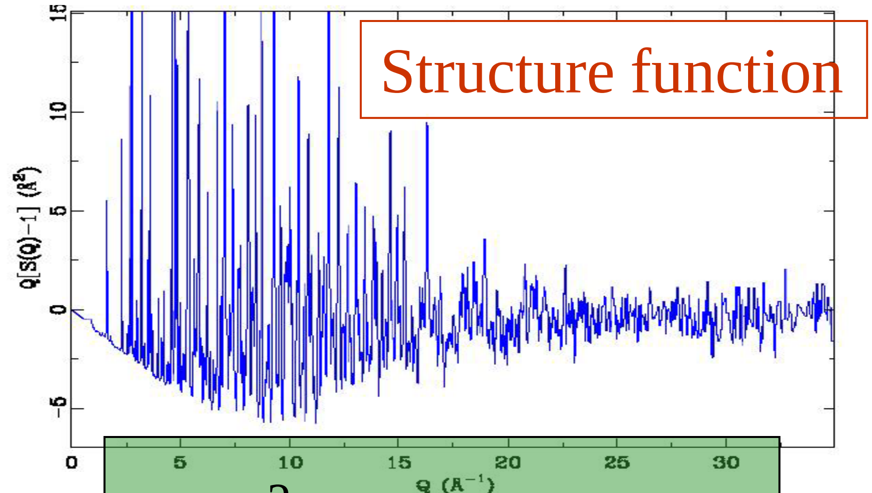
HTTP://thebillingegroup.com

Overview

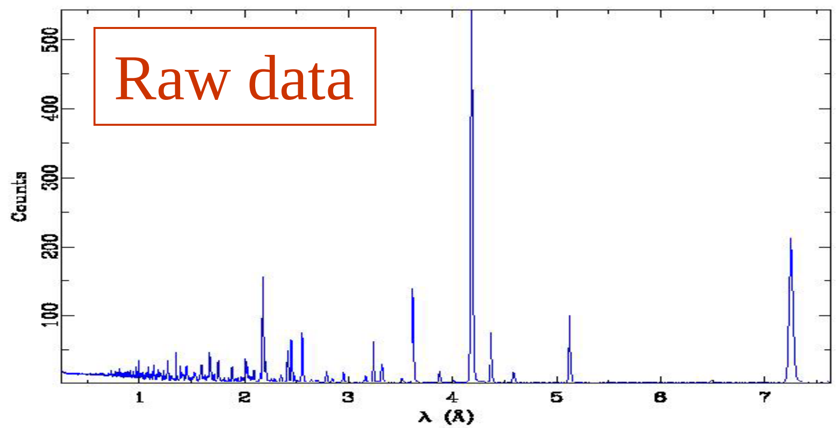
1. Getting Started in PDF-land
2. High throughput powder diffraction and PDF
    1. Why bother
    2. How
3. Billinge-group software development workflow

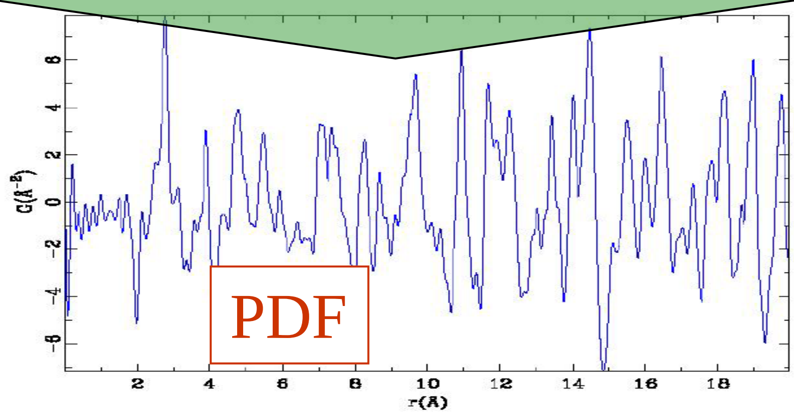
The atomic Pair Distribution Function





$$G(r) = \frac{2}{\pi}\int_0^\infty Q[S(Q)-1]\sin Qr\,dQ$$



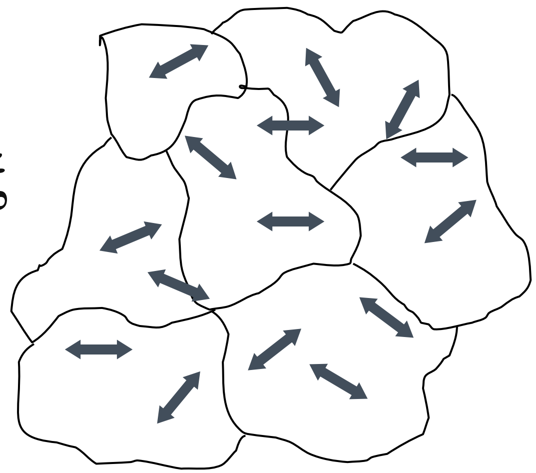


COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

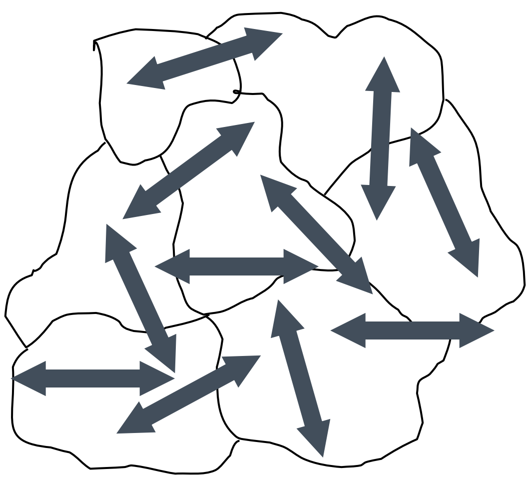
BROOKHAVEN NATIONAL LABORATORY

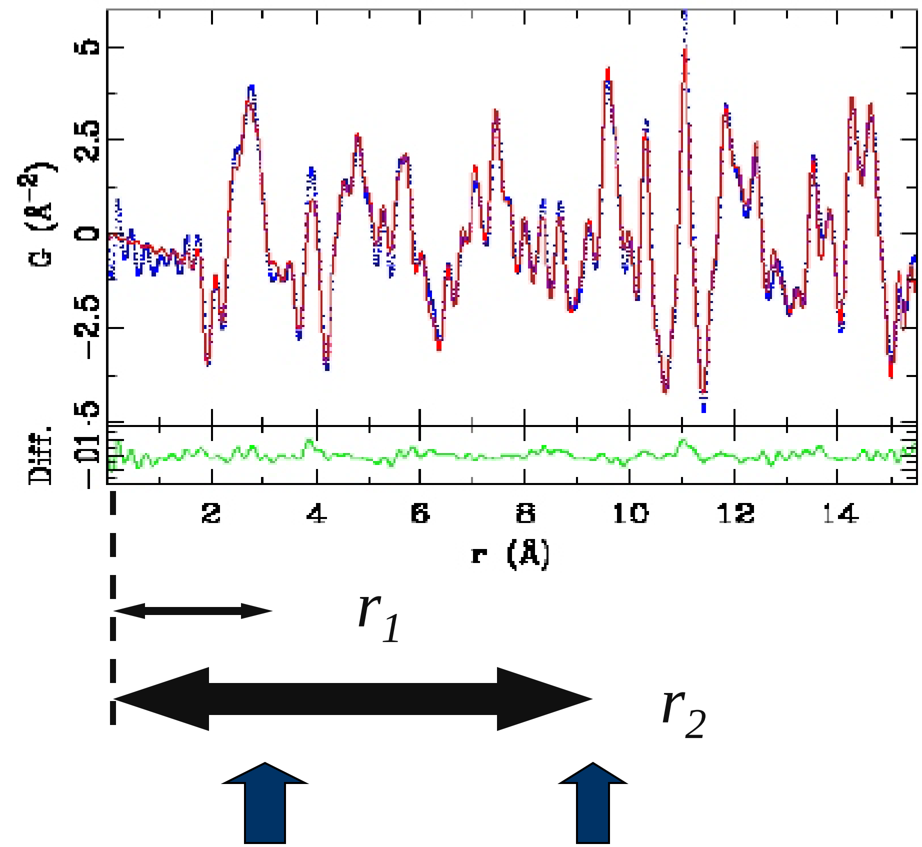
HTTP://thebillingegroup.com

$r_1 << \xi$



$r_2 \sim \xi/2$





Intra-domain structure

Inter-domain structure

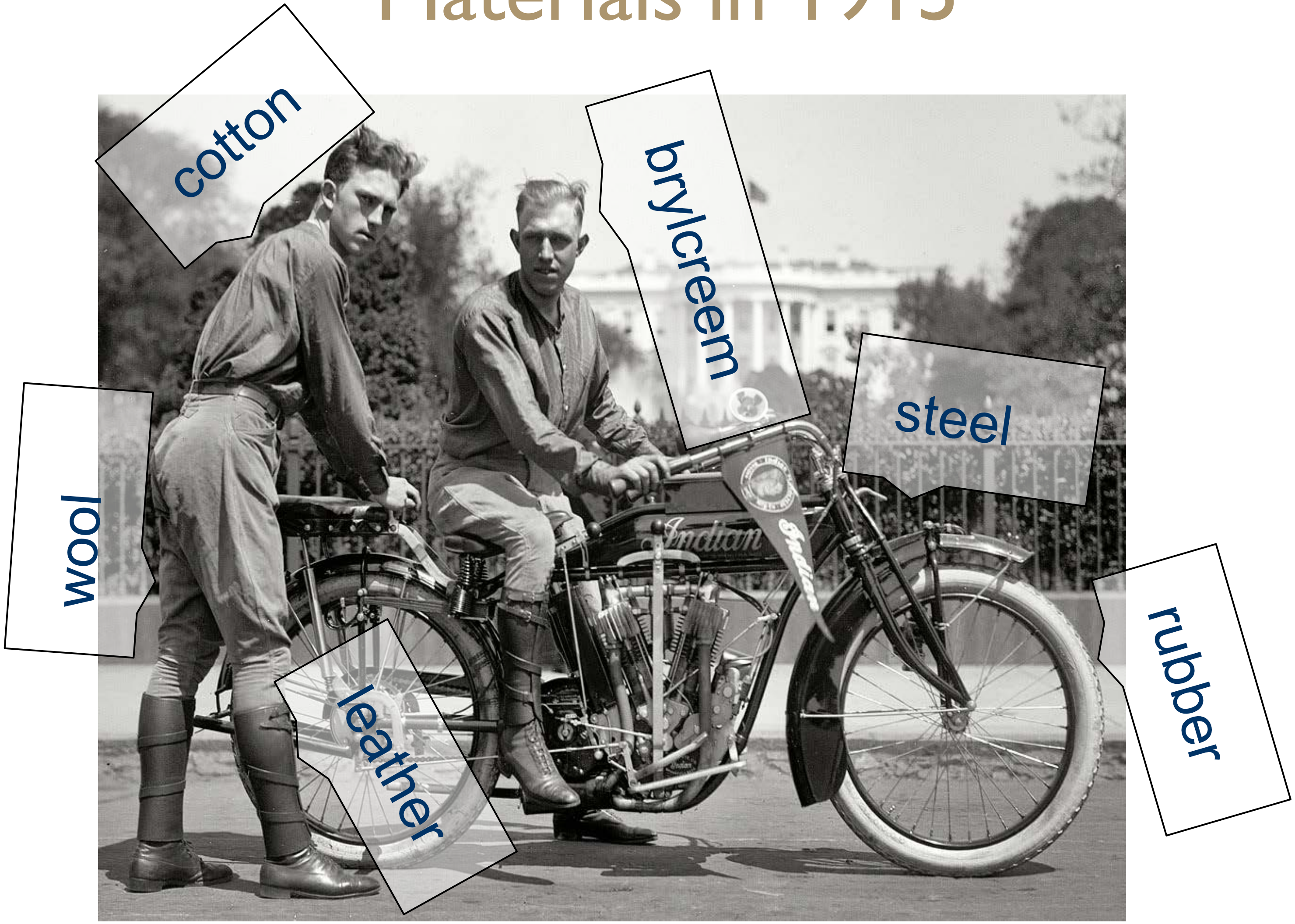
Getting Started in PDF

1. Email me (sb2896@columbia.edu)
    1. Remind me that you were at this meeting and you want to measure a sample with x-rays to get PDF data
2. I will put you in touch with a student in the group and you can make arrangements to ship samples to us
3. We will measure preliminary data and send it to you
4. We will help you to get going with the analysis on your own using our programs
5. We can help with data interpretation, but you must learn to do the analysis yourself
6. If more experiments are needed we can help you to write a general user proposal and could possibly collect data for you if successful

High Throughput Powder/PDF

- Why?
- How?
- Computational issues

Materials in 1915

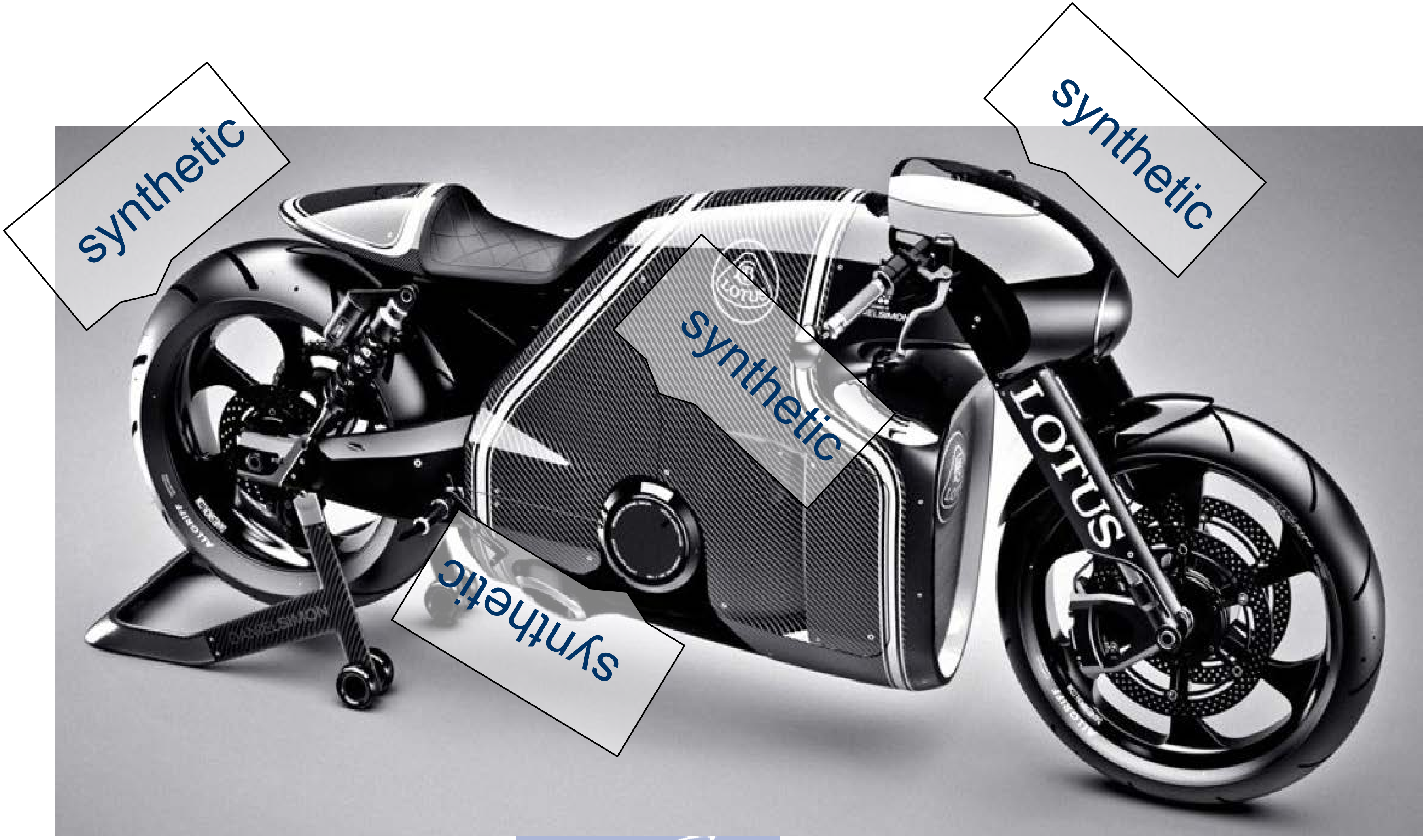


COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

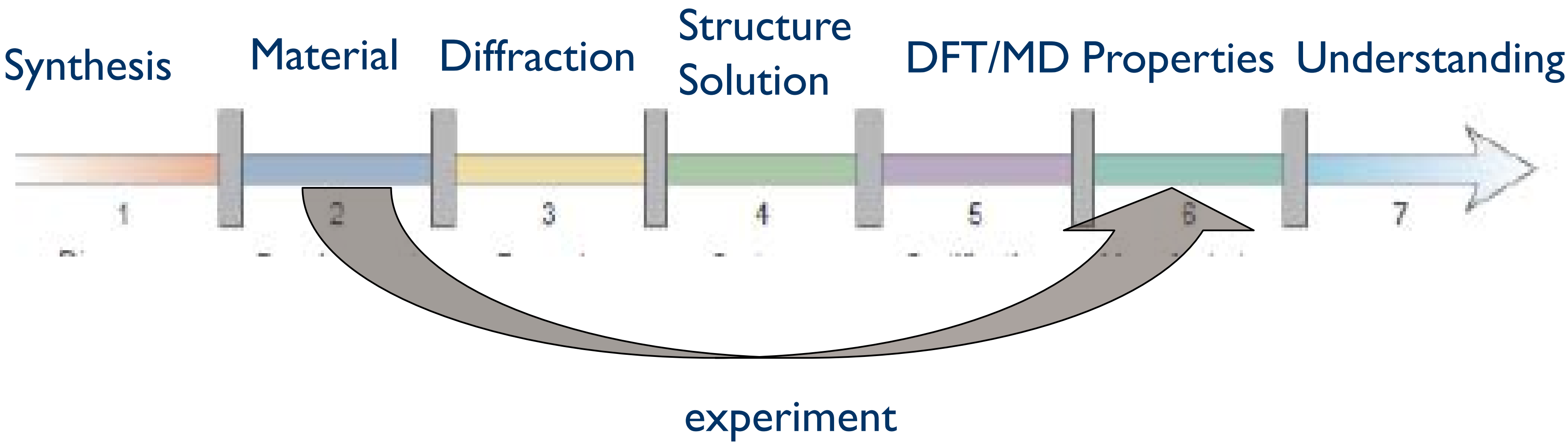
BROOKHAVEN
NATIONAL LABORATORY

HTTP://thebillingegroup.com

Materials 2015



Current paradigm



- Key step is Diffraction -> Structure solution. Understanding flows from that.
- The structure solution step is crystallography
  - According to IUCr, 48 nobels associated with crystallography (some loosely!)
- Crystals are idealizations of real material structure

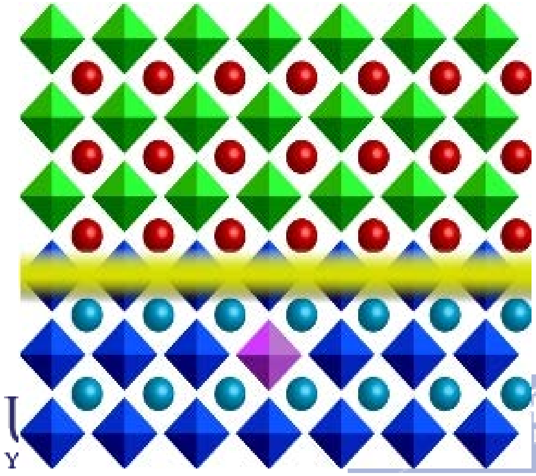
Real Materials: more complex than ideal crystals

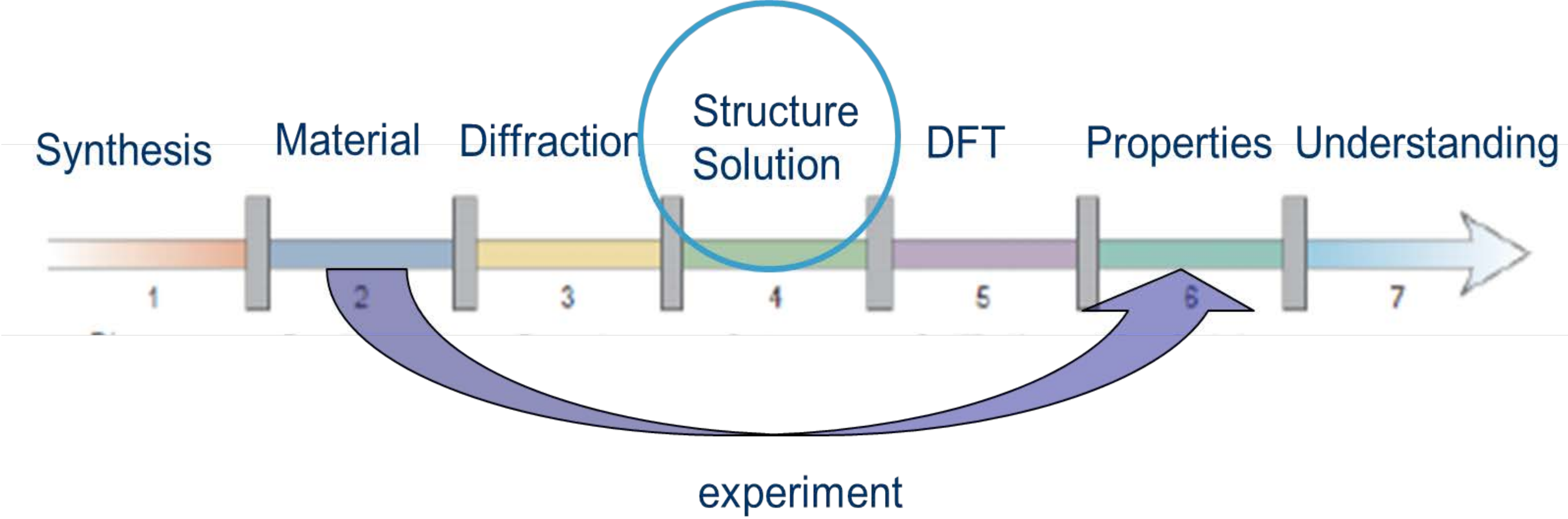
- Real-Material Structure model:
  - Crystal structure (if there is one)
  - Morphology (could be nano)
  - Surface reconstruction
  - Surface termination/dressing (ligands etc.)
  - Interfaces
  - Heterogeneities, phase separation
  - Point defects
  - Extended defects
  - Chemical short-range order
  - Distortive short-range order
  - …

Real material properties depend sensitively on crystalline imperfections

Examples

- Optical properties of quantum dots depend on presence or absence of surface trap states
- Photovoltaic performance depends on charge transfer and charge extraction
- Catalysis depends on surface structure
- Battery electrodes depend on access of lithium
- Broken symmetry states in the PG phase of HTSCs
- Place two insulators together and get superconductivity in the interface



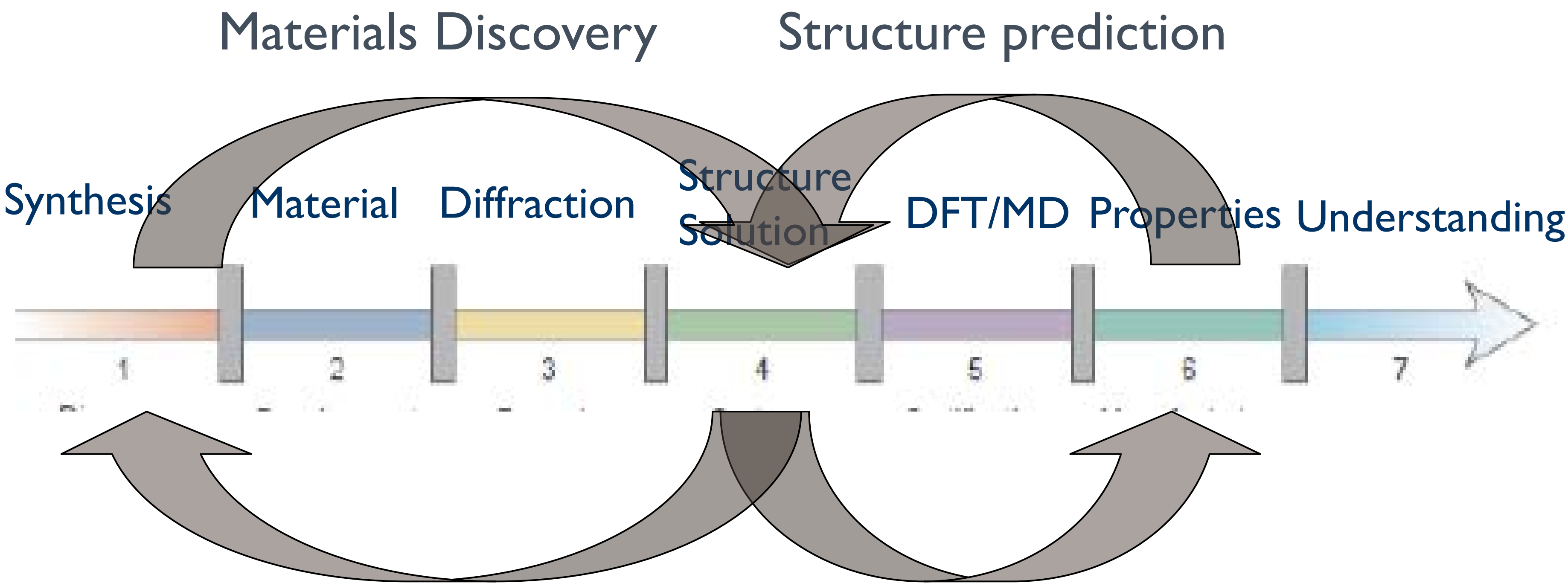


We need to do structure solution from

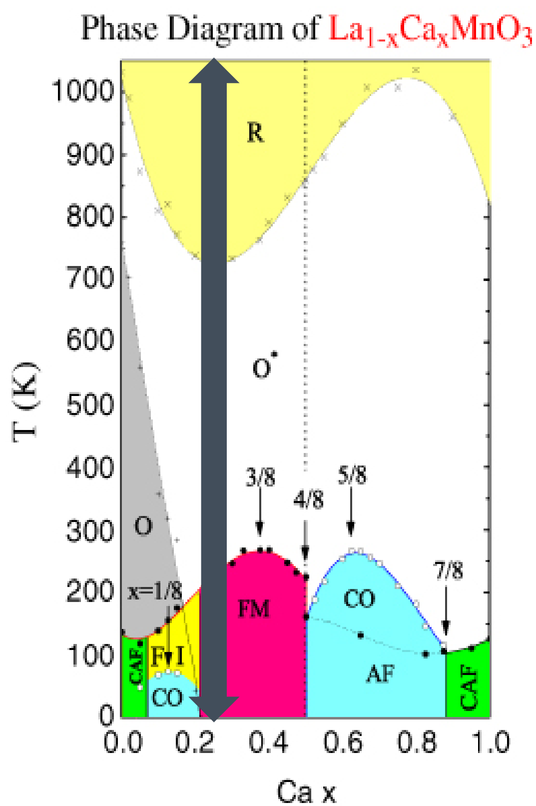
Real Materials

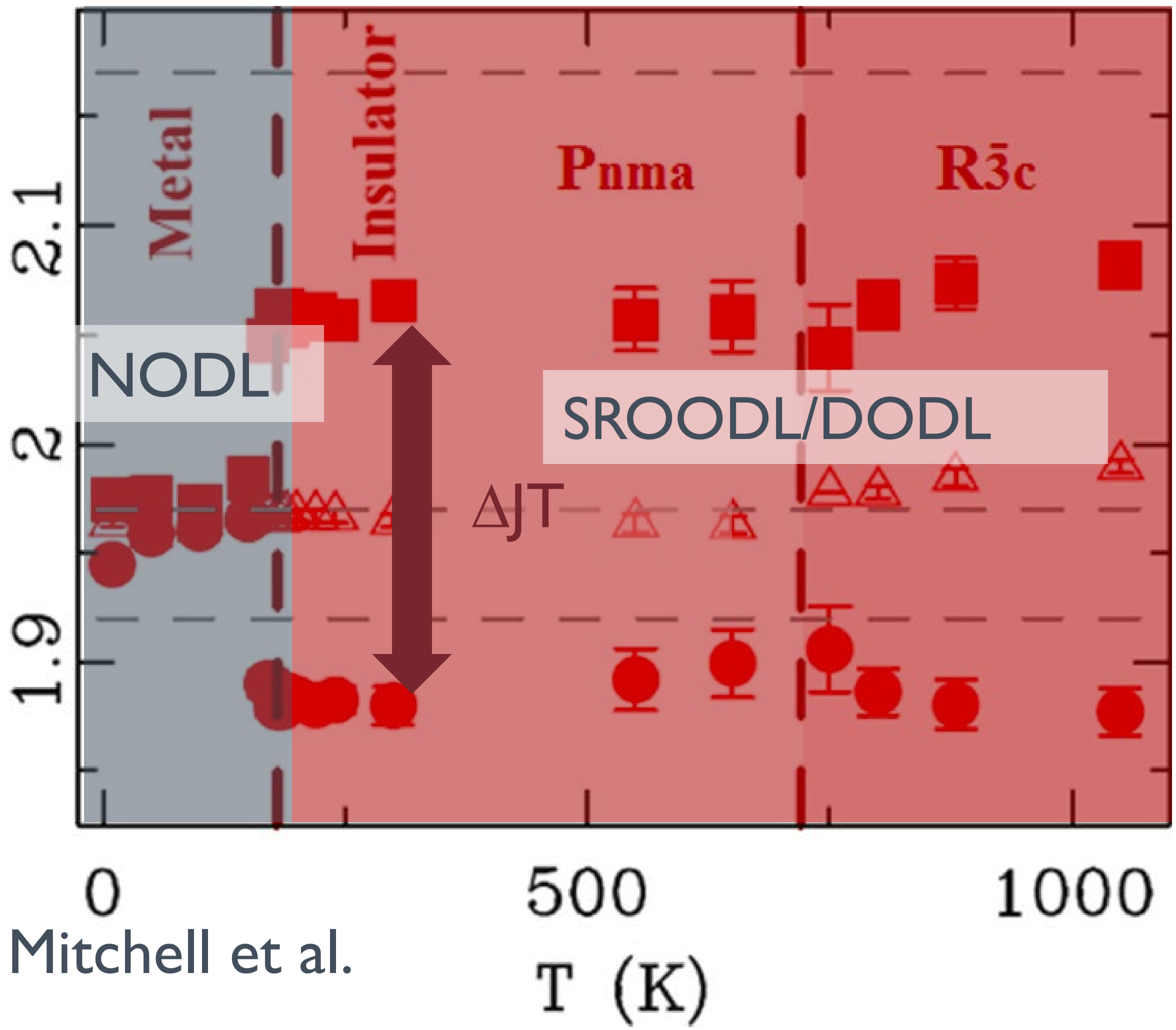
3D xyz coordinates of atoms with high precision

Materials Genomics



Polaron "liquid" phase in manganites





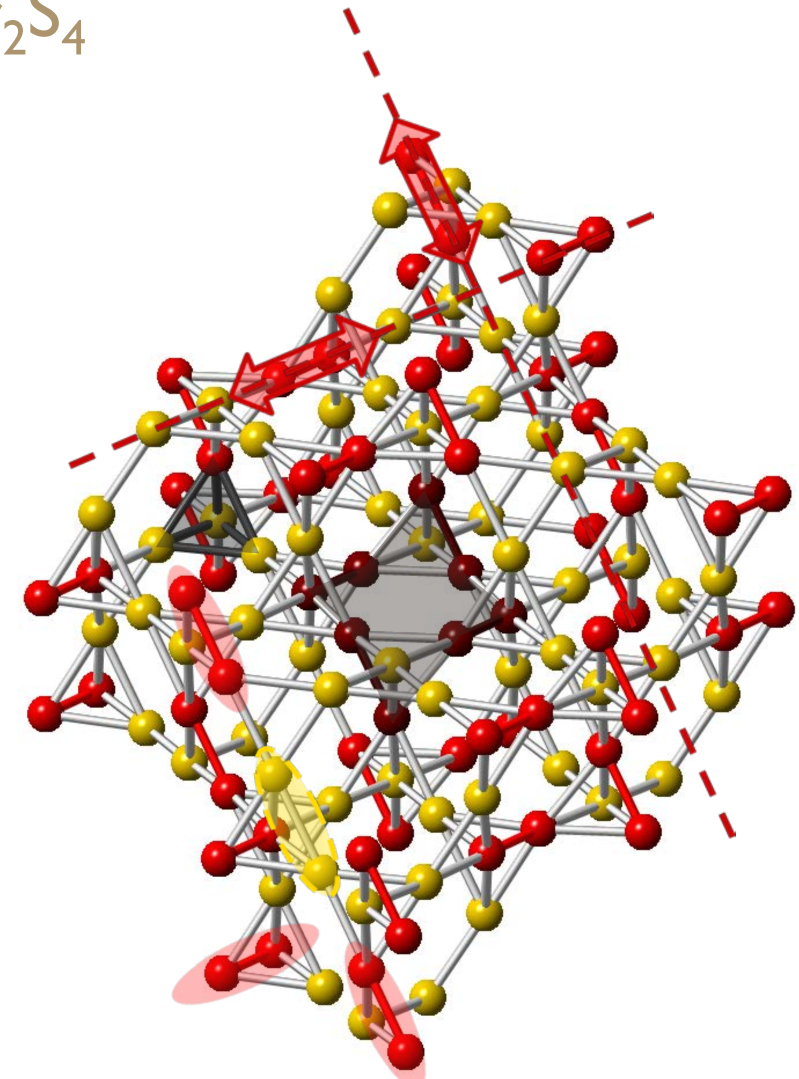
X=0.22

Masadeh, Bozin, SJLB, Mitchell et al. PRB (2016)

ODL and the Metal-Insulator Transition in $CuIr_2S_4$

Emil Bozin
w/ JF Mitchell, ANL
M. Abeykoon, BNL

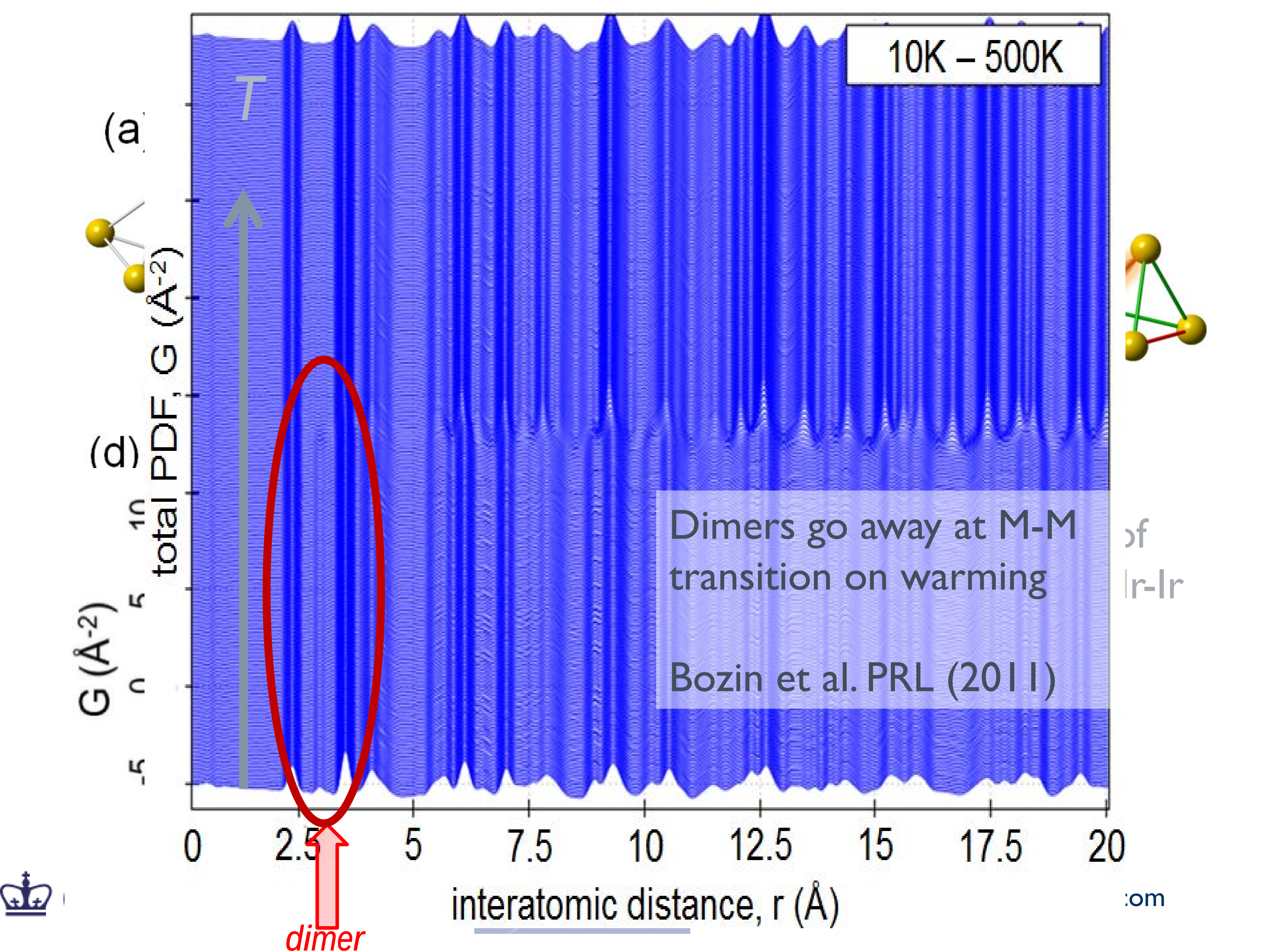
Data from 28-ID-2 NSLS-II



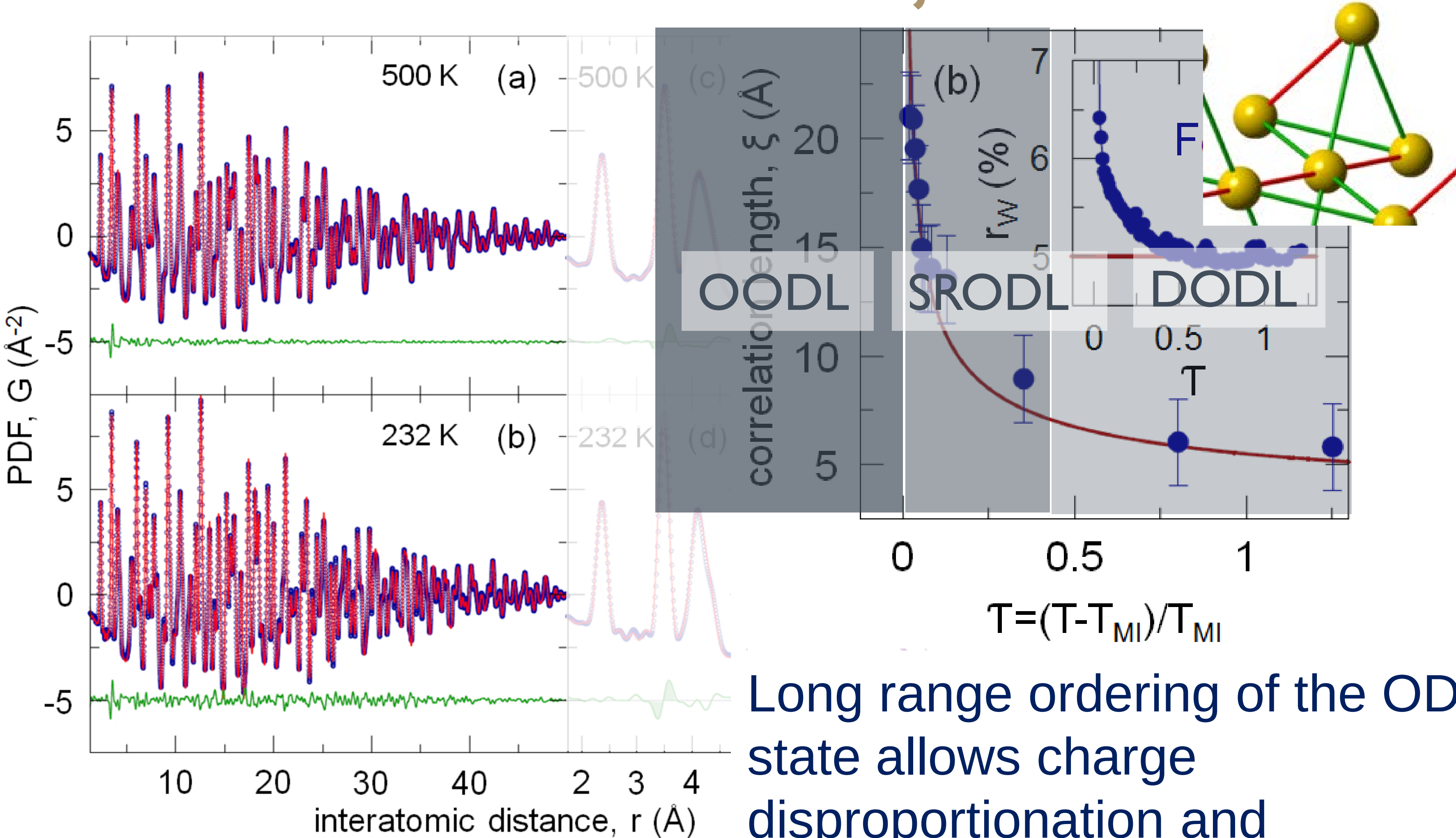
COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

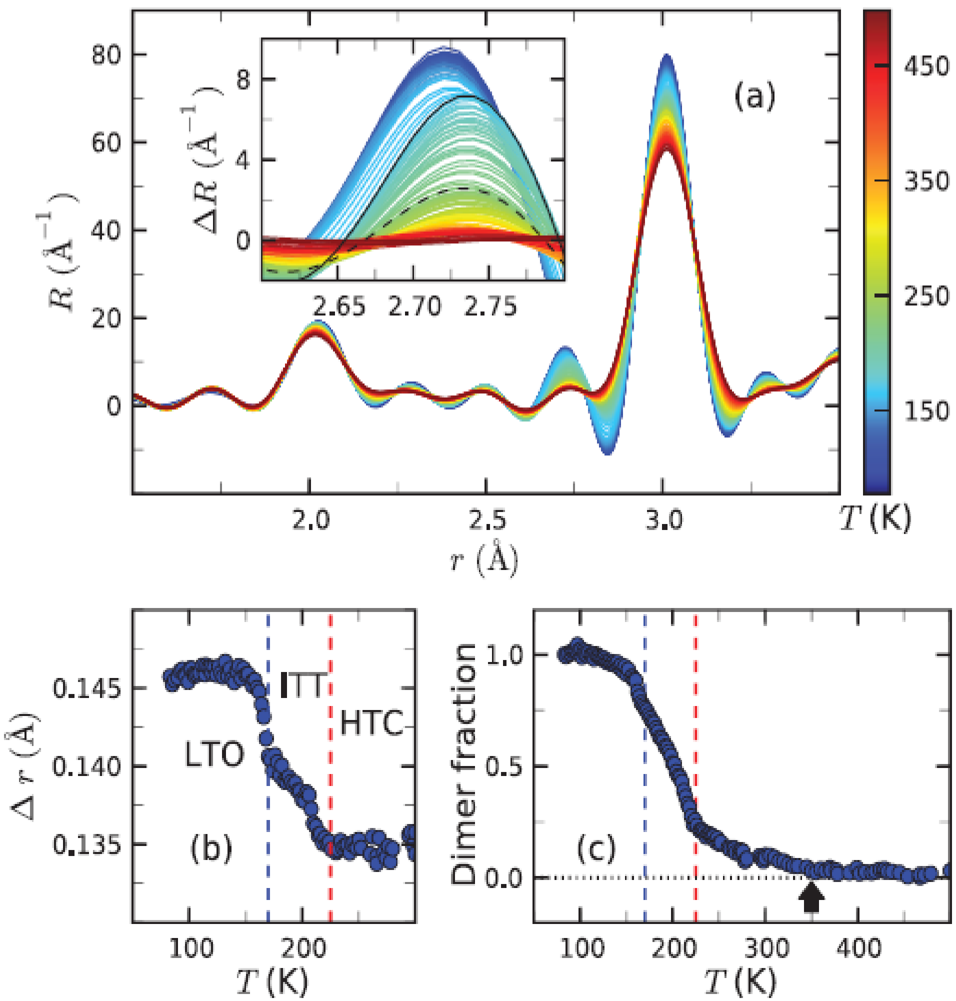

10K – 500K
T
(a)
(d)
total PDF, G (Å^{-2})
G (Å^{-2})
Dimers go away at M-M transition on warming
Bozin et al. PRL (2011)
0
2.5
5
7.5
10
12.5
15
17.5
20
interatomic distance, r (Å)
dimer

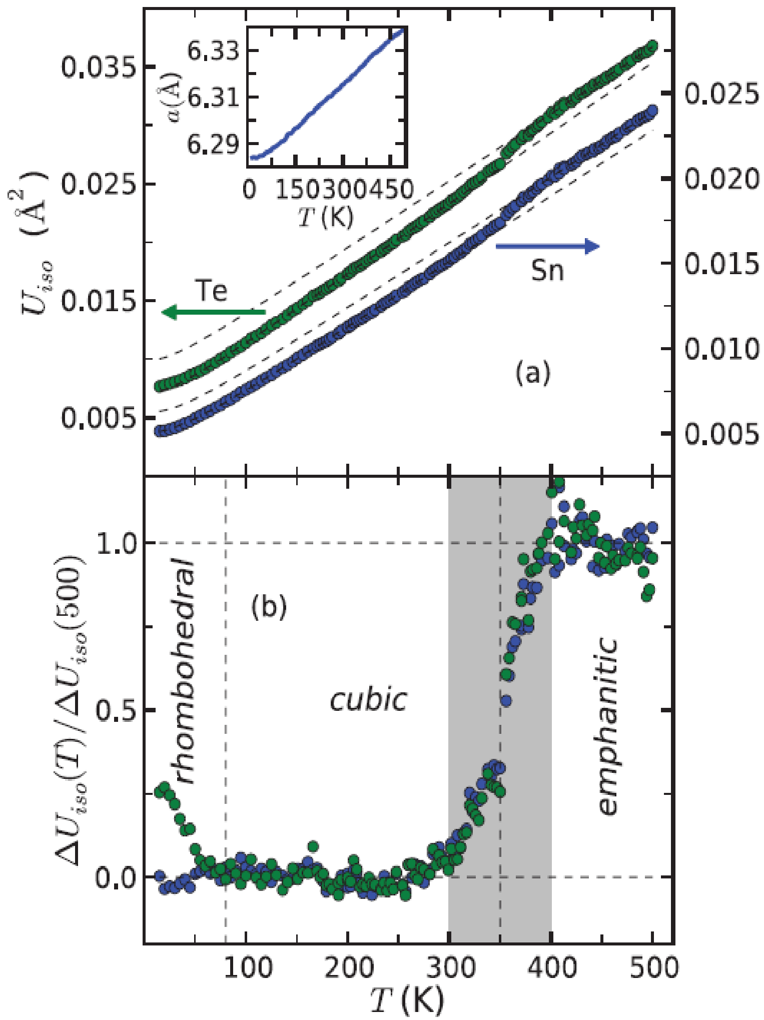
What happens to the correlation length of SRO of the ODL objects?



Long range ordering of the ODL state allows charge disproportionation and dimerization

High throughput allows detailed temperature studies



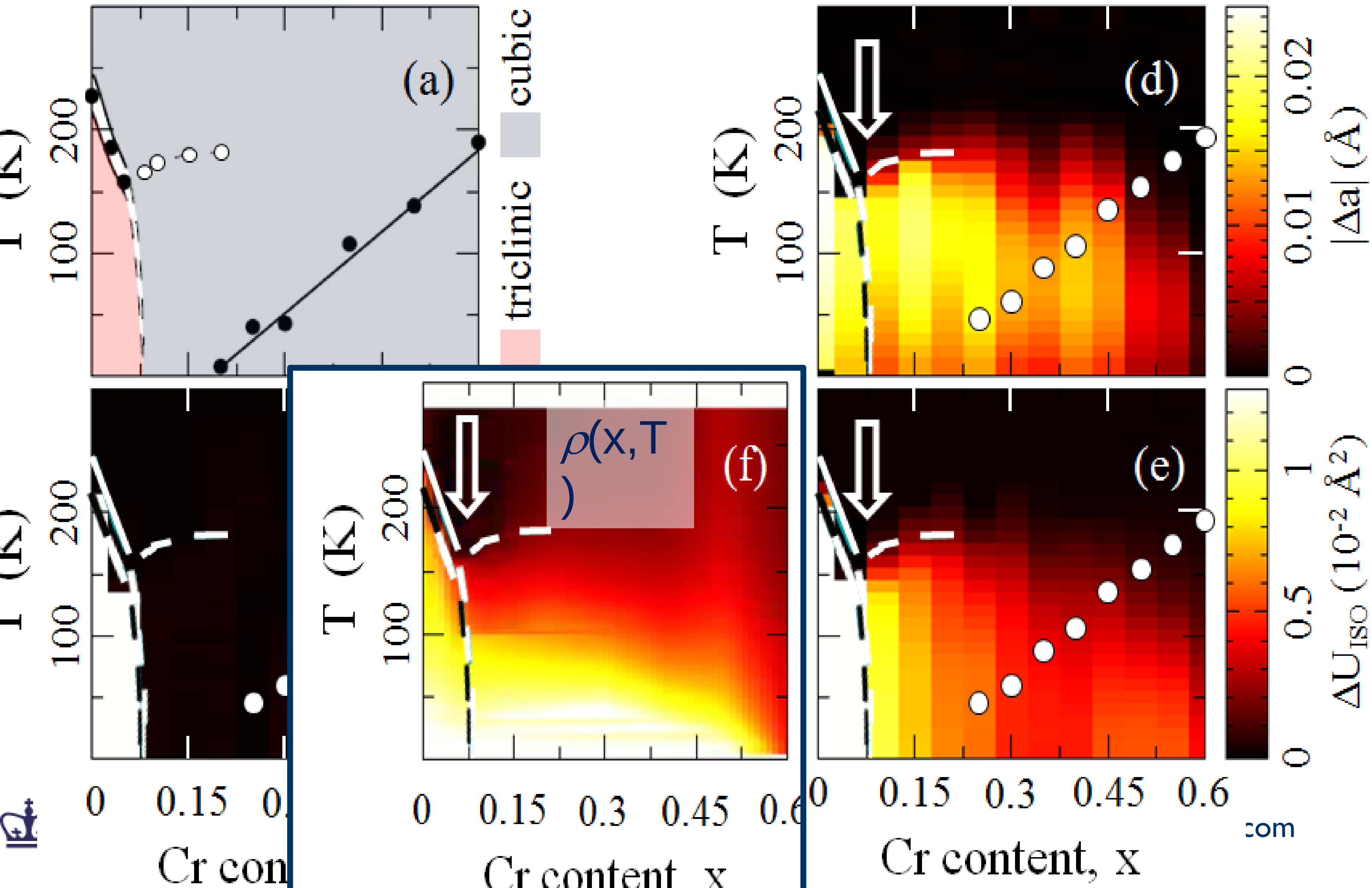


- Evolution of Ir dimers vs T in Li_2RhO_4
- Knox, Bozin, sjlb et al, PRB 13

- Evolution of anharmonicity/off centering in SnTe
- Knox, Bozin, SJLB et al PRB 14

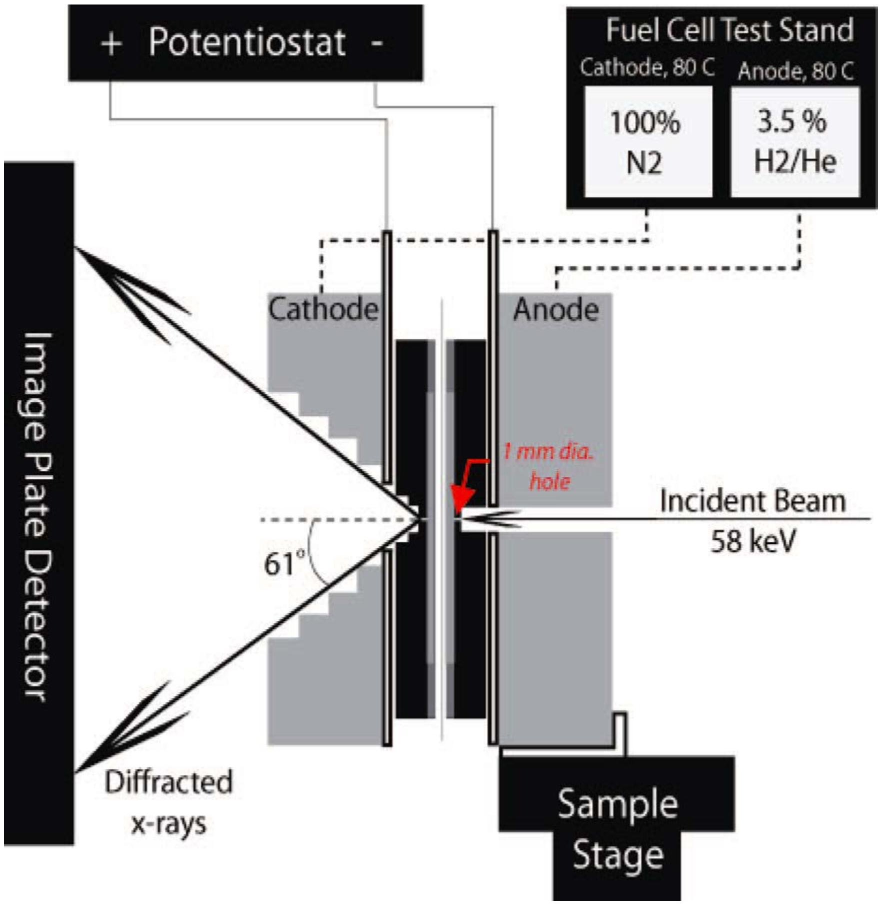
Avg str

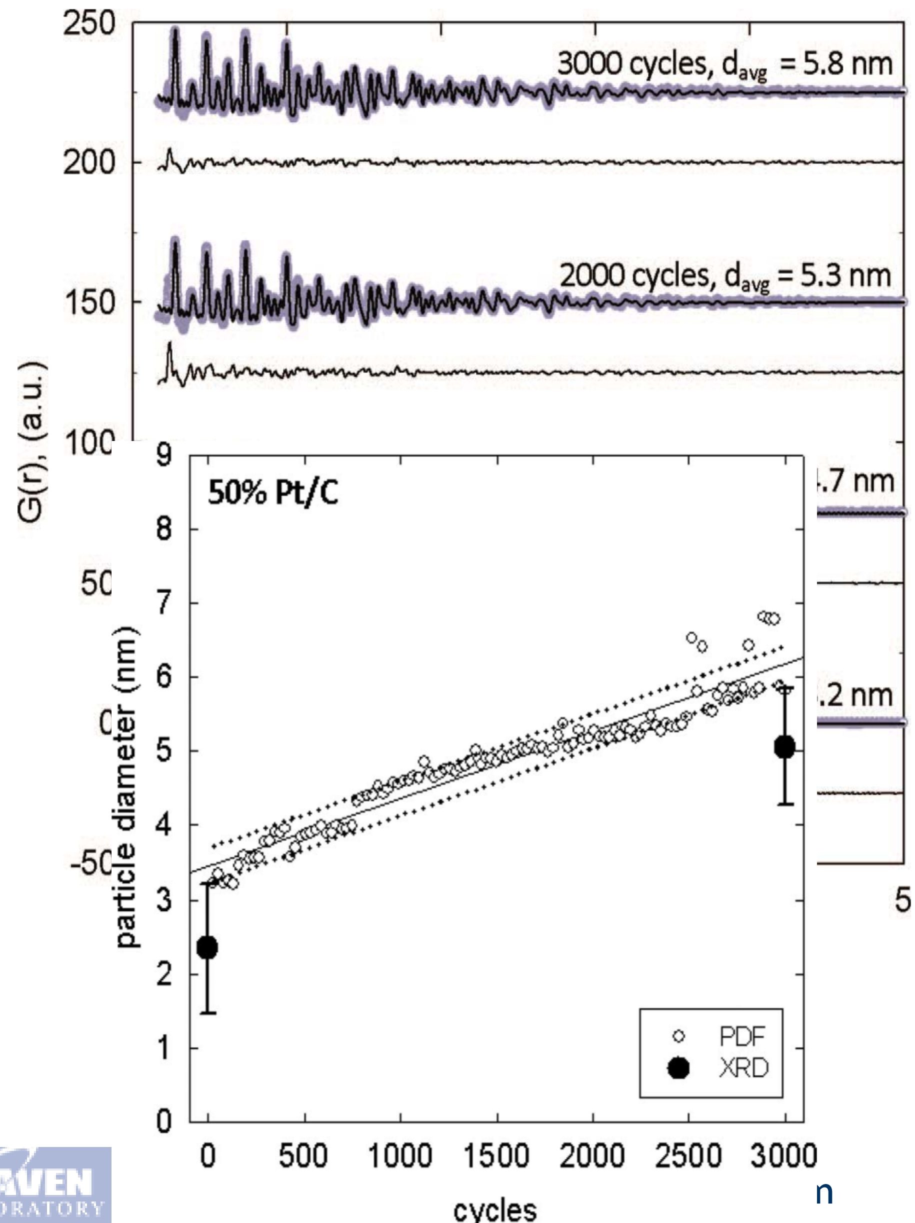
Local str



.com

In Situ



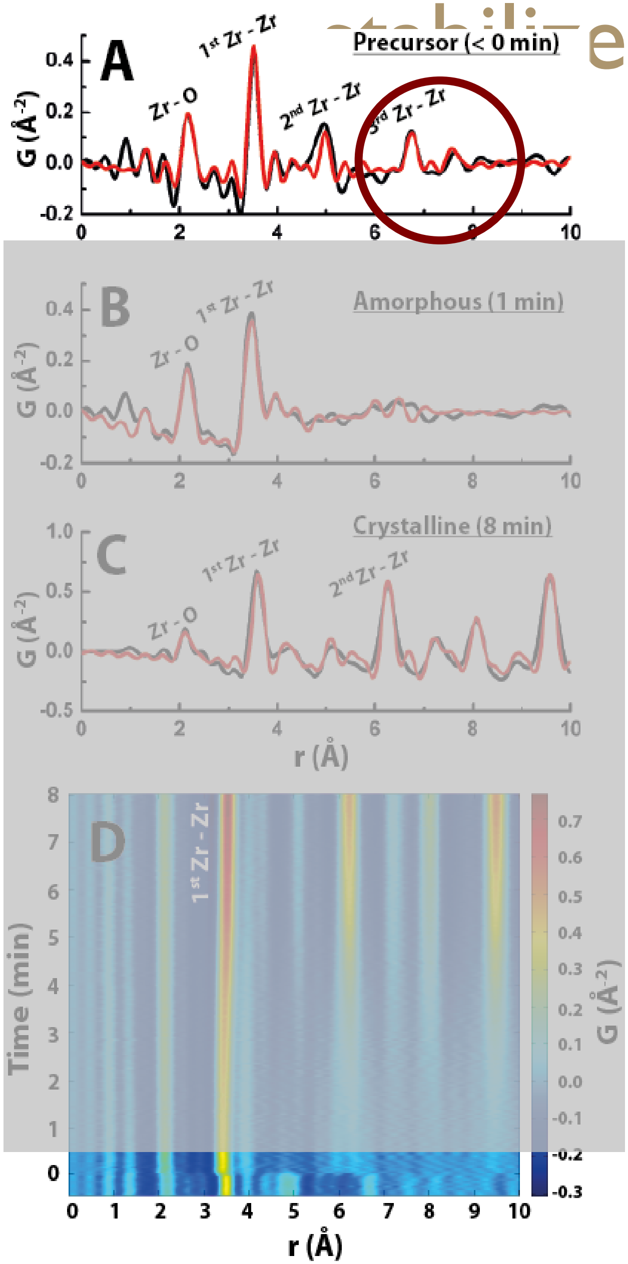


- Redmond, et al., *Electrochem. Solid St., 15* (5), B72–B74.

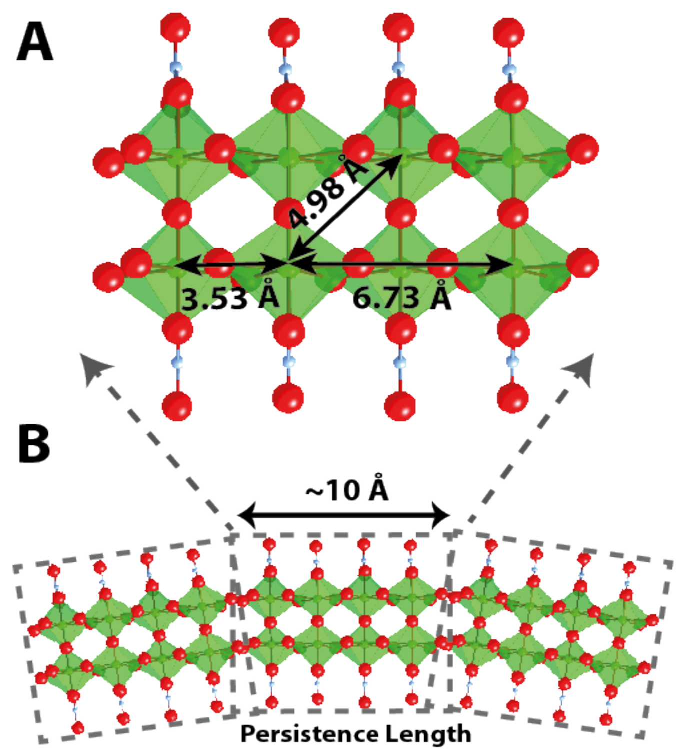
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

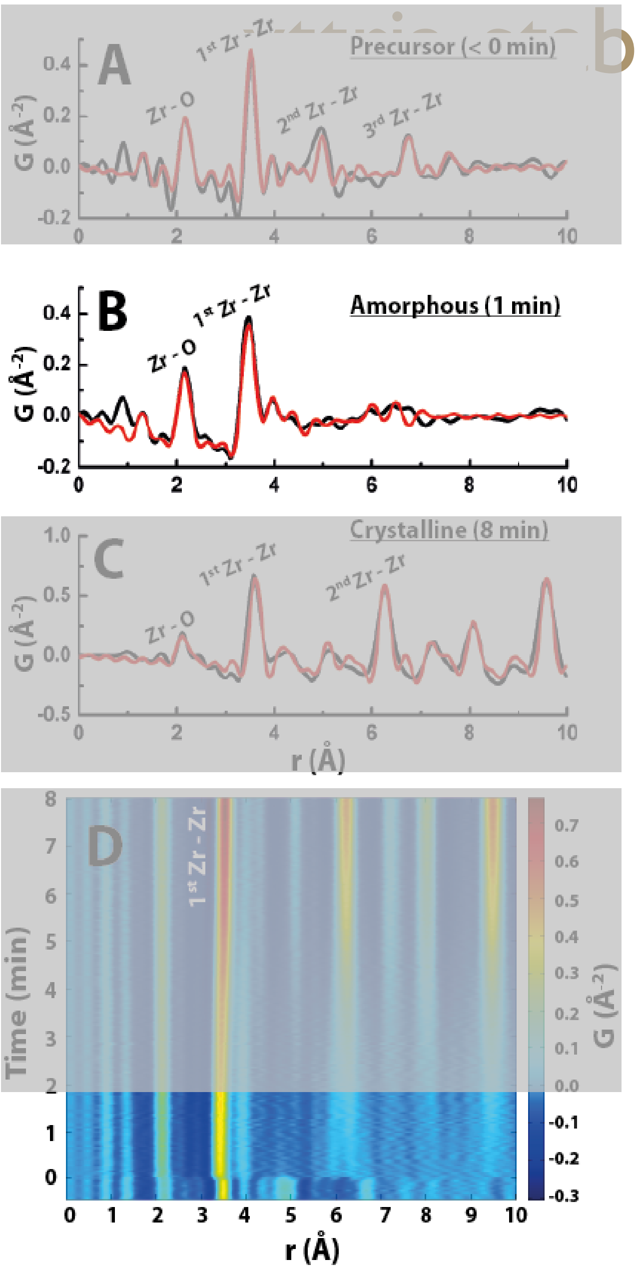
In-situ study of hydrothermal synthesis of yttria-stabilized zirconia nanoparticles



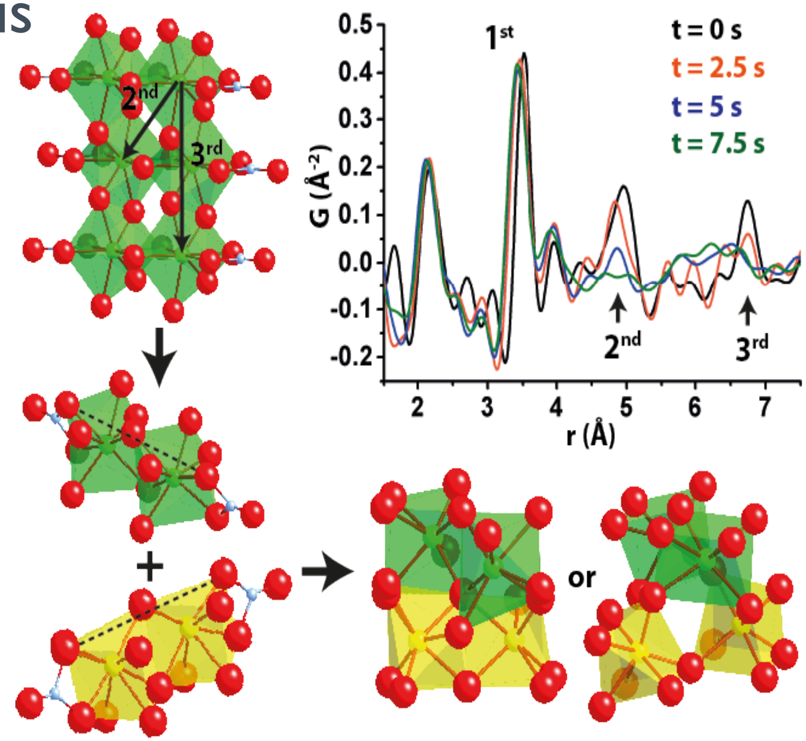
- Christoffer Tyrsted, Nina Lock, Kirsten M. \O. Jensen, Mogens Christensen, Espen D. Bøjesen, Hermann Emerich, Gavin Vaughan, Simon J. L. Billinge and Bo B. Iversen, *IUCrJ.* **1**, 165-171 (2014)
- Precursor



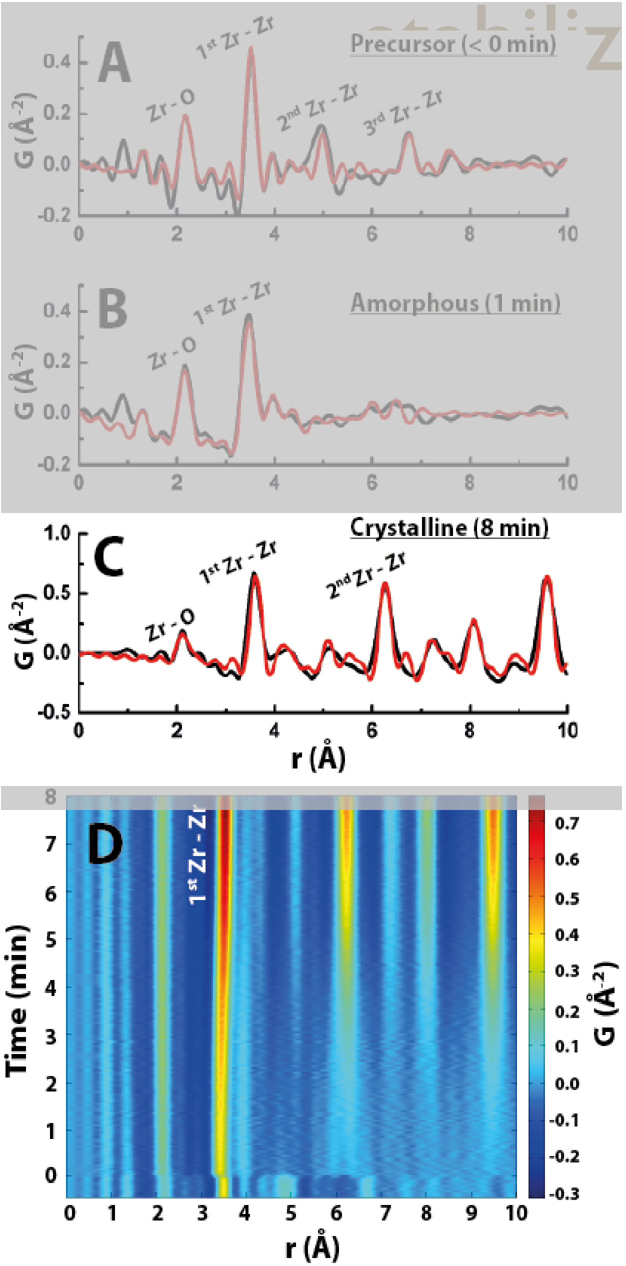
In-situ study of hydrothermal synthesis of bilized zirconia nanoparticles



- Christoffer Tyrsted, Nina Lock, Kirsten M. \O. Jensen, Mogens Christensen, Espen D. Bøjesen, Hermann Emerich, Gavin Vaughan, Simon J. L. Billinge and Bo B. Iversen, *IUCrJ.* 1, 165-171 (2014)
- Amorphous intermdiate

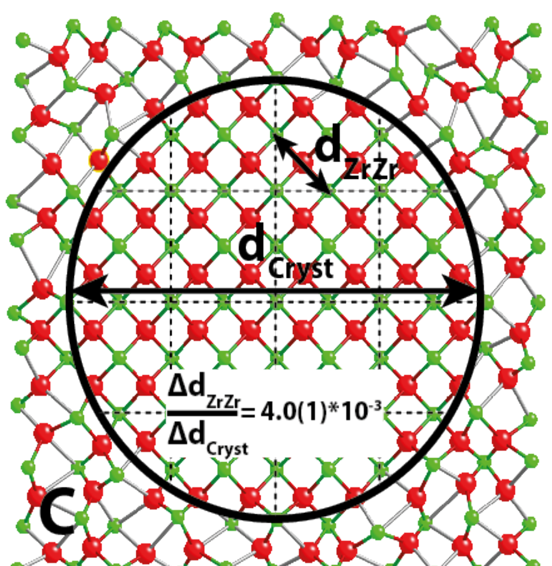


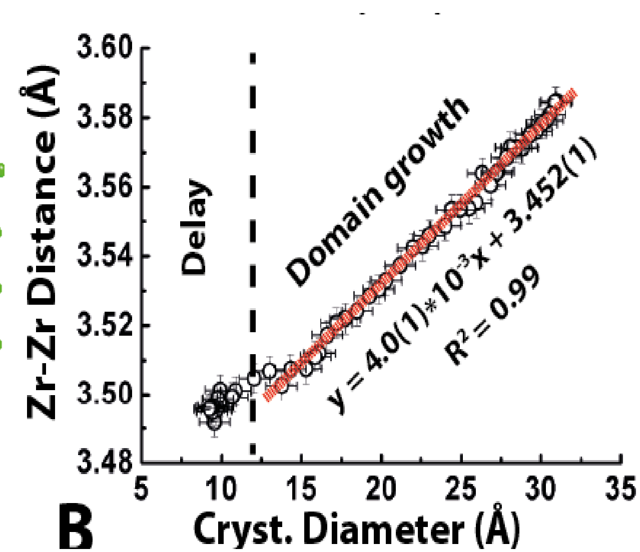
In-situ study of hydrothermal synthesis of yttria-stabilized zirconia nanoparticles



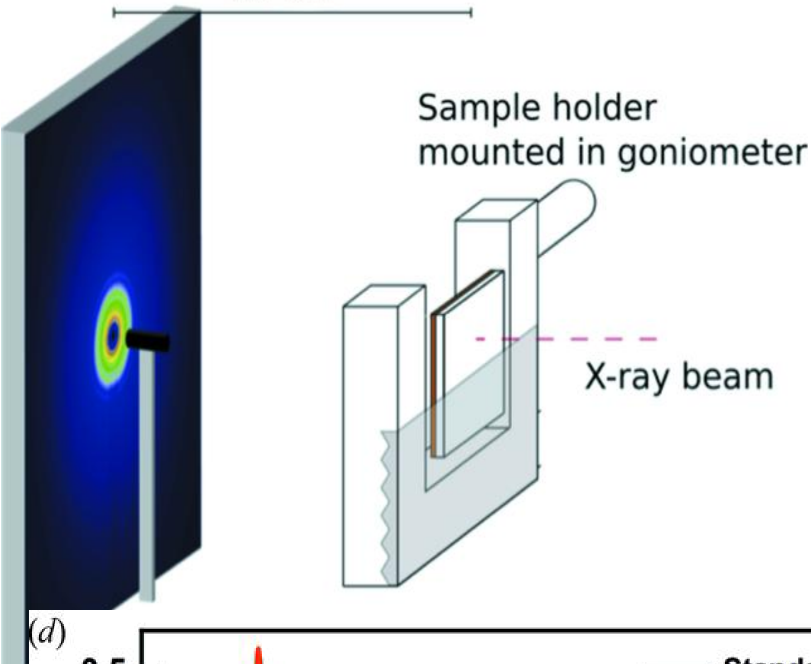
- Christoffer Tyrsted, Nina Lock, Kirsten M. \O. Jensen, Mogens Christensen, Espen D. Bøjesen, Hermann Emerich, Gavin Vaughan, Simon J. L. Billinge and Bo B. Iversen, , *IUCrJ.* **1**, 165-171 (2014)
- Nanocrystalline

final product

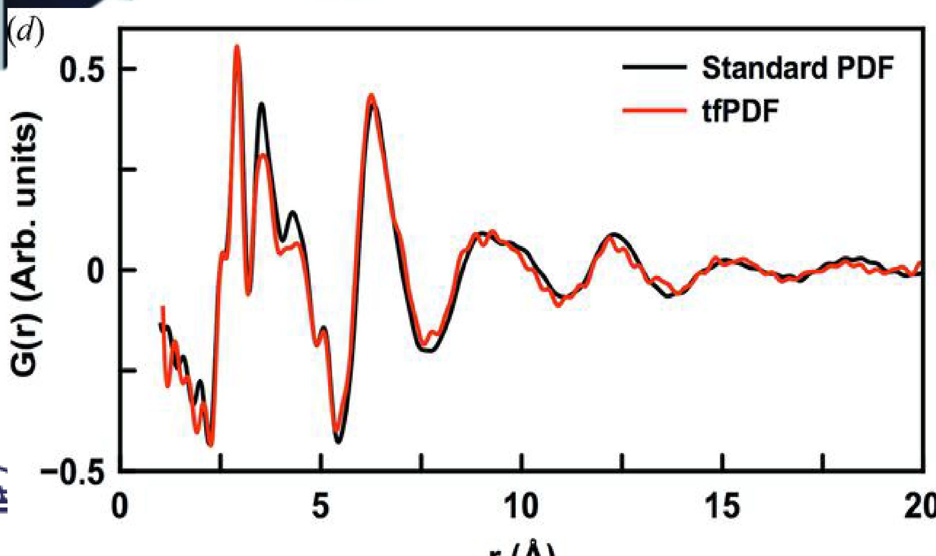


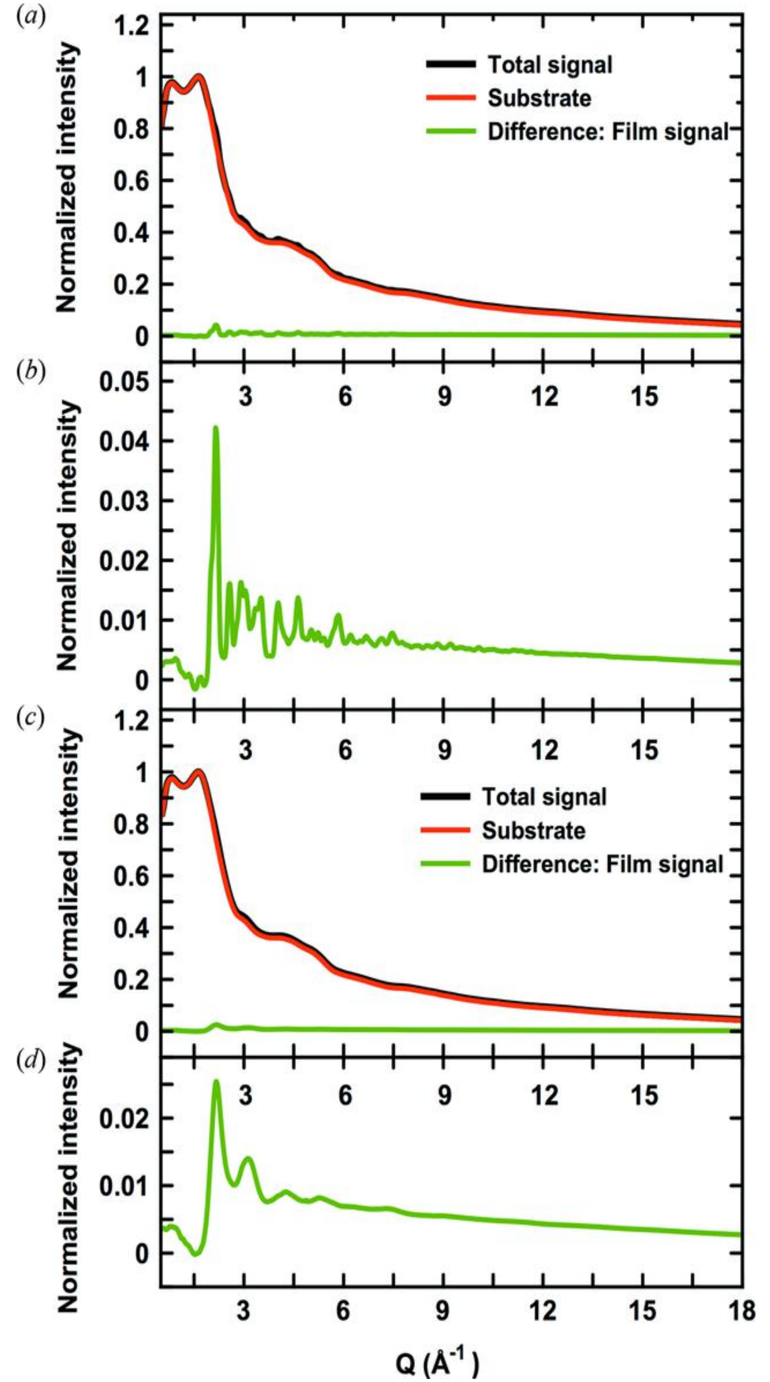


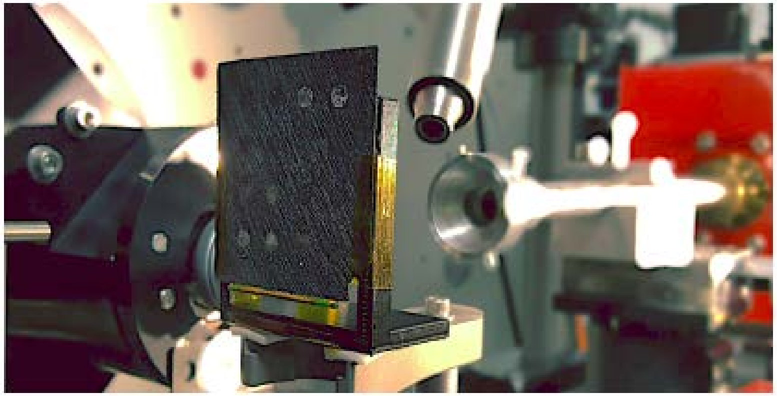
Thin film PDF at normal incidence



Jensen, Iversen, Johnson, Dooryhee SJLB et al. *IUCrJ* 2015

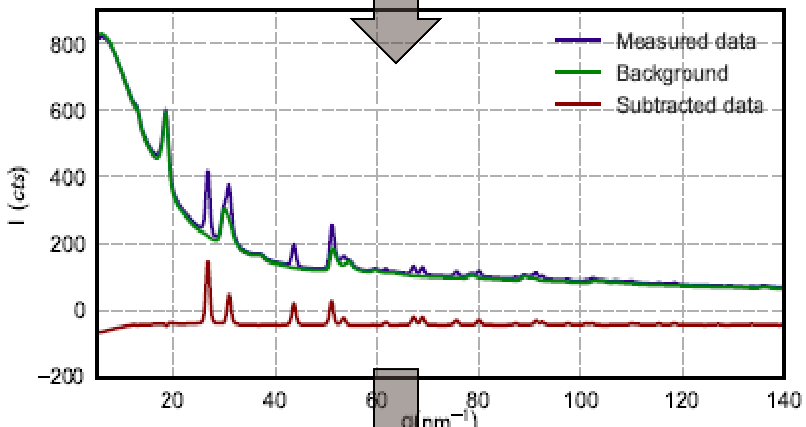


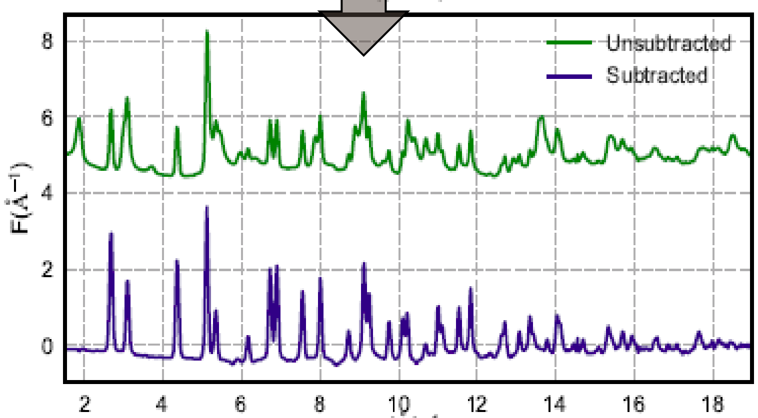


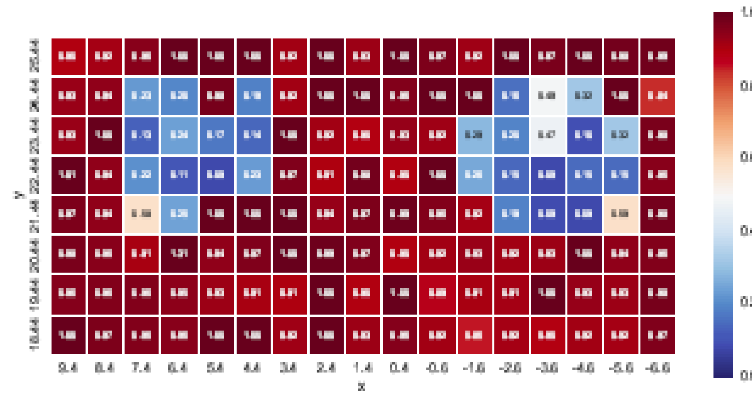


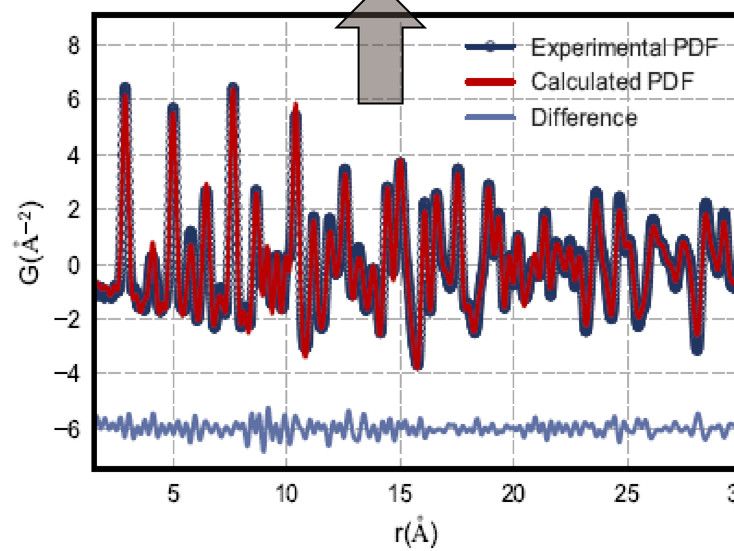
Spatially Resolved PDFs

- Anton Kovyakh, Soham Banerjee, Chia Hao Liu



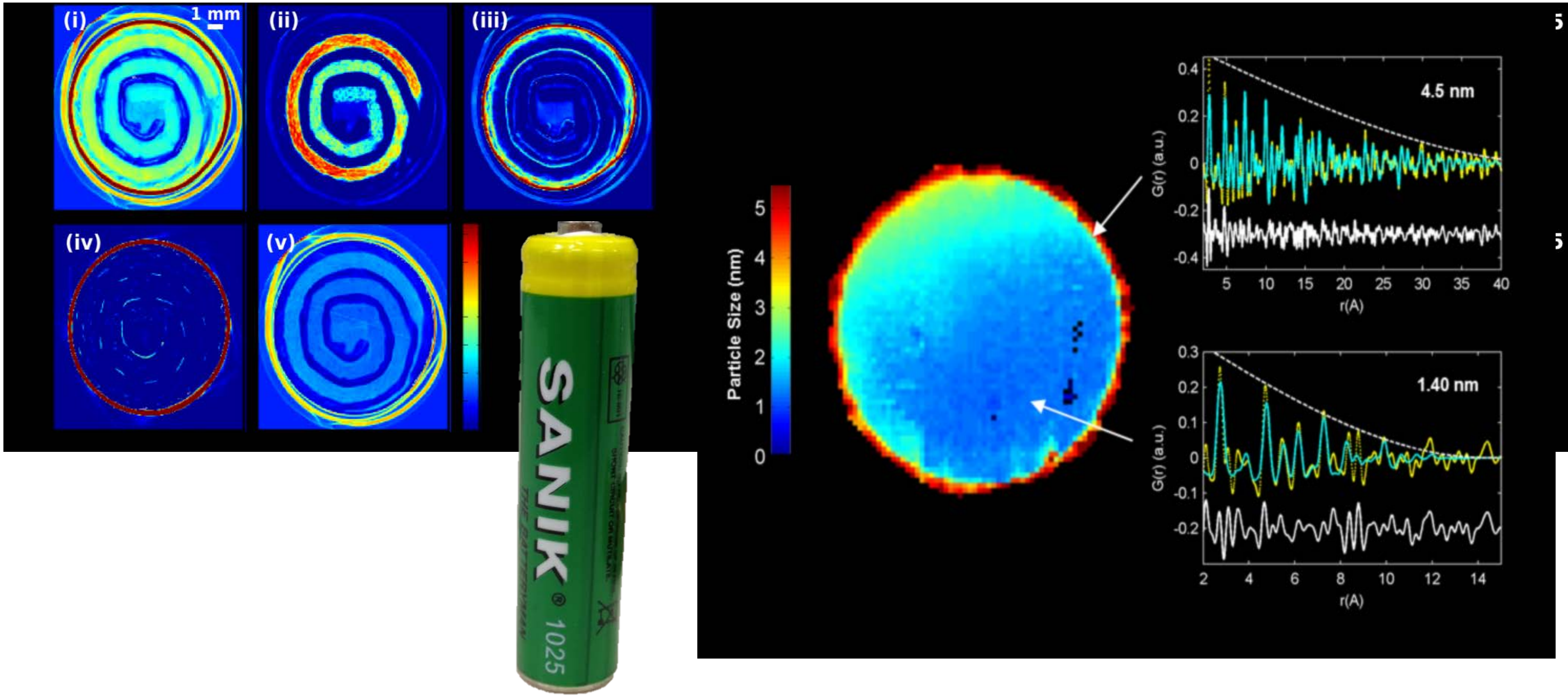






http://thebillingegroup.com

Batteries



10,000 2D datasets per image, 30 mins per image
~10Tb/day

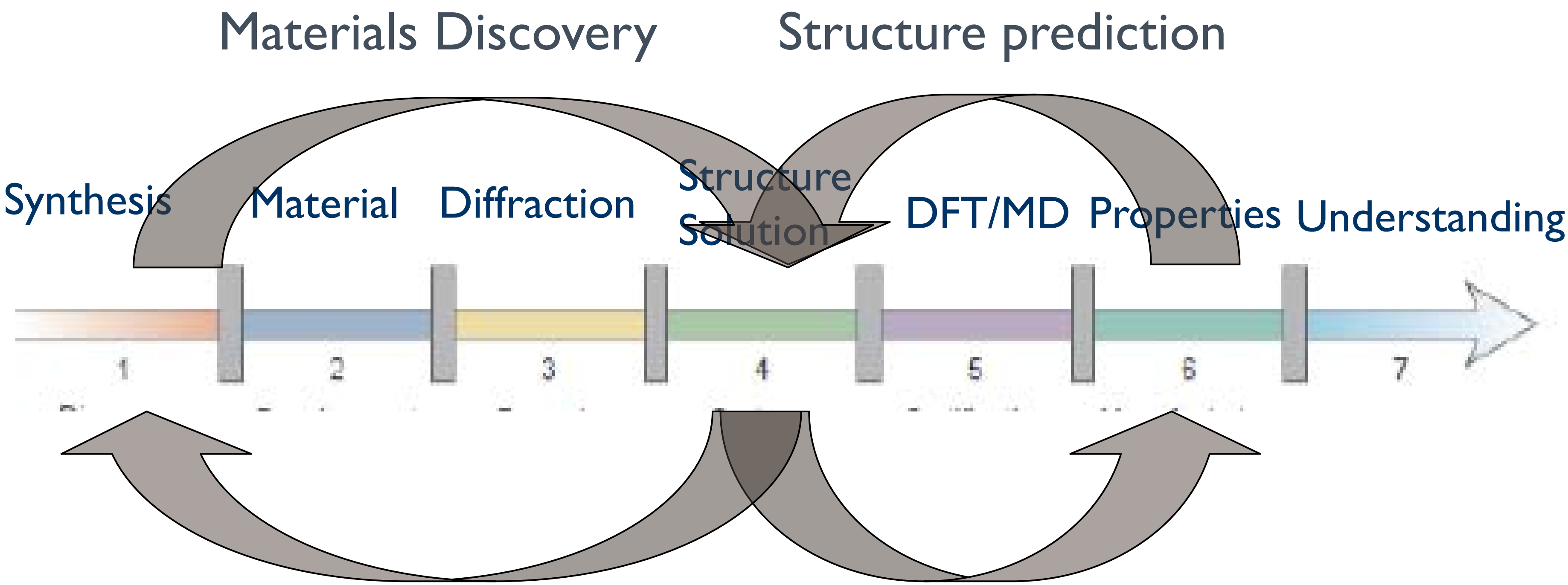
Jensen, Corr, Di Michiel, SJLB et al., *J. Electrochem. Soc*. (2015)

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

Materials Genomics



1. High throughput, automated data analysis
2. Sample/experiment metadata + data captured in databases
3. Link outcomes to inputs => autonomous experimentation

XPD@NSLS-II, BL 28ID-2

- Partner User Agreement between us and XPD and PDF beamlines at NSLS-II
- Methodological contributions to beamline development and commissioning:
  - Automated data acquisition protocols
  - Low-T cryostat
  - Background reduction
- Software for XPD
  - xpdAcq
  - SHED
  - xpdAn

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

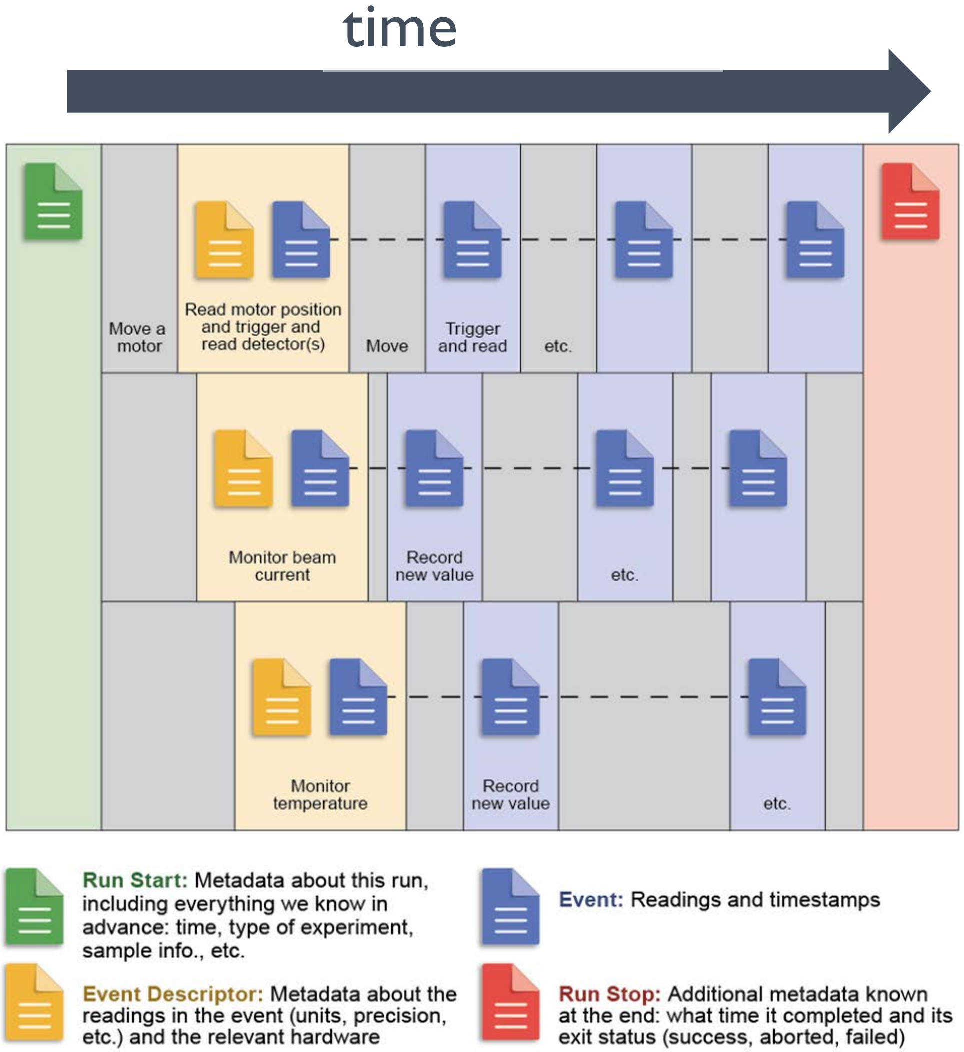
BROO
NATIONAL

Python stack for solving the high throughput

- Bluesky
  - The event model
  - NSLS-II software team
- xpdAcq
  - Collect high-throughput data with large amount of metadata capture but low experimenter overhead
- SHED
  - "Streaming heterogeneous event data"
  - Allows complex, configurable data processing graphs to be built for highly heterogeneous data streams
- xpdAn
  - UI for the stack

Bluesky and the Event Model

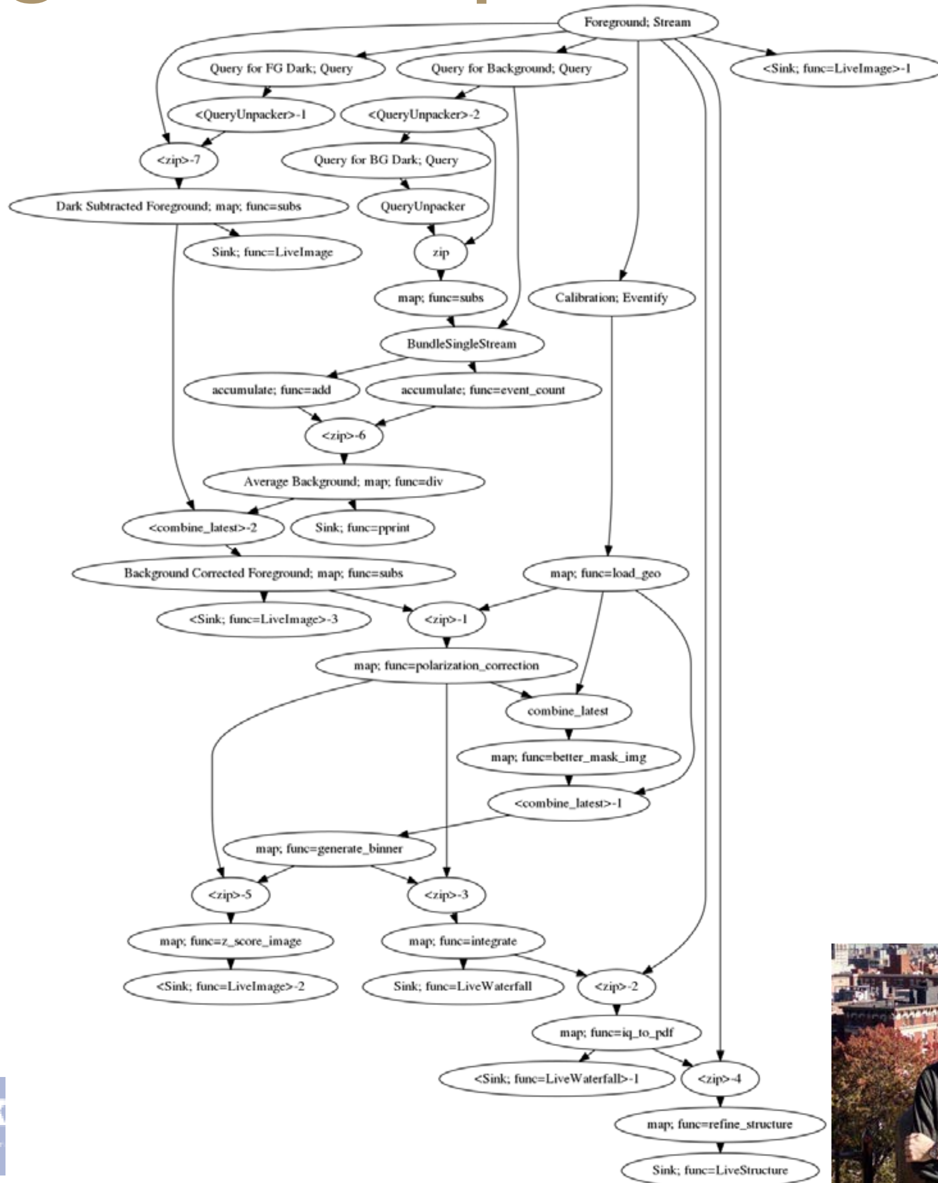
- Life is a series of fortunate (quite probably asynchronous) events
  - Asynchronous Data Acquisition
  - Rich, Searchable Metadata
  - Hardware Agnostic
- Event data is Heterogeneous



Modified from https://nsls-ii.github.io/architecture-overview.html

SHED: Heterogeneous Pipelines

- Most workflow tools don’t work with Event Model
- Nodes are transformations
- Build graph first, then stream data through it
- Send data to live visualization, database, etc



Columbia University in the City of New York

Brookhaven National Laboratory

SHED: Technical bits

- Base library "dask-streams" handles homogeneous data
- In this case each stream node know how to:
  - Update themselves with new data (performing an operation on the data)
  - Emit processed data to other nodes which are subscribed to it
  - Provide lazy data processing and backpressure (we won't memory crash!)
- The SHED library extends this to work with the Event Model
- Each node knows how to dispatch information to different methods within the class depending on what type of document we are looking at
- The node then passes data downstream in the Event Model.
- Since each node "speaks" the event model any node can
  - Attach directly to the running experiment
  - Attach to database
  - Attach to other code which supports the event model, including existing tools
    - Live visualization
    - Live computation (potentially on HPC cluster)
    - ZeroMQ integration

SHED: Example Pipeline

Purpose:
import libraries

define a data processing function

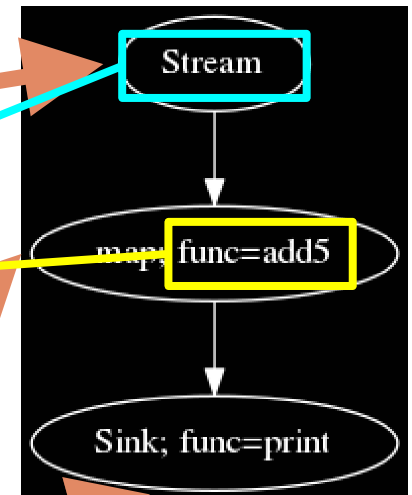
source where we will put data

node which maps onto the data
function to apply to data
Source of data
data key -> kwarg
describe the data from the func

print the data after processing

push data into the pipeline
could be attached to async
source (eg running experiment)

```
from streams import Stream
import shed.event_streams as es

def add5(img):
    return img + 5

source = Stream()

m = es.map(
    add5,
    source,
    input_info={'img': 'pe1_image'},
    output_info=[('img_plus_5'),
                    {'dtype': 'array',
                'source': 'pipeline'}])

m.sink(print)

my_datastream = <some event stream>
for data in my_datastream:
    source.emit(data)
```



xpdAn: A library of tubes

- *Complete user facing pipelines*
  - For common data systems
- *Smaller pipeline chunks*
  - Mix and match approach to pipeline creation
- *Advanced tools for data analysis*
  - Functions which run inside pipelines and standalone
    - Automated masking
    - Automated integration resolution
    - ***Machine learning and statistical analysis***

A day in the life of an Experiment

1. Collect setup scans and data
   1. Calibrations
   2. Dark frames
   3. Background frames
   4. Etc.
2. Collect production data
   1. Actual sample, could be temperature dependence, etc etc.
3. Save and collect relevant metadata
   1. What is the sample made of? Who owns the sample, who made it? When was it measured? What temperature was it measured at? Is the structure known?
   2. Link setup scans to production scans and apply corrections etc.
4. Stream to partially or fully corrected data and visualize and save
5. Restream with different analysis parameters if necessary

A Day in the Life of an Experiment

Collect rich metadata while keeping the experiment easy

```
bt.list()
```

```
ScanPlans:
0: 'ct_5'
1: 'ct_30'
2: 'ct_1'
3: 'ct_10'
4: 'ct_60'
5: 'ct_0.1'
6: 'Tramp'
Samples:
0: Ni
1:TiO2
2: GaAs
3: In0.5Ga0.5As
4: In0.25Ga0.75As
```

A Day in the Life of an Experiment

Running scans is then very easy, but rich metadata is saved

```
xrun(1, 6)
```

```
('start', {'beamline_id': 'xpd',
  'bkgd_sample_name': 'kapton_1mmOD',
  'bt_experimenters': ['Soham', 'Banerjee',],
  'bt_piLast': 'Billinge',
  'bt_wavelength': 0.1847,
  'sample_name': 'EuTiO3',
  #... (and much more)
  'sp_Nsteps': 50,
  'sp_computed_Tstep': 5.0,
  'sp_computed_exposure': 60.0,
  'sp_endingT': 330,
  'sp_num_frames': 600.0,
  'sp_plan_name': 'Tramp',
  'sp_requested_Tstep': 5,
  'sp_startingT': 85,
  'sp_time_per_frame': 0.1,
  'sp_type': 'Tramp'})
```

A Day in the Life of an Experiment

The output from the scan is an event stream.

It has a descriptor

```
xrun(1, 6)
```

```
('descriptor',
 {'data_keys':
     {'pe1_image':
         {'dtype': 'array',
          'object_name': 'pe1',
          'shape': [2048, 2048, 0],},
      'temperature':
         {'dtype': 'number',
          'object_name': 'cs700',
          'precision': 2,
          'units': 'degK'},
#… (and much more)
      'time': 1490578113.9974933,
      'uid': 'eaf45c6c-2e82-43de-b75d-
1c3e00beb8a5'})
```

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

A Day in the Life of an Experiment

…and a series of events

```
xrun(1, 6)
```

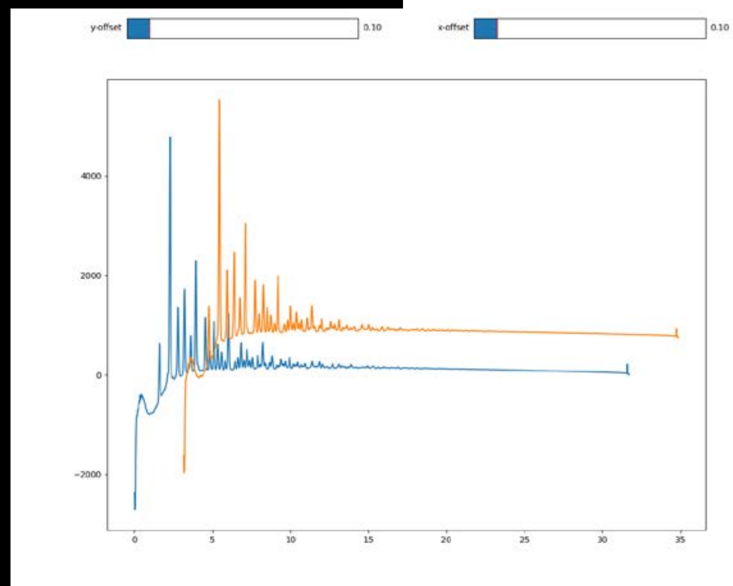

```
('event',
 {'data':
{'pe1_image':
array([[ 4848.03857422,  4852.92333984,
4862.05517578, ..., 4816.79980469,
4830.50830078,  4854.02685547]],
dtype=float32),
        'temperature': 85.8,
        'temperature_setpoint': 85.0},
 #... (and much more)
  'seq_num': 1,
  'time': 1490578114.0032613,
  'uid':
'b5fc4caa-0409-41d2-9aba-ee0bb30b8db9'})
```


COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

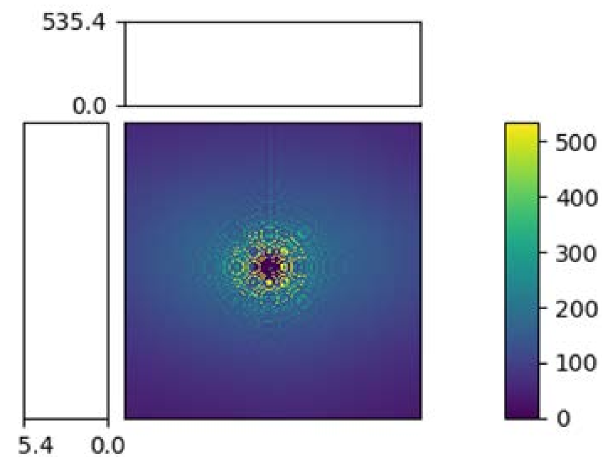
A Day in the Life of an Experiment

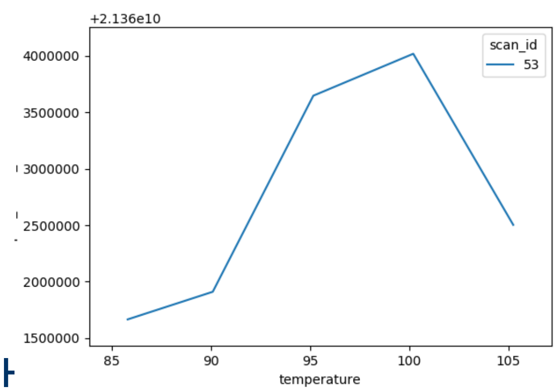
Depending how we set up the analysis graph, we can output live graphics at different points by subscribing to the stream

```
xrun(1, 6)
```



Output:
Live Visualization kicks in





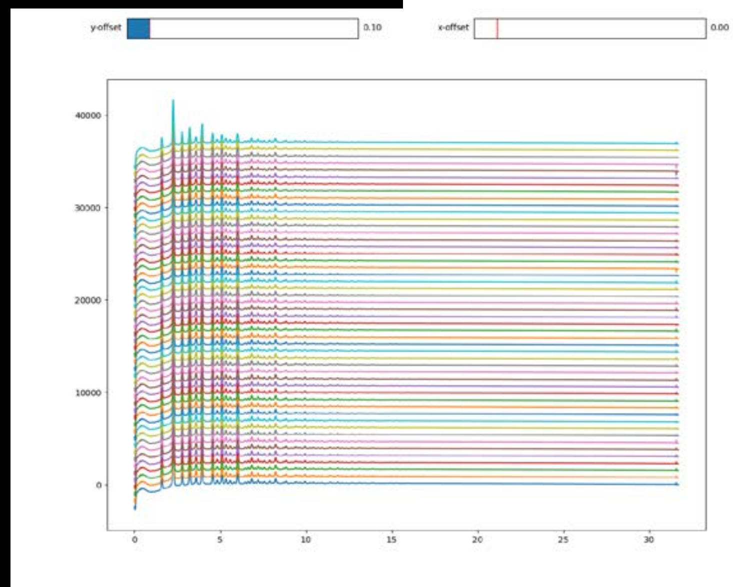
Columbia University in the City of New York
Brookhaven National Laboratory

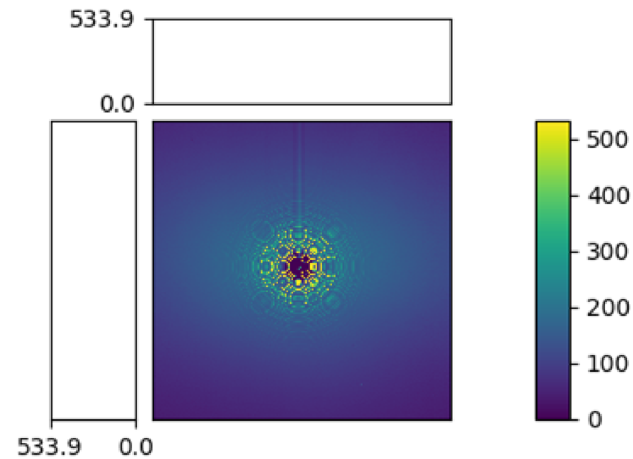
A Day in the Life of an Experiment

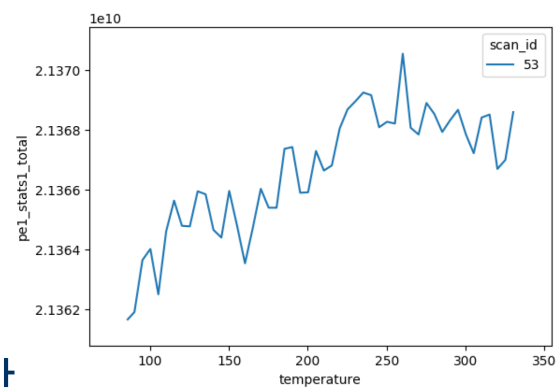
And the graphics can update as the stream passes through

```
xrun(1, 6)
```

Output:
Live Visualization continues







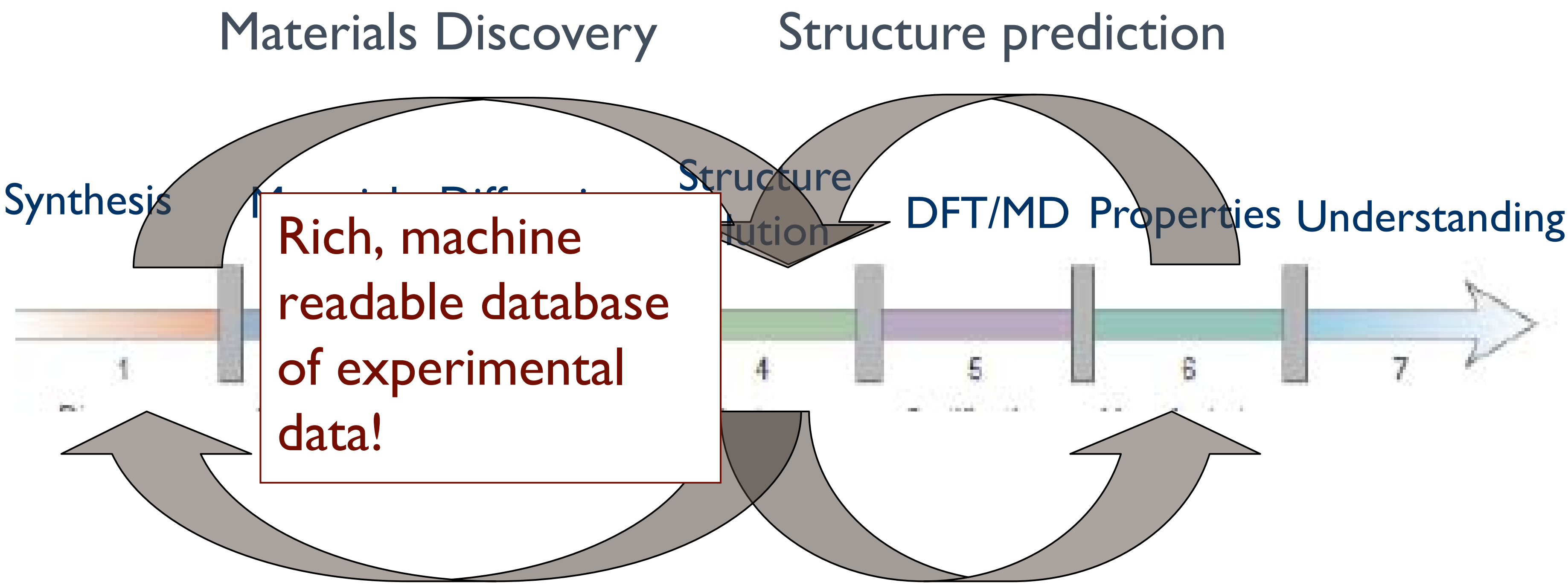
A Day in the Life of an Experiment

Scan is ended, so we find a Stop statement which terminates the stream

```
xrun(1, 6)
```

('stop',
 {'exit_status': 'success',
  'time': 1490583897.7904909,
  'uid': '0eec9e25-56f2-43d2-a09f-aac65c789ab7'})

Summary


Materials Discovery
Structure prediction
Synthesis
Structure
DFT/MD
Properties
Understanding
1
4
5
6
7
Rich, machine readable database of experimental data!

Acknowledgements

- A special thank you to all my current and former students and post-docs
- NSLS-II and DASK development teams
- Also my many wonderful collaborators, mentioned during the talk
- Facilities:
  - APS, CHESS, NSLS (and people therein)
  - MLNSC, ISIS, SNS (and people therein)
- Funding: DOE-BES and NSF-DMR

HTTP://thebillingegroup.com