Recent and future developments in PDF-land

S.J.L. Billinge

Columbia University,

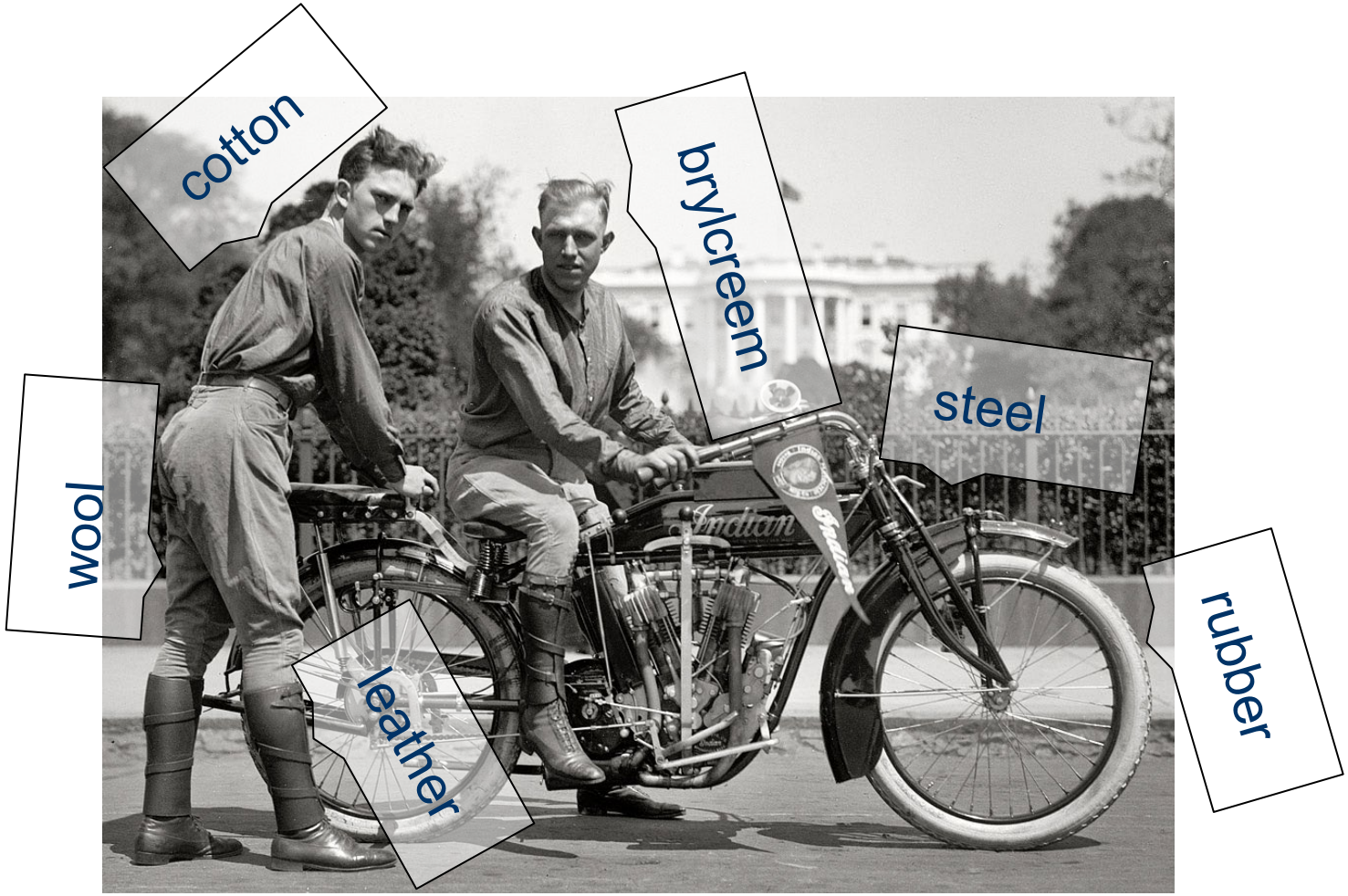
Brookhaven National Laboratory

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

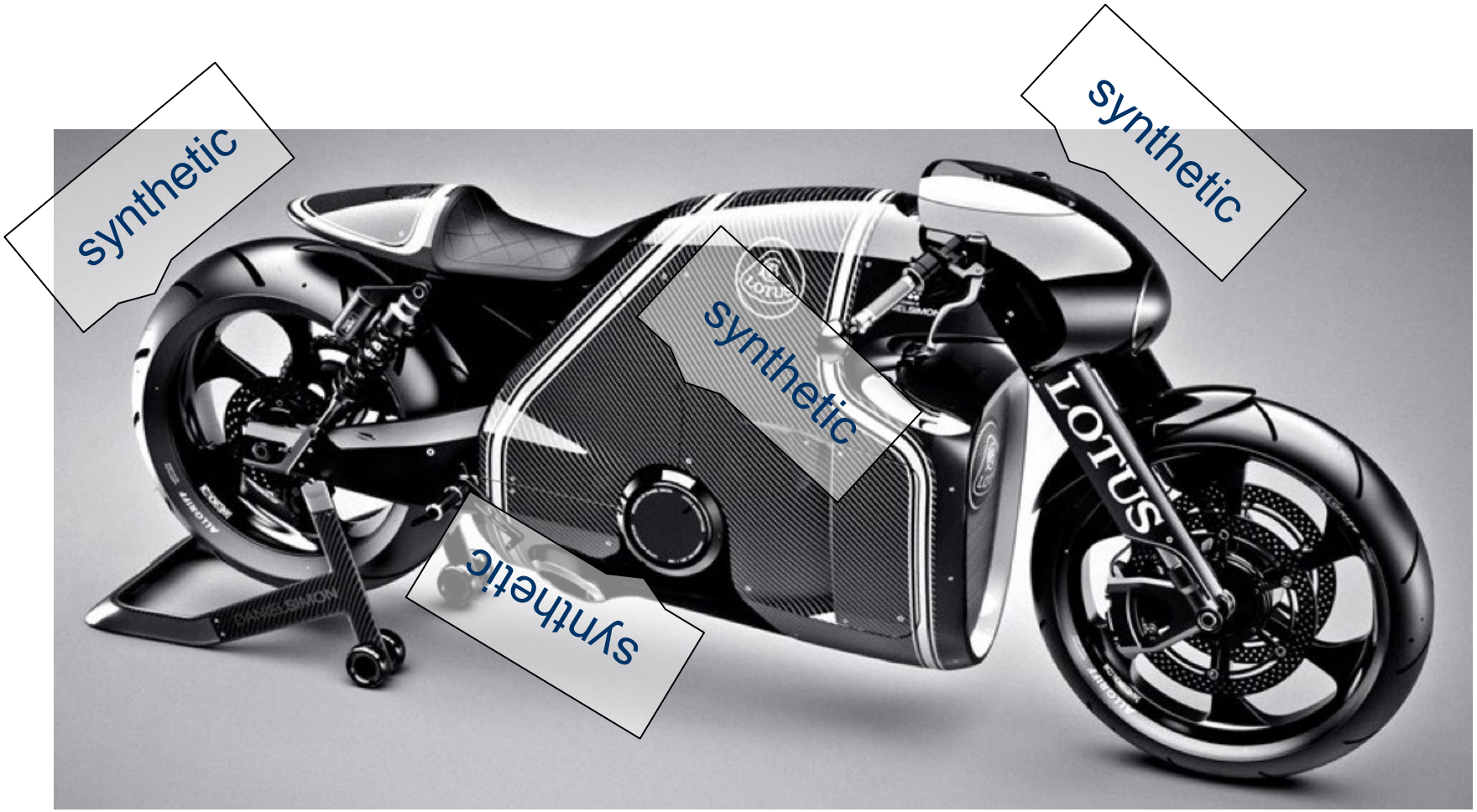
BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

Materials in 1915



Materials 2015



Modern materials and devices are complex!

- Photovoltaics with improved efficiency
  - Nanoparticles in the light collecting layer
- High energy density batteries
  - Electrodes
  - Electrolytes
- Fuel cells for transportation applications
  - Electrodes
  - Electrolytes
  - Catalysts
  - Hydrogen storage
- Sequestration
  - Functionalized mesoporous materials

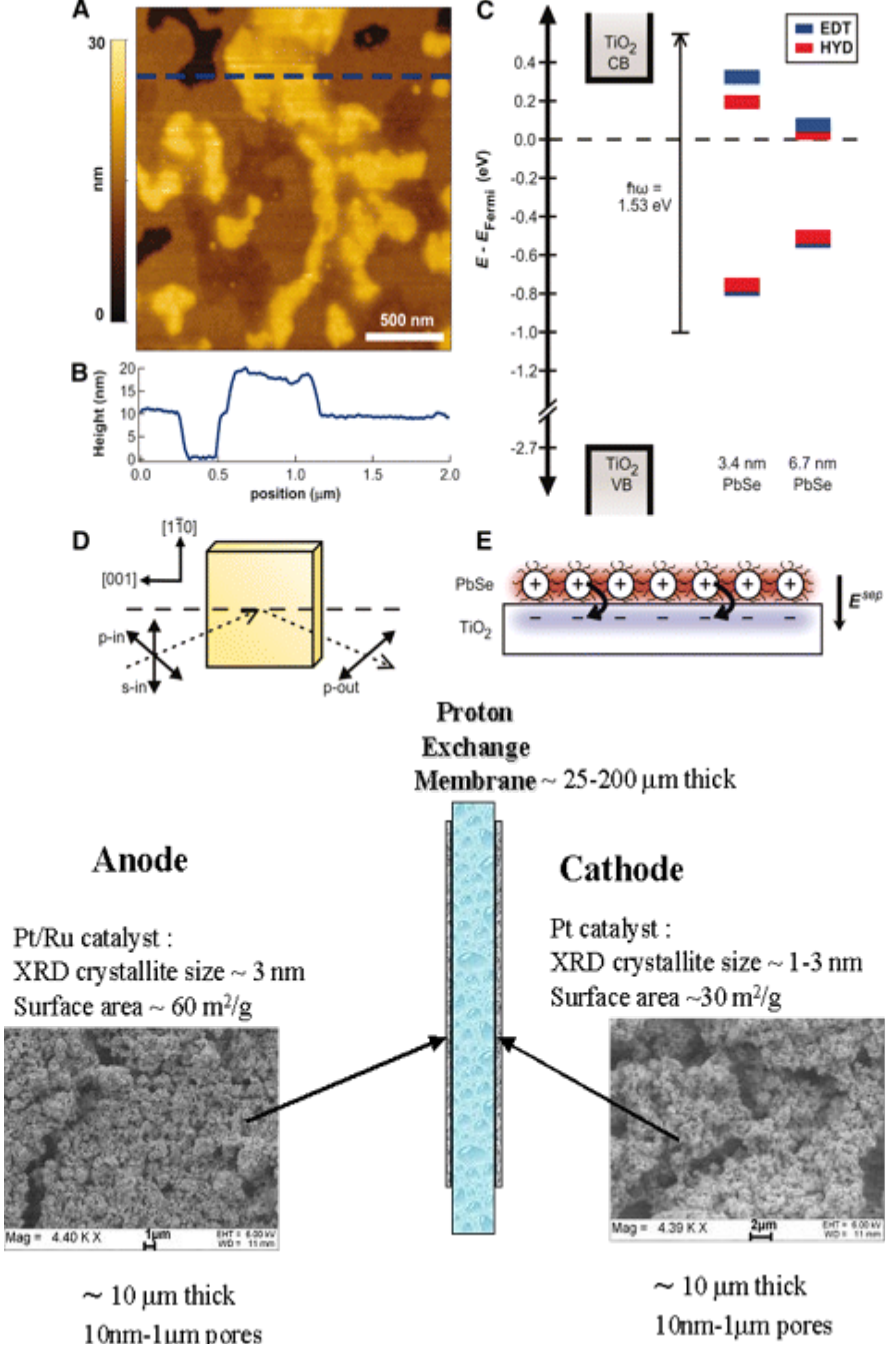


Image credits: 10.1126/science.1185509 U. Uppsala

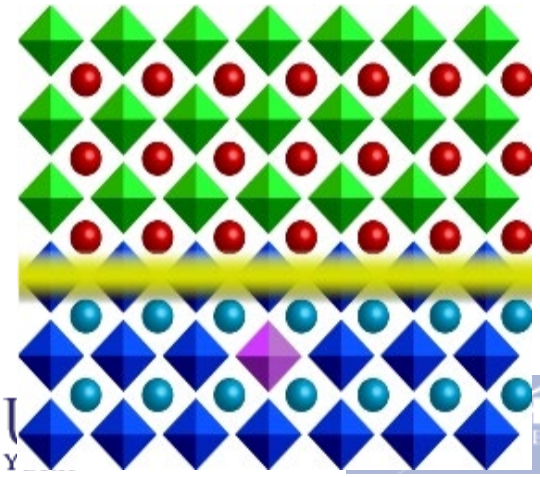
Real Materials

- 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



21st Century Crystallography

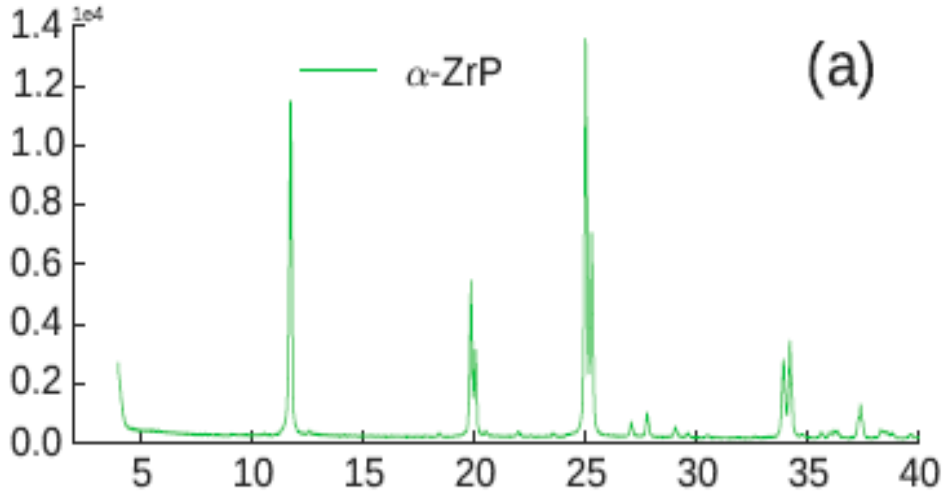
The "Crystallography" of Real-Materials

Has to go well beyond perfect crystals

Including:

- Defects
- Morphology
- Nanostructure
- Surfaces and interfaces
- Complexity!

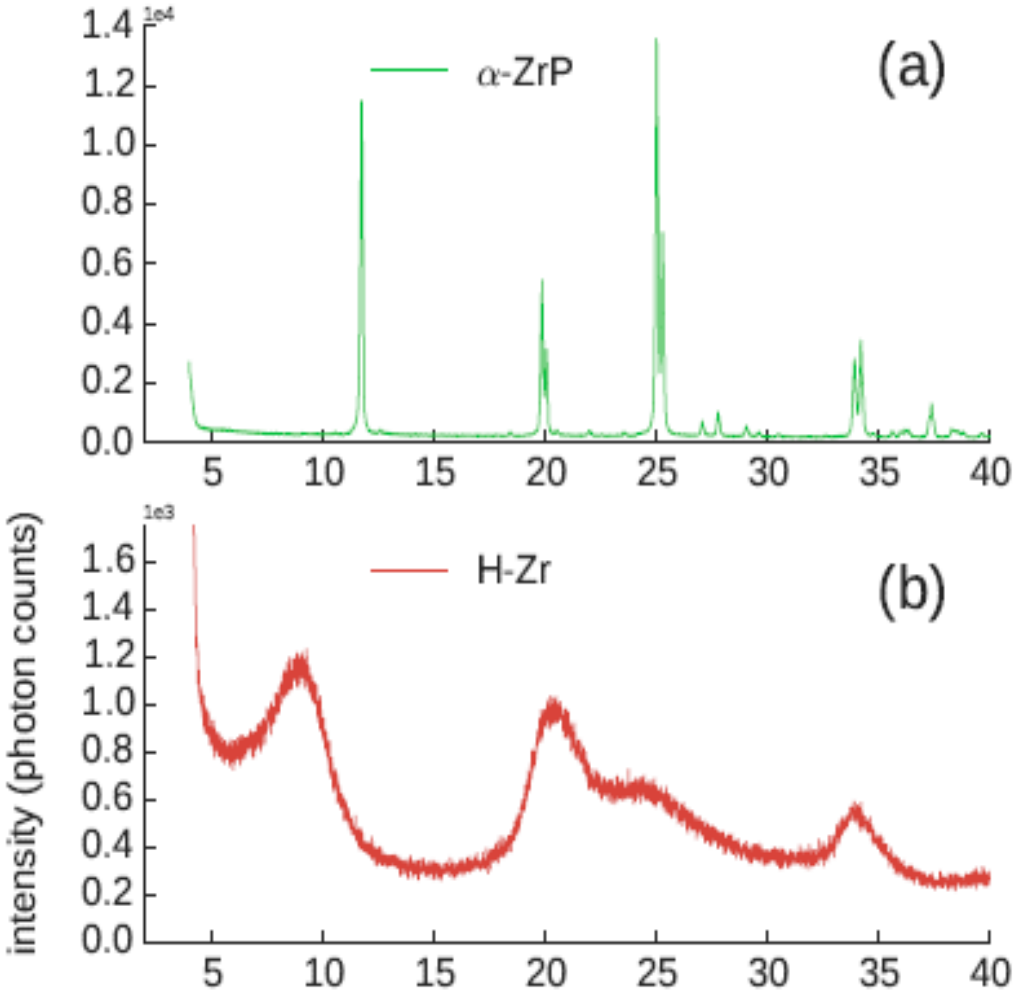
The Challenge with Real Material Structure Determination



XRPD

crystalline material

The Challenge with Real Material Structure Determination



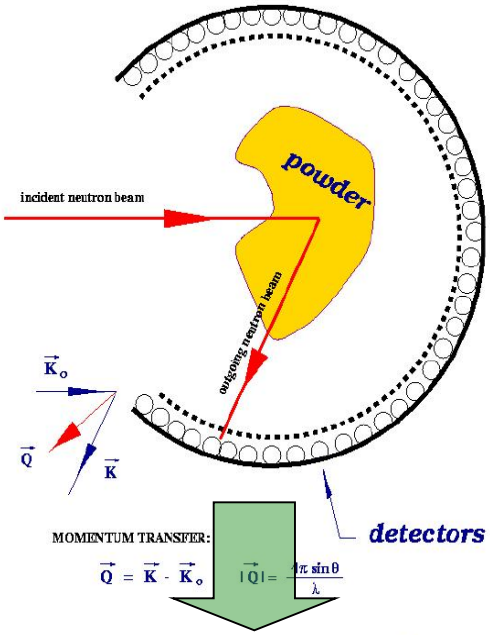
XRPD

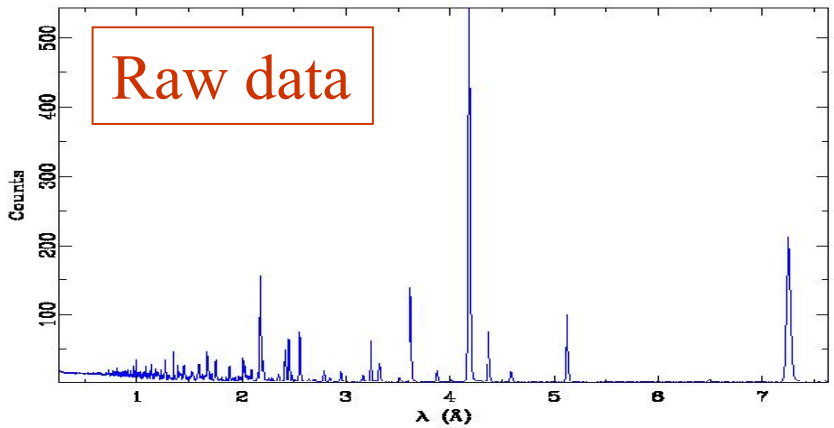
crystalline material

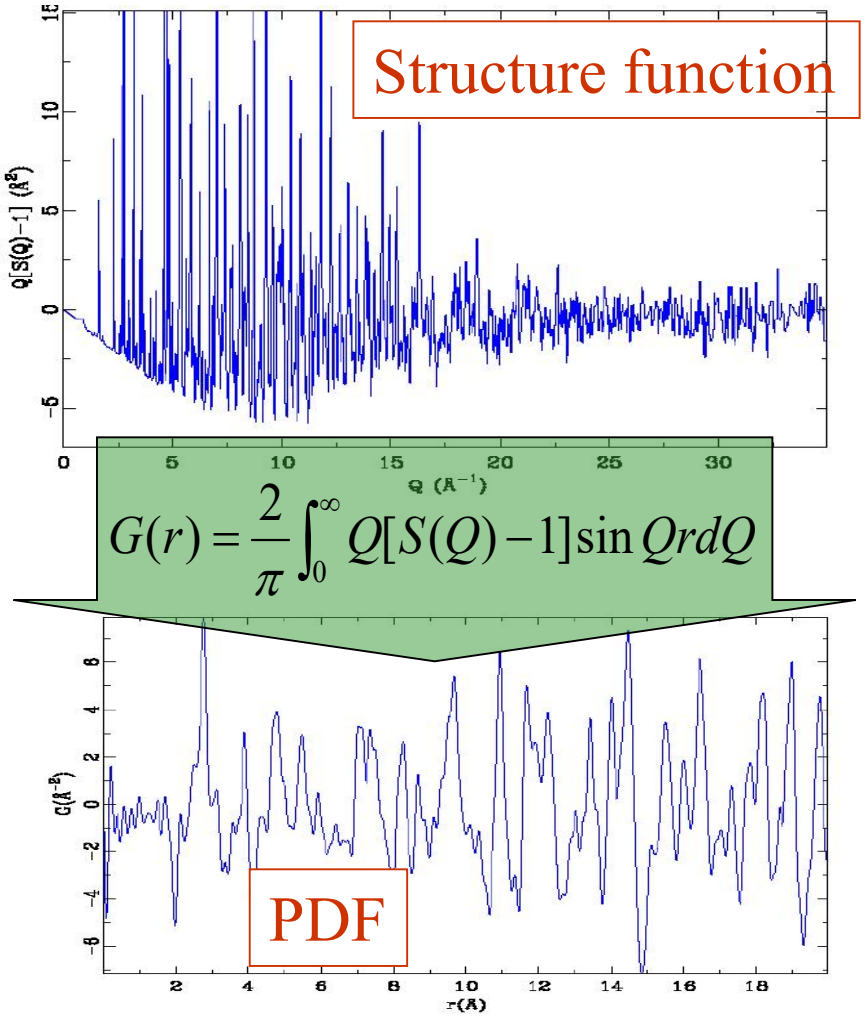
nanocrystalline material

HTTP://thebillingegroup.com

Local atomic structure from the Atomic Pair Distribution Function (PDF)





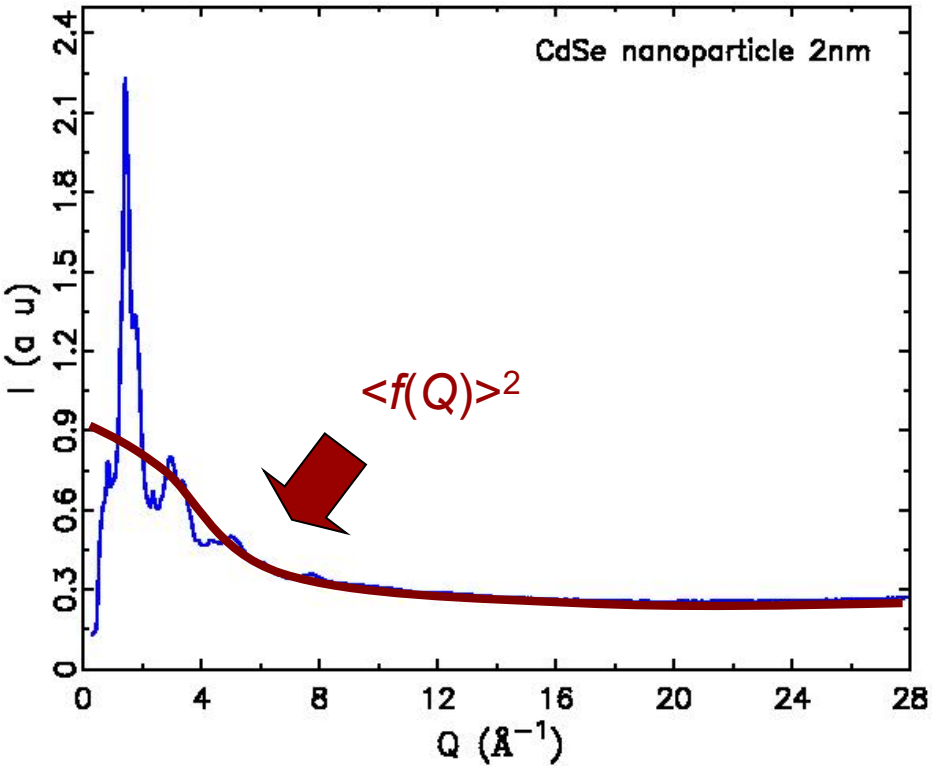


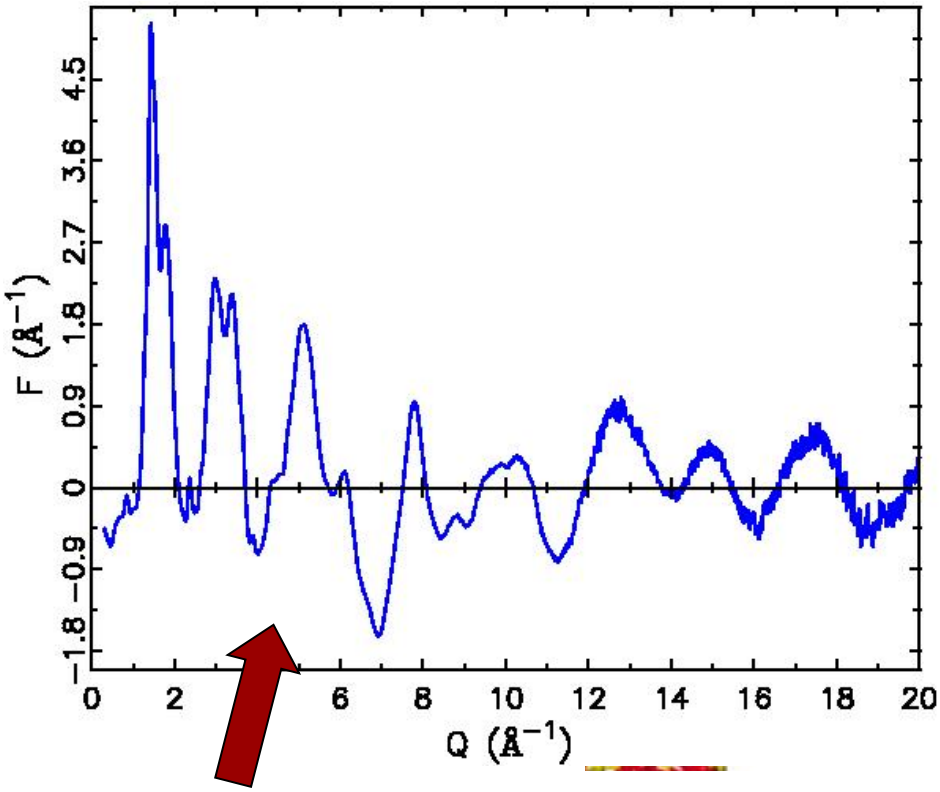
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

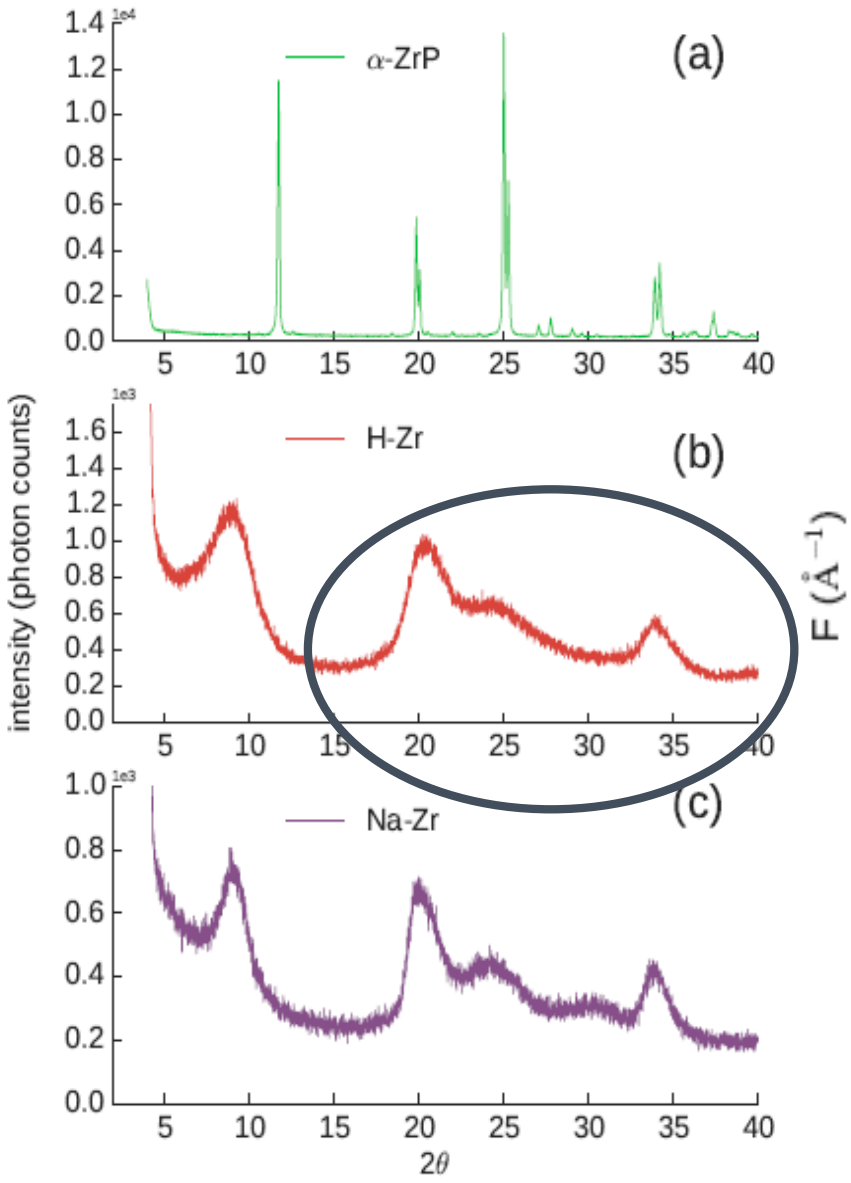
But there is no information at high-Q…?

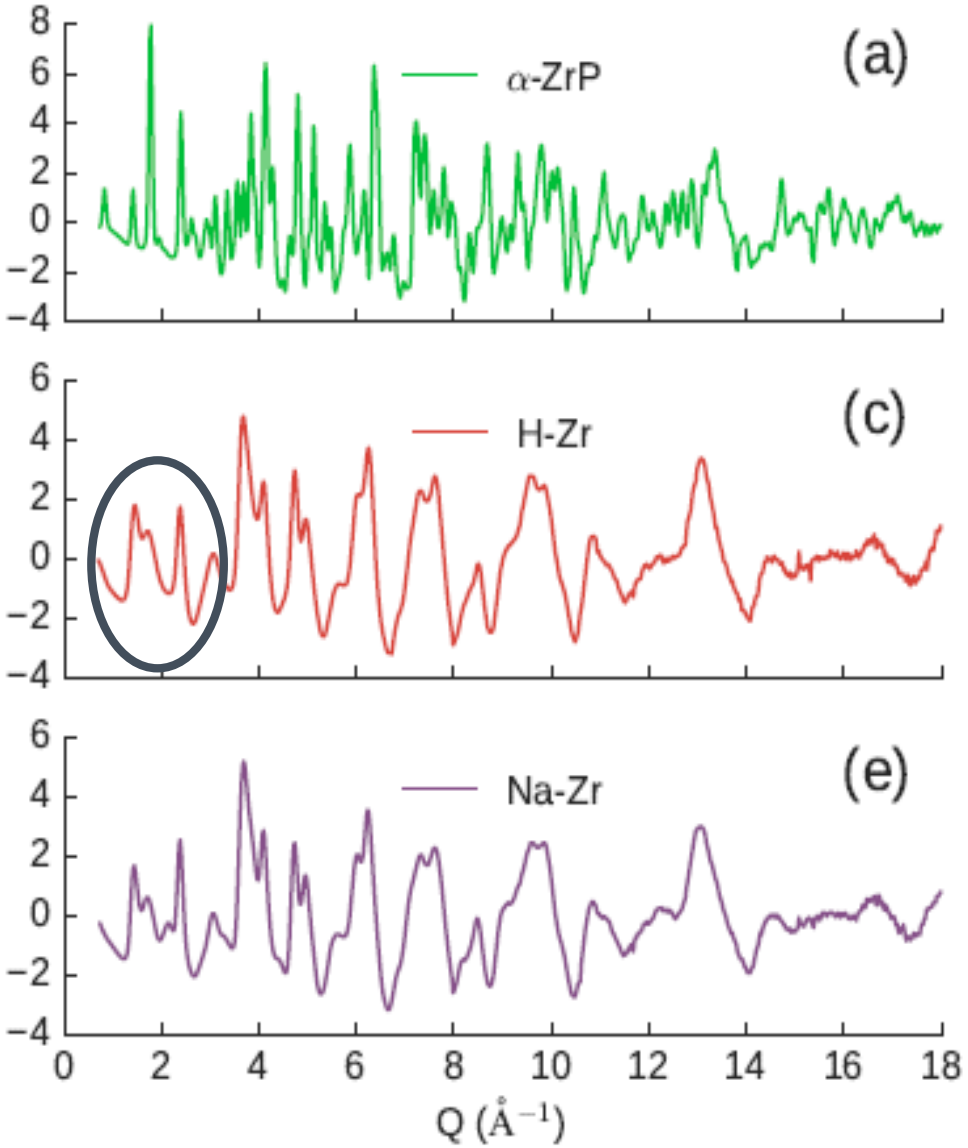




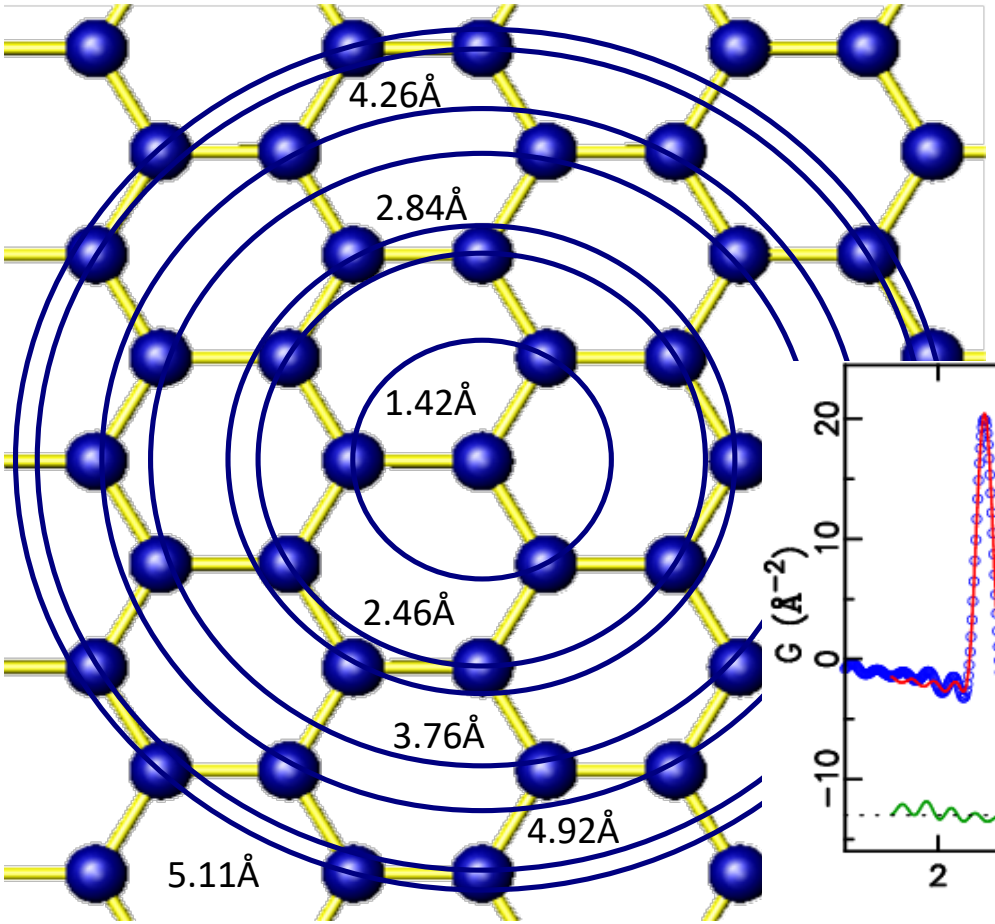
$$G(r) = \frac{2}{\pi}\int_0^\infty Q[S(Q)-1]\sin Qr\,\mathrm{d}Q$$

The Challenge

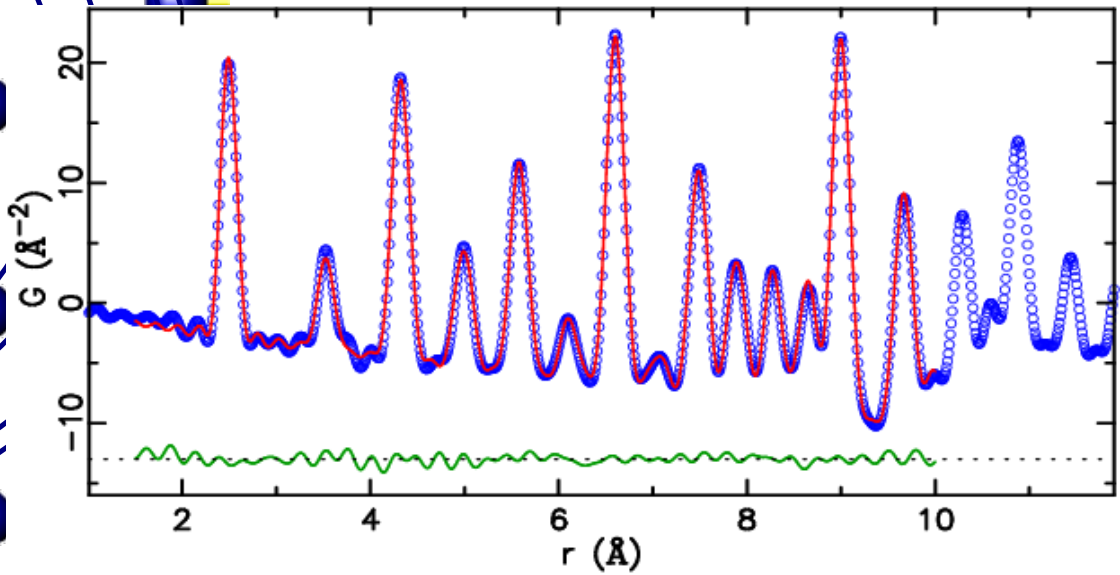




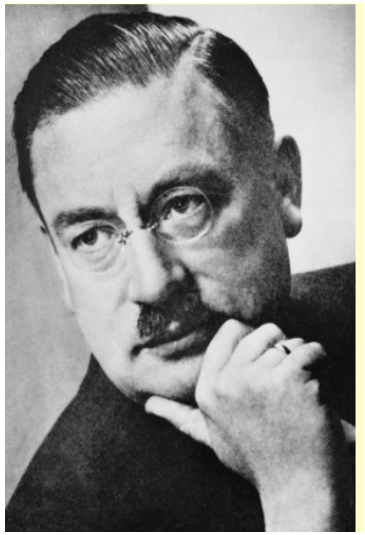
Nanostructure refinement



Pair distribution function (PDF) gives the probability of finding an atom at a distance “r” from a given atom.



PDF has been around for a long time



- Pieter Debye, 1915:

$$I = \sum_n \sum_m f_m f_n^* \frac{\sin qr_{mn}}{qr_{mn}}$$

- Fritz Zernike and Jon Prins, 1927:

$$4\pi r^2 \rho(r) = 4\pi r^2 \rho_a + \frac{2r}{\pi} \int_0^\infty qi(q) \sin qr dq$$

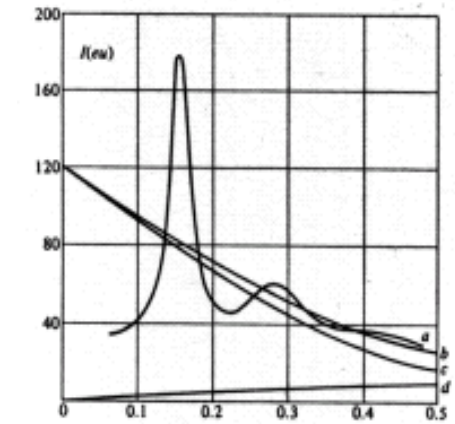
History

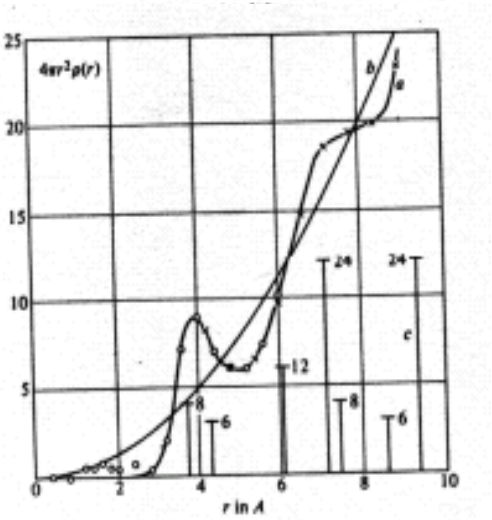
Debye and Menke, Z. Phys. (1930)

PDFs of mercury

Tarasov, L. P., and Warren, B. E., (1936) *J. Chem. Phys.*, **4**, 236.

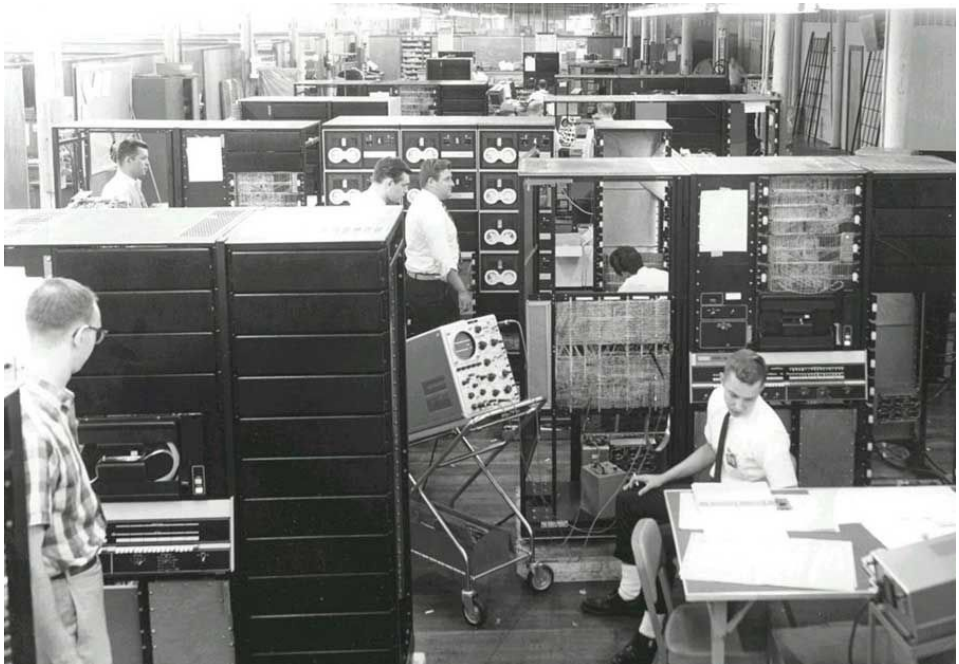
X-ray PDFs of molten sodium





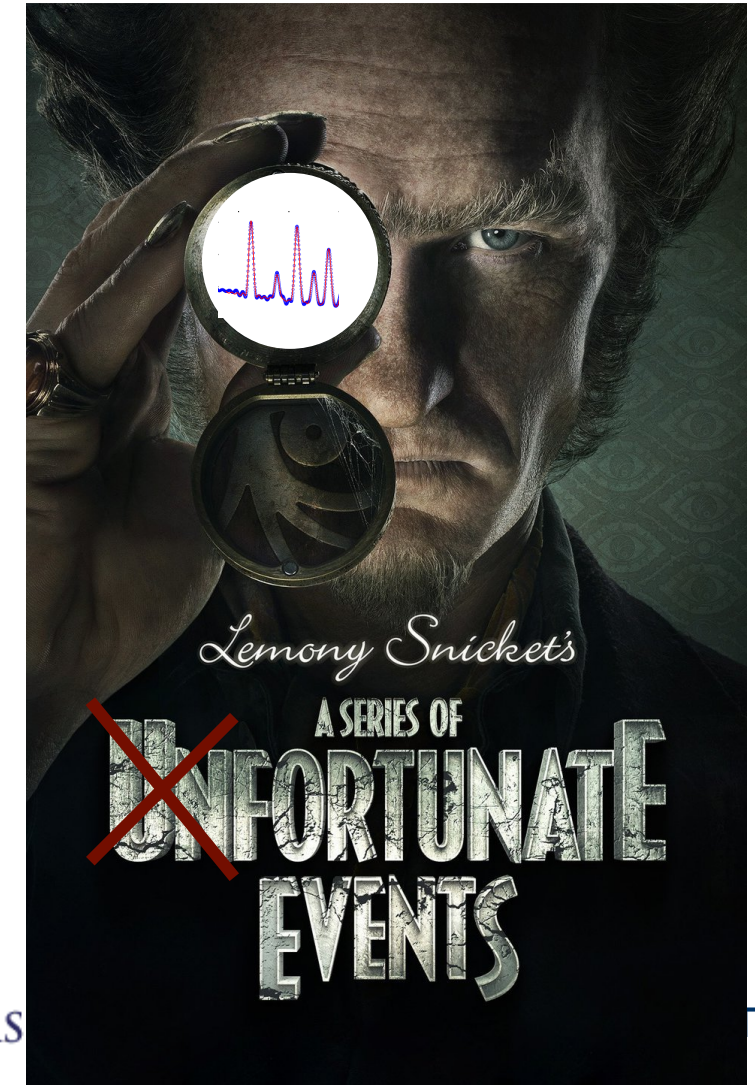
Studying (nano)crystalline materials with PDF

- 1950's Rosalind Franklin
  - Carbon black, disordered carbon materials
  - (also Bert Warren!)
- 1960's Roy Kaplow, Stephen Strong and Benjamin L. Averbach and computers
  - Simple metals, e.g., Lead. Interested in using them as models for liquid and glassy metals



PDP-9 factory test area, ~1967

PDF



A series of fortunate events

Web of science (x-ray + PDF) OR (x-ray + pair distribution function)

Total Publications
2,470

***h*-index**
96

Average citations per item
22.8

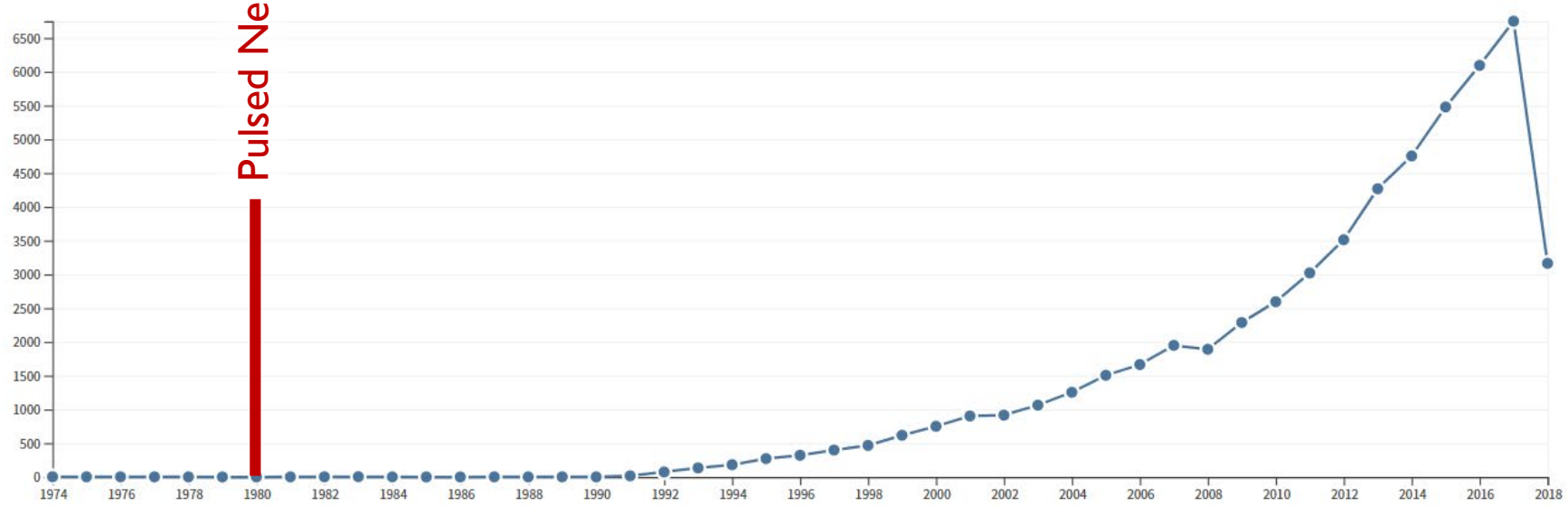
Sum of Times Cited
56,306

Without self citations
50,593

Citing articles
43,758

Without self citations
42,303

Sum of Times Cited per Year



Studying (nano)crystalline materials with PDF

- 1950's Rosalind Franklin
  - Carbon black, disordered carbon materials
- 1960's Roy Kaplow, Stephen Strong and Benjamin L. Averbach and computers
  - Simple metals, e.g., Lead. Interested in using them as models for liquid and glassy metals
- 1980's Takeshi Egami and pulsed neutrons
  - Use of the newly available pulsed neutron source IPNS for quasicrystals then high temperature superconductors when they were discovered

A series of fortunate events

Web of science (x-ray + PDF) OR (x-ray + pair distribution function)

Total Publications
2,470

***h*-index**
96

Average citations per item
22.8

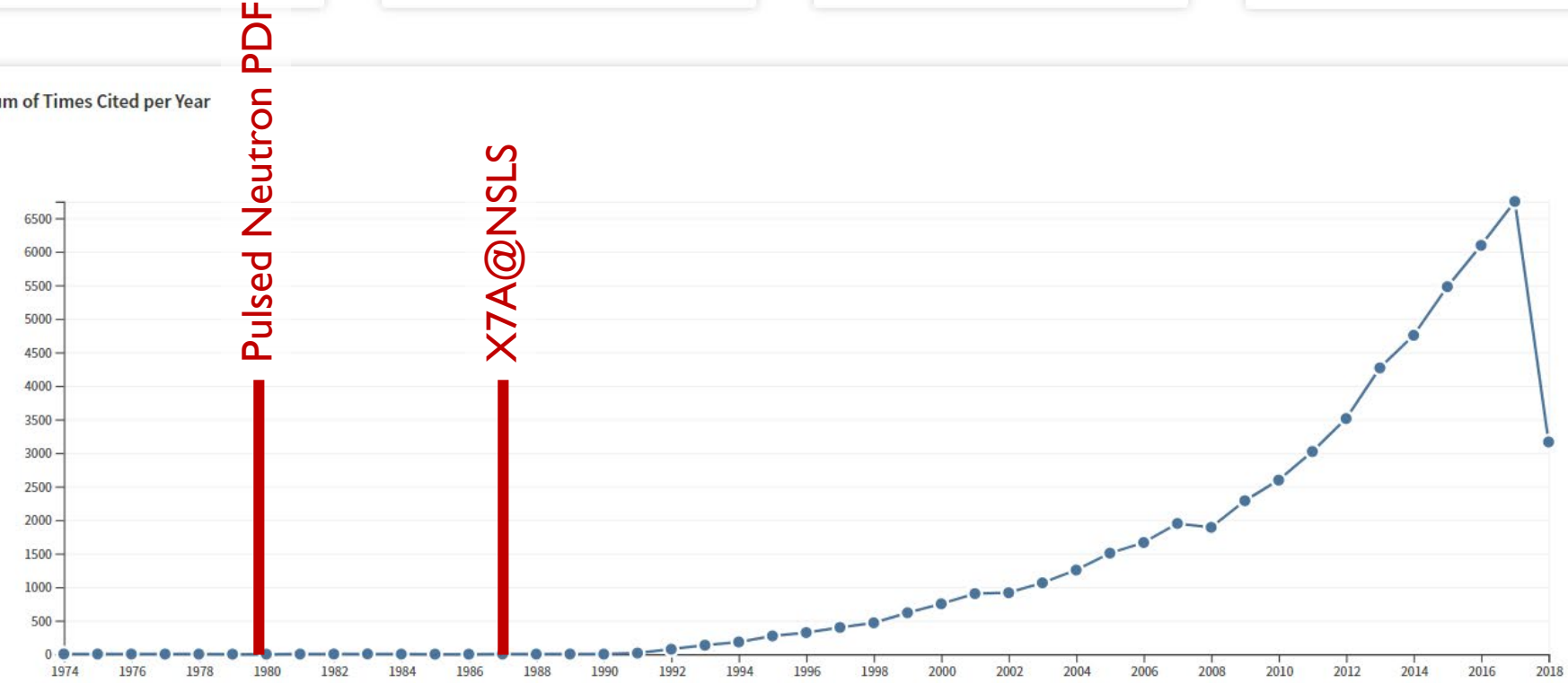
Sum of Times Cited
56,306

Without self citations
50,593

Citing articles
43,758

Without self citations
42,303

Sum of Times Cited per Year



Synchrotron PDF

- Early x-ray PDF measurements at X7A, NSLS-II (20 keV, Ge point detector), led by Takeshi Egami

A series of fortunate events

Web of science (x-ray + PDF) OR (x-ray + pair distribution function)

Total Publications
2,470

***h*-index**
96

Average citations per item
22.8

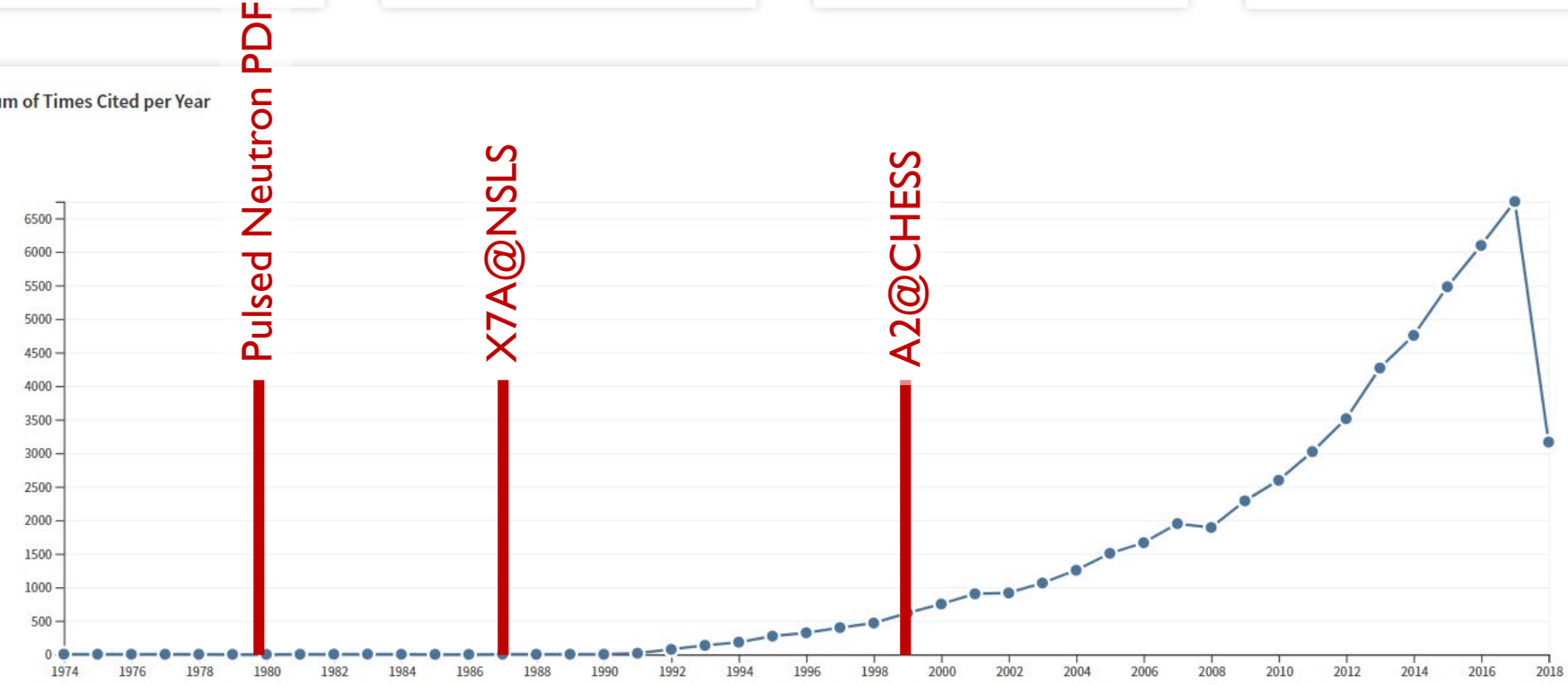
Sum of Times Cited
56,306

Without self citations
50,593

Citing articles
43,758

Without self citations
42,303

Sum of Times Cited per Year



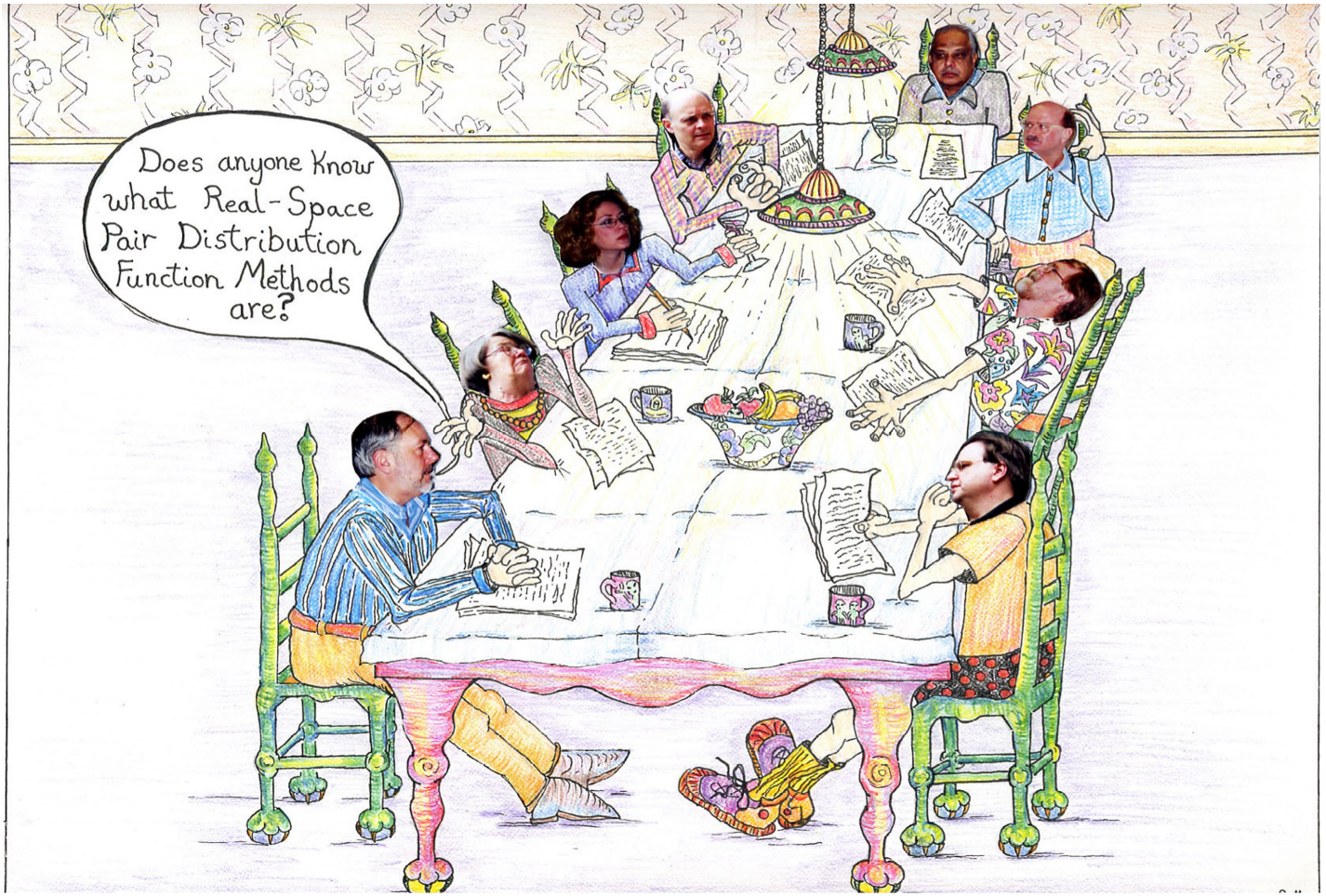
High energy synchrotron PDF

- Early x-ray PDF measurements at X7A, NSLS-II (20 keV, Ge point detector), led by Takeshi Egami
- First high resolution x-ray PDF measurement (Q_{max} = 45 A^{-1}!) using CHESS, Ge point detector, with Stefan Kycia

COLUMBIA U
IN THE CITY OF NEW YO

HTTP://thebillingegroup.com

Workshop at the 2001 ACA annual meeting



Cartoon by Stacey Morrison, shared with me by Connie Rajnak (Chidester)

A series of fortunate events

Web of science (x-ray + PDF) OR (x-ray + pair distribution function)

Total Publications
2,470

***h*-index**
96

Average citations per item
22.8

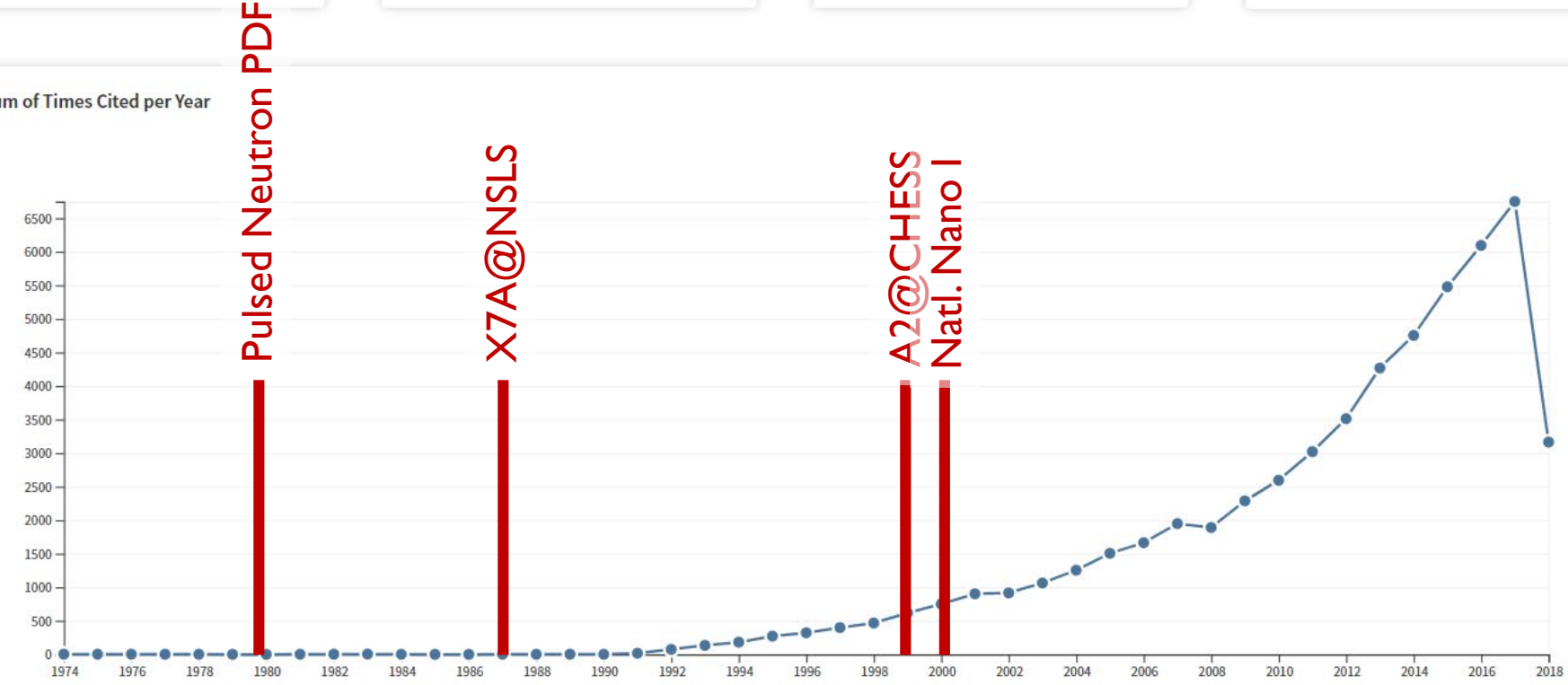
Sum of Times Cited
56,306

Without self citations
50,593

Citing articles
43,758

Without self citations
42,303

Sum of Times Cited per Year



Synchrotron PDF

- Early x-ray PDF measurements at X7A, NSLS-II (20 keV, Ge point detector), led by Takeshi Egami
- First high resolution x-ray PDF measurement (Q_{max} = 45 A^{-1}!) using CHESS, Ge point detector, with Stefan Kycia
- National Nanotechnology Initiative 2000

COLUMBIA
IN THE CITY OF NEW Y

PDF landscape in 2018

Web of science (x-ray + PDF) OR (x-ray + pair distribution function)

Total Publications
2,470

***h*-index**
96

Average citations per item
22.8

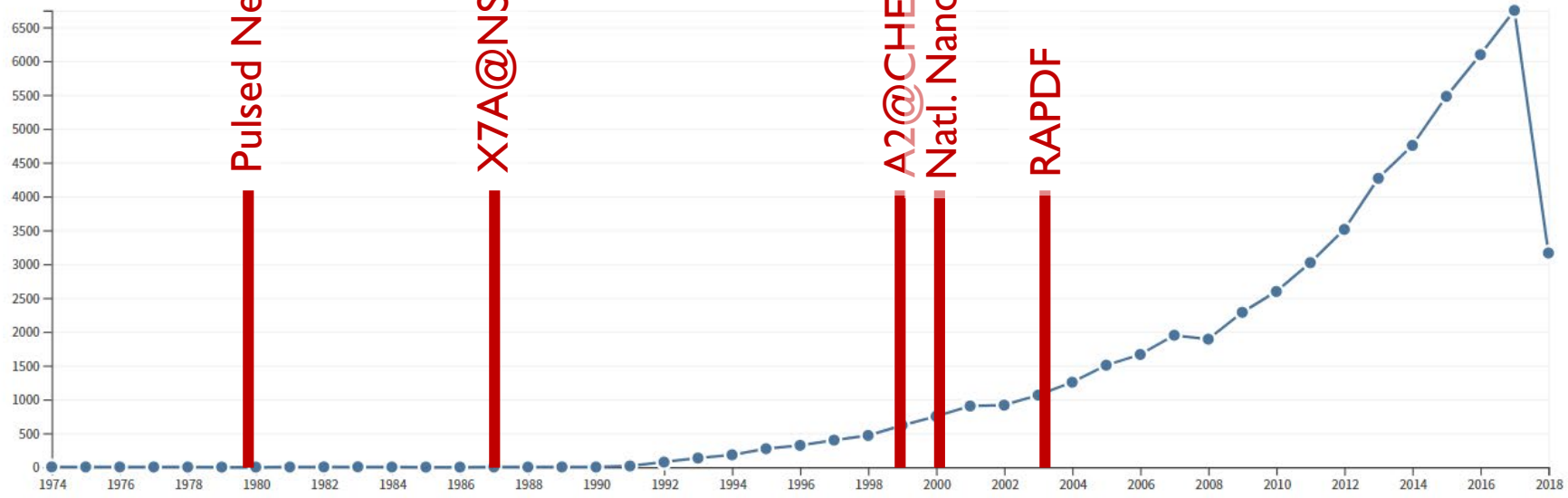
Sum of Times Cited
56,306

Without self citations
50,593

Citing articles
43,758

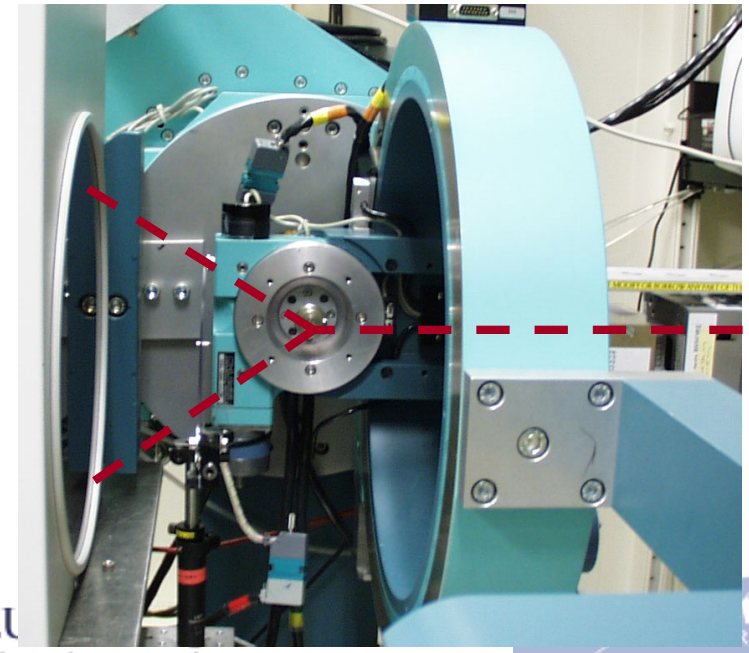
Without self citations
42,303

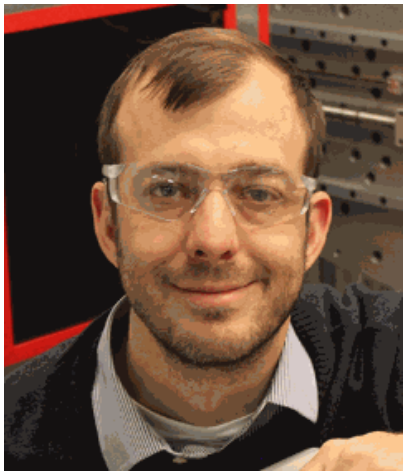
Sum of Times Cited per Year



Synchrotron PDF

- Early x-ray PDF measurements at X7A, NSLS-II (20 keV, Ge point detector), led by Takeshi Egami
- First high resolution x-ray PDF measurement (Q_{max} = 45 A^{-1}!) using CHESS, Ge point detector, with Stefan Kycia
- National Nanotechnology Initiative
- Rapid Acquisition PDF (RAPDF) developed at 6ID-D at APS, with Pete Chupas, 2D 2003





HTTP://thebillingegroup.com

PDF landscape in 2018

Web of science (x-ray + PDF) OR (x-ray + pair distribution function)

Total Publications
2,470

***h*-index**
96

Average citations per item
22.8

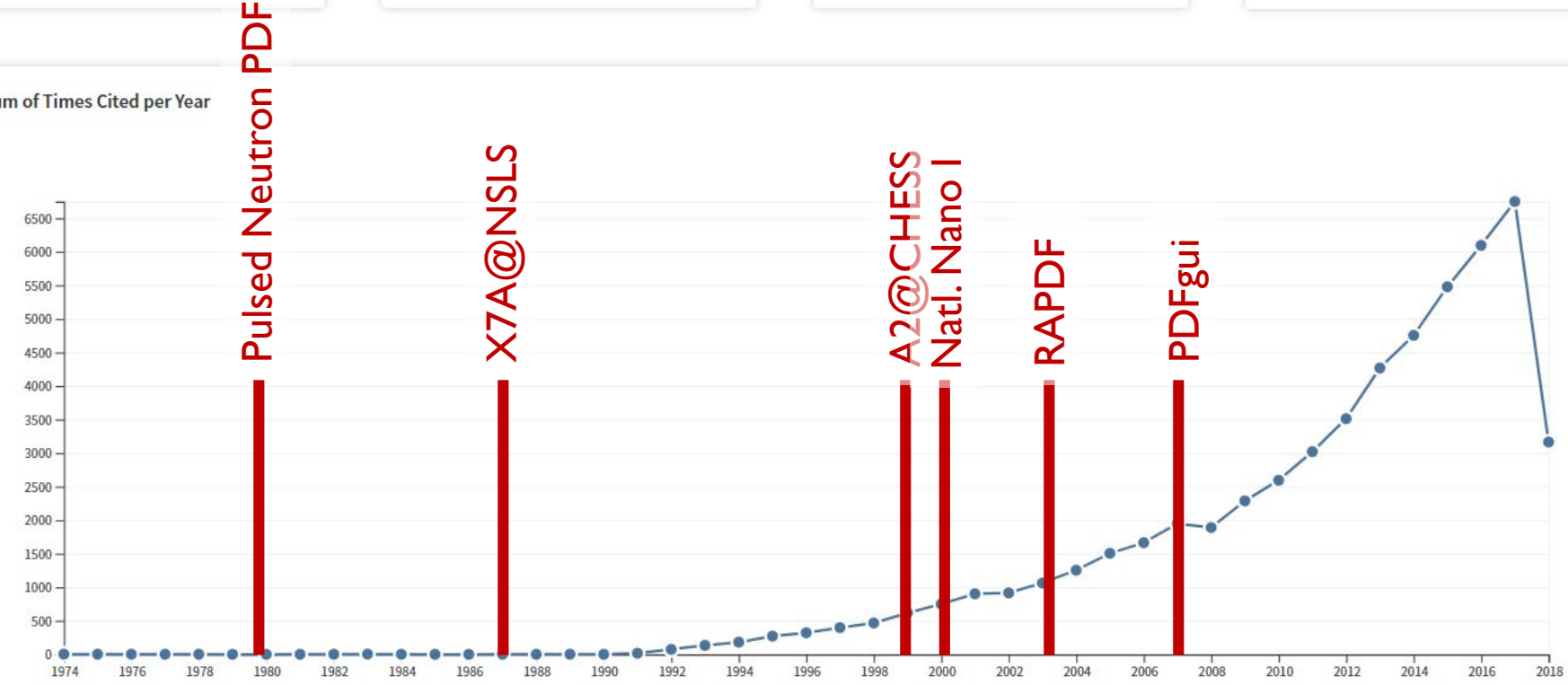
Sum of Times Cited
56,306

Without self citations
50,593

Citing articles
43,758

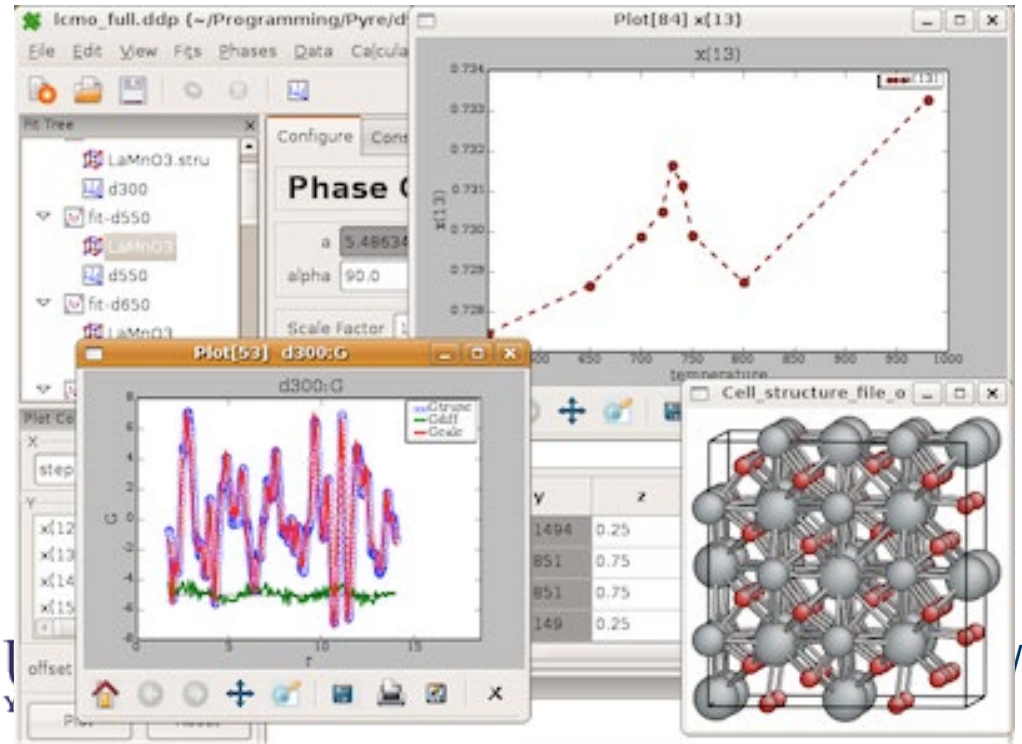
Without self citations
42,303

Sum of Times Cited per Year

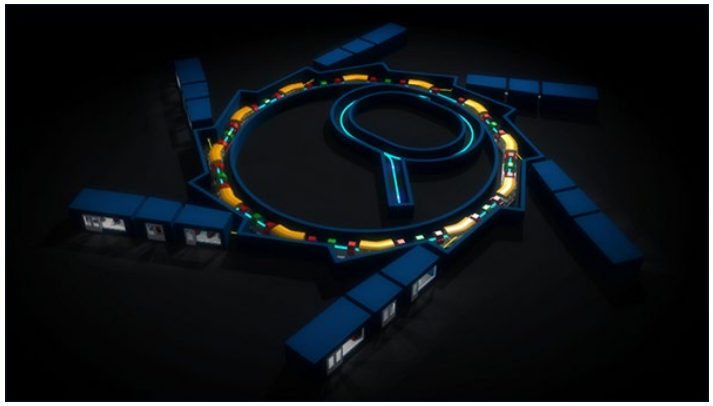


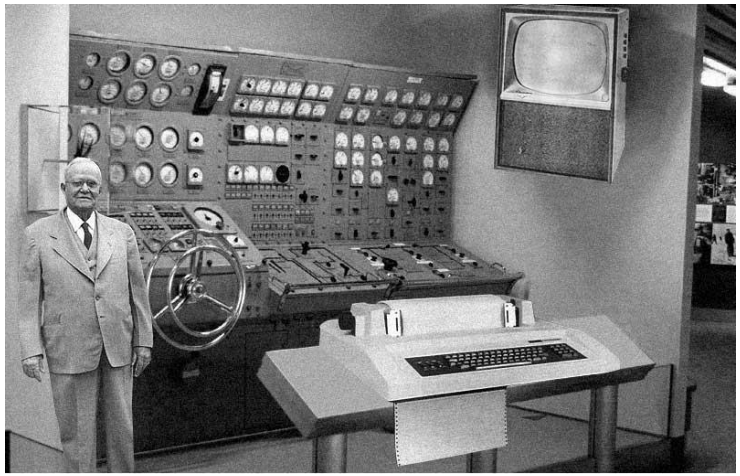
Synchrotron PDF

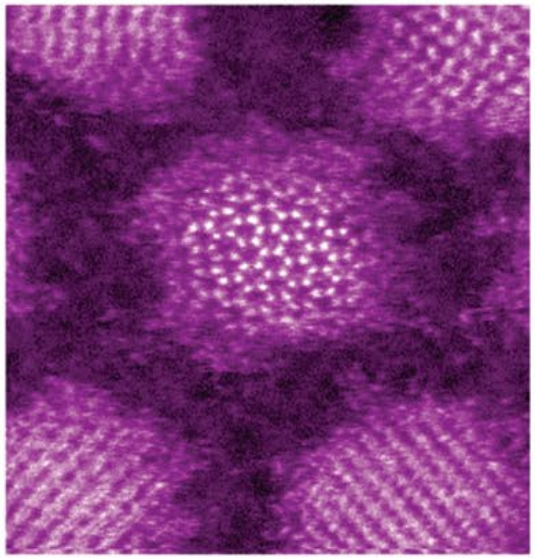
- Early x-ray PDF measurements at X7A, NSLS-II (20 keV, Ge point detector), led by Takeshi Egami
- First high resolution x-ray PDF measurement (Q_{max} = 45 A^{-1}!) using CHESS, Ge point detector, with Stefan Kyci
- National Nanotechnology Initiative
- Rapid Acquisition PDF (RAPDF) developed at 6ID-D at APS, 2003 with Pete Chupas, 2D
- PDFgui



The perfect storm

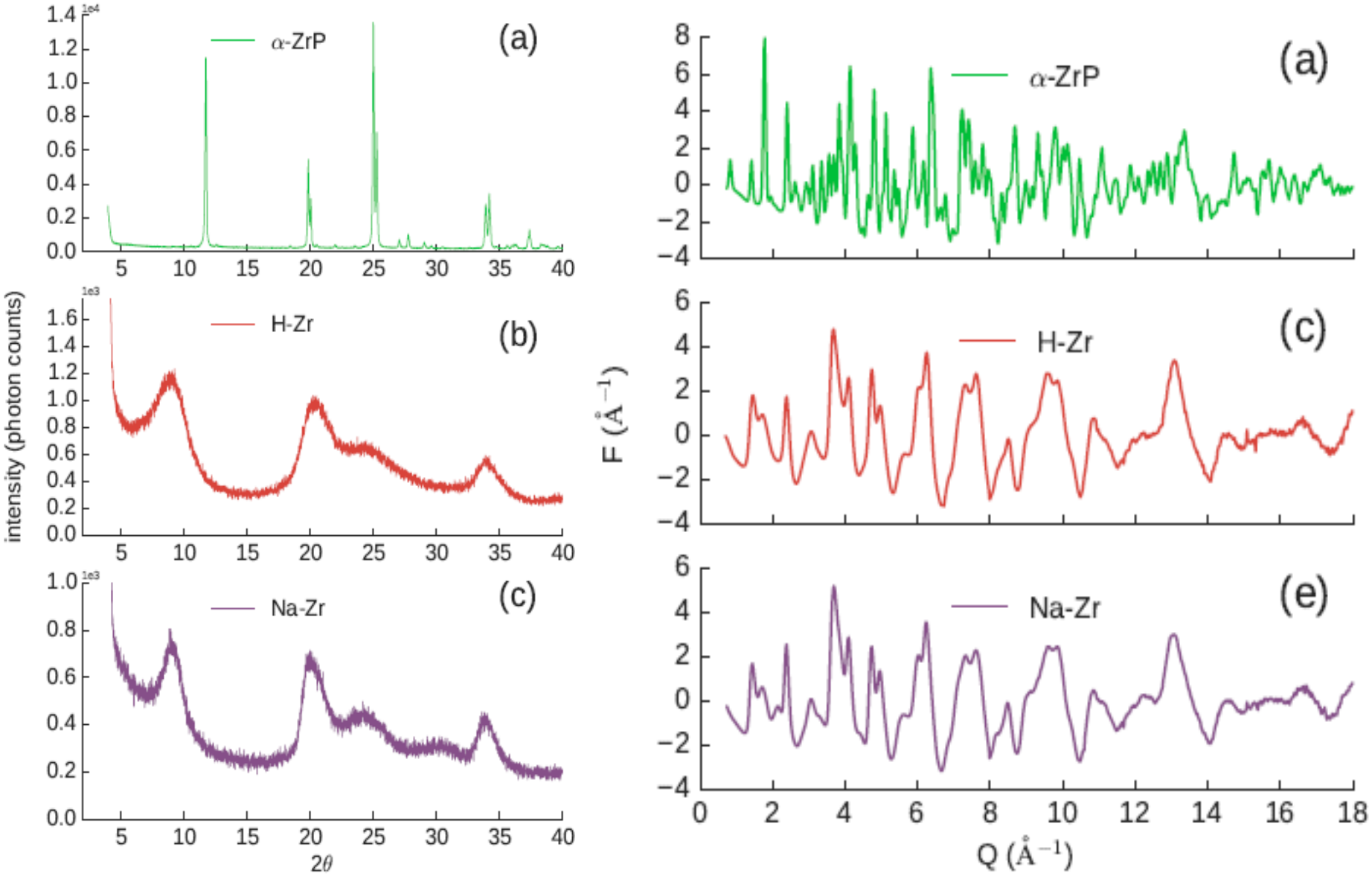




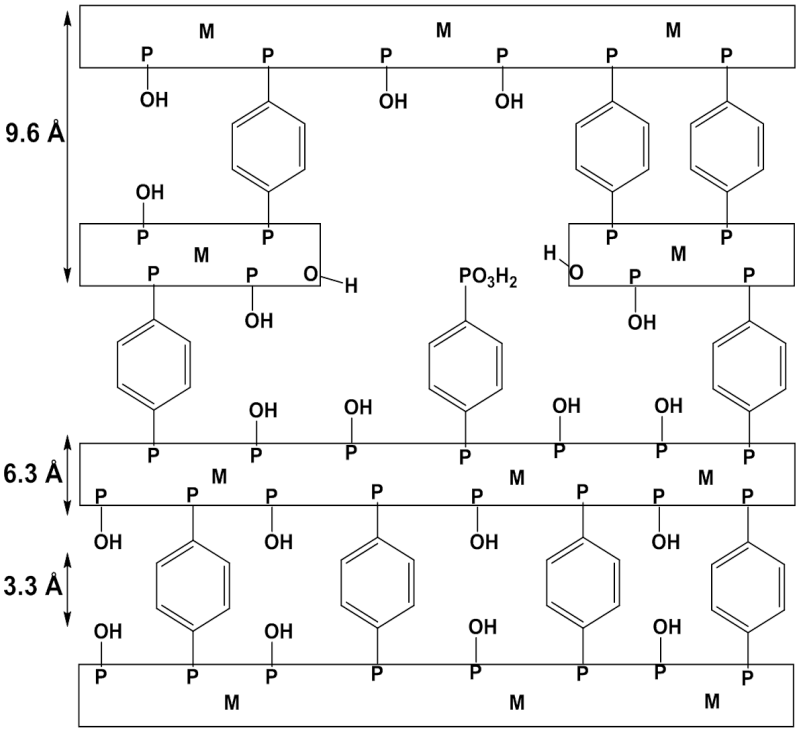


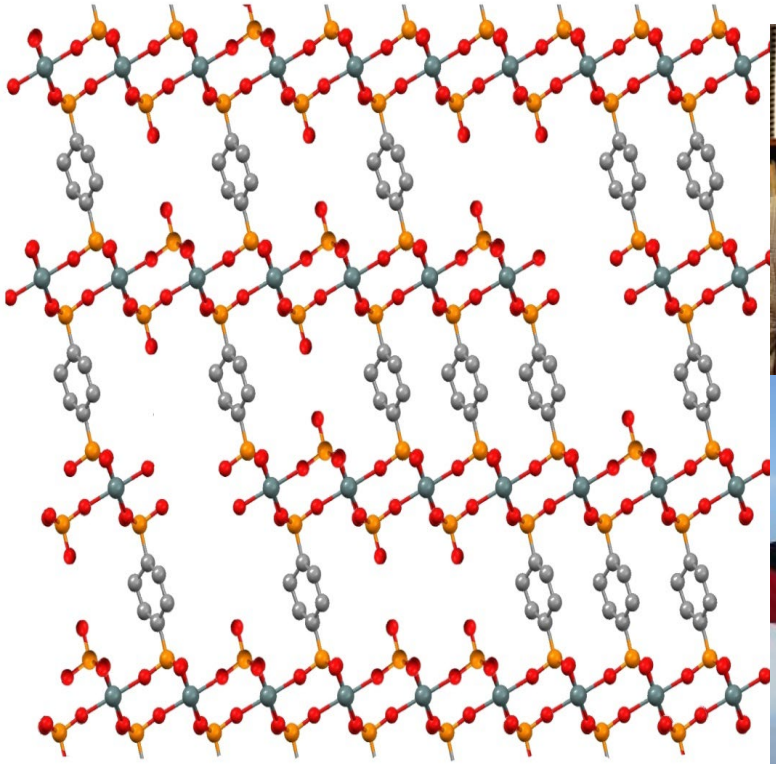
Sources/detectors + computing + nanoscience

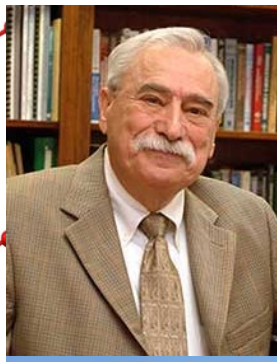
You must remember this



Zr/Sn phenyl phosphate, unconventional MOFs







Clearfield, A. *Dalton Trans*. 2008, **44**, 6089-6102.

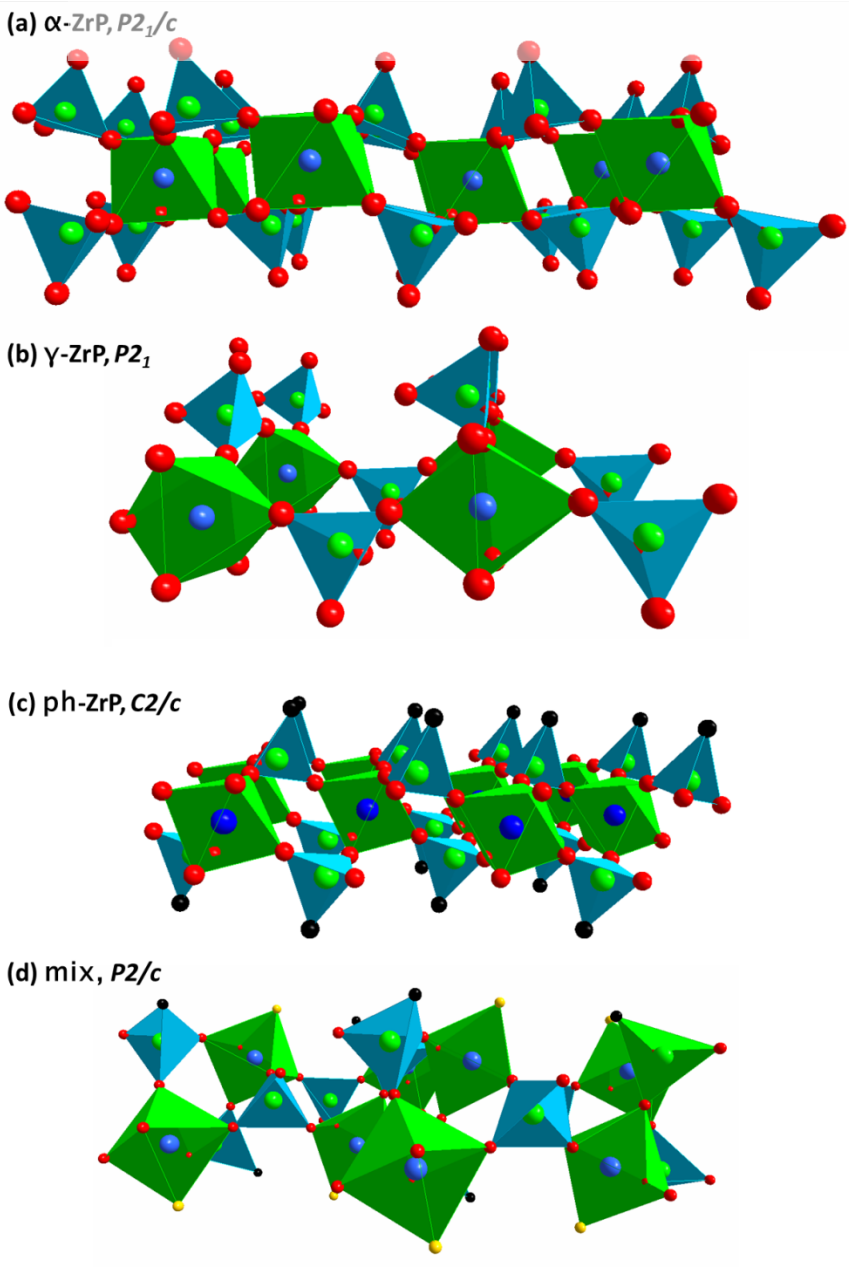
- Collaboration with the group of Abe Clearfield, Texas A&M
- Work of Chenyang Shi, Max Terban and Rita Silbernagel

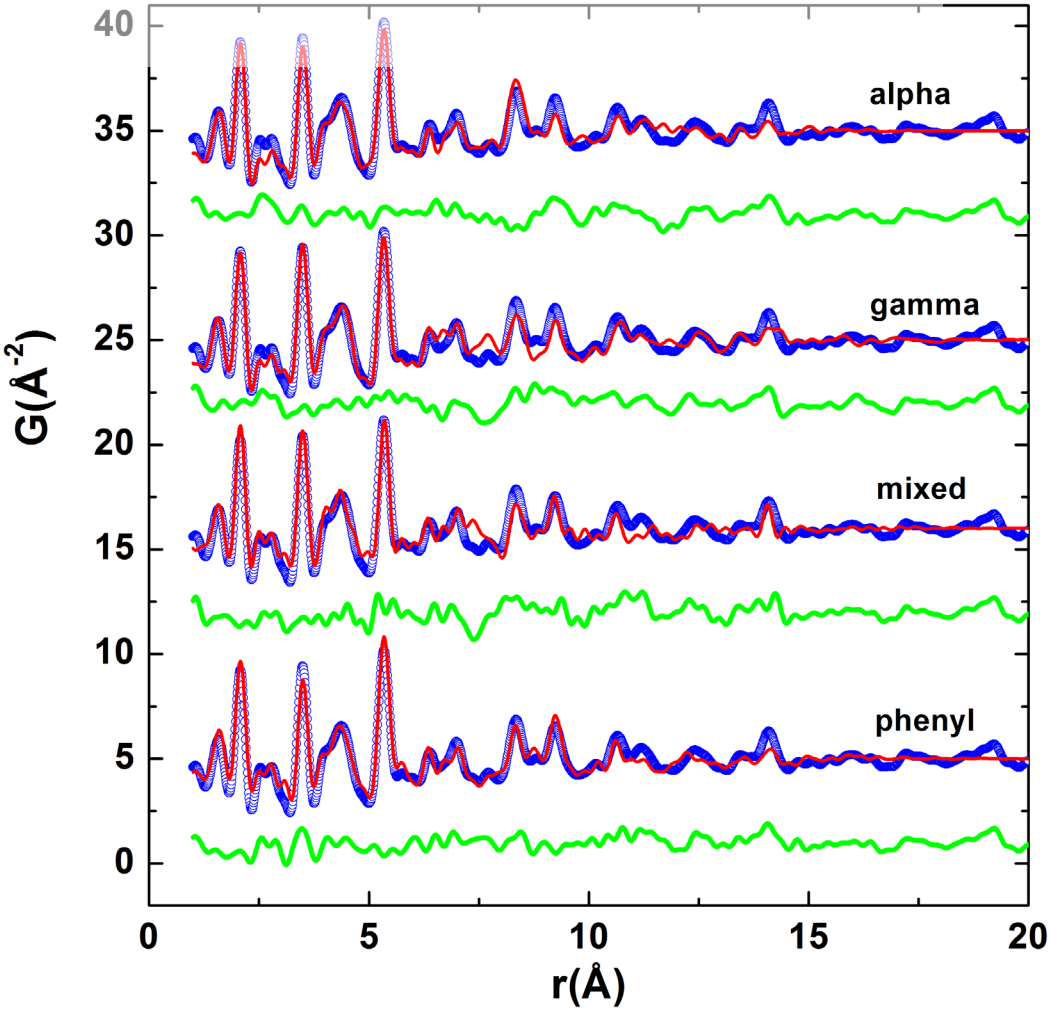
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegro

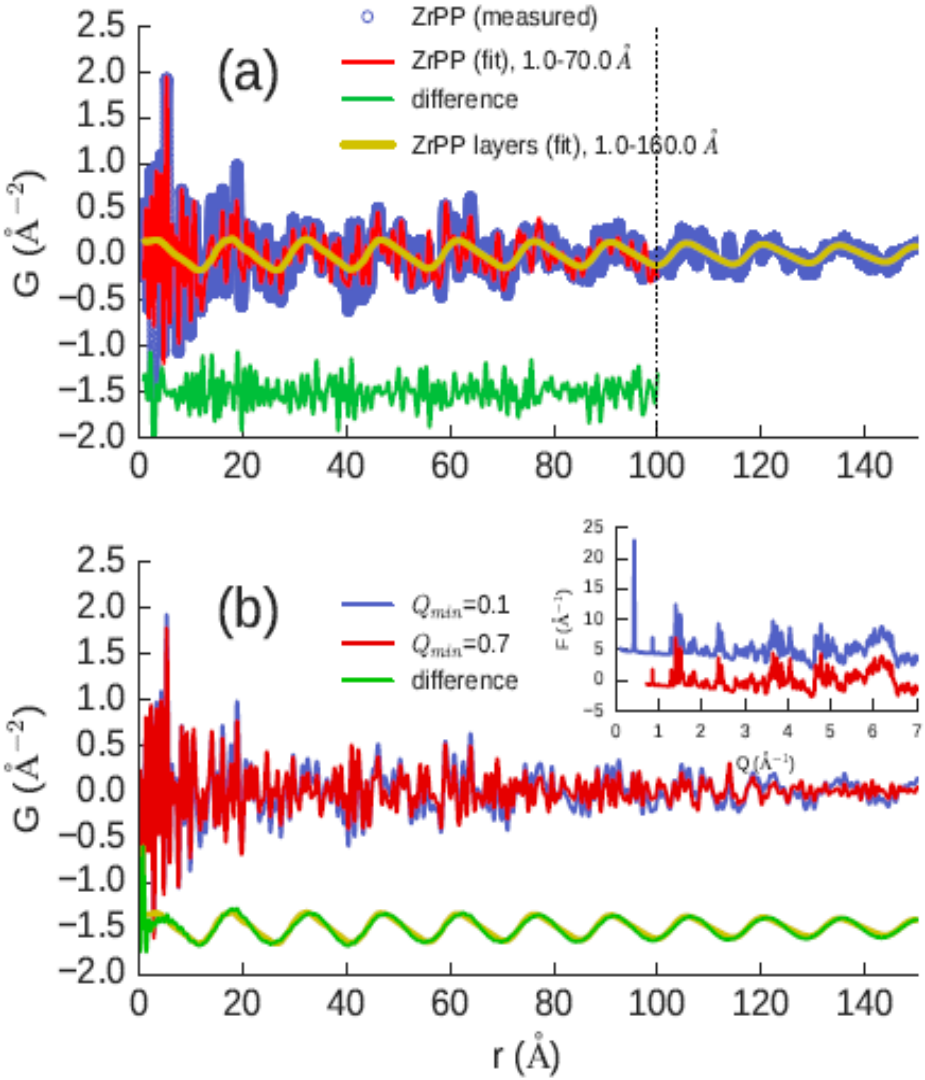
Try ZrP structures on H-Zr sample, PDFgui





- Fits are good up to 7 Å: basic motifs, i.e. ZrO_6 and PO_4 are similar in all models.
- Fits of phenyl and alpha ones are similar
- Fits of phenyl and alpha are better.

Studying the layer stacking



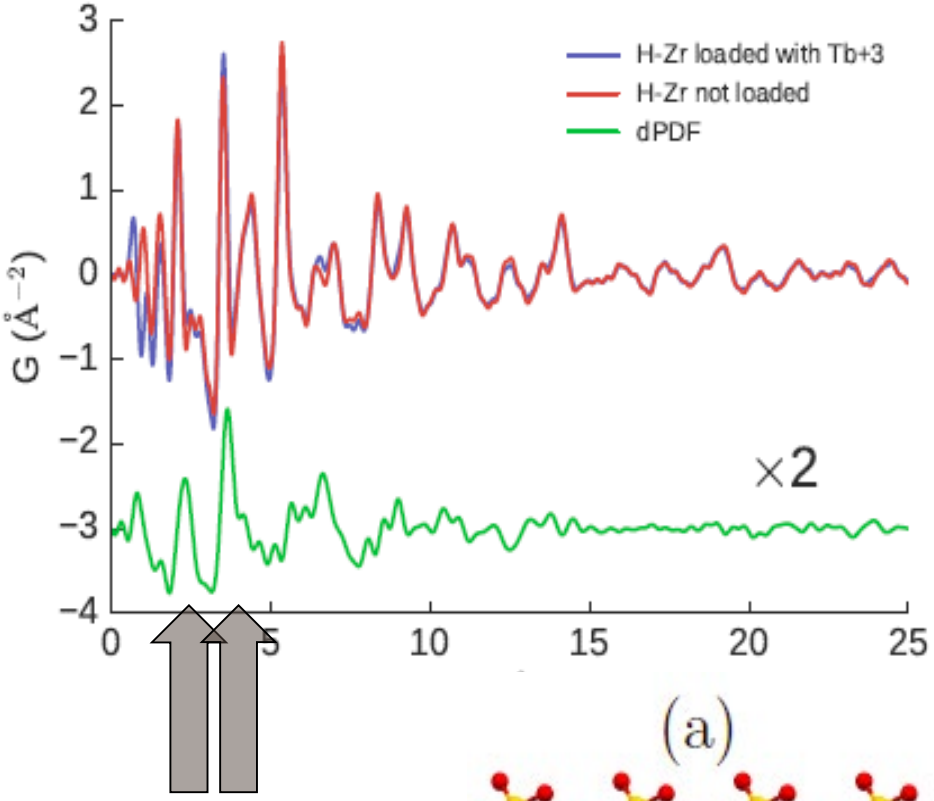
- ZrPP layers
- Smear out electron-density *in*, but *not perpendicular* to the layers (gold)
- Get beautiful "Warren lineshape" signal extending to high-r
- Comes from the layer stacking, regardless of any intra-order

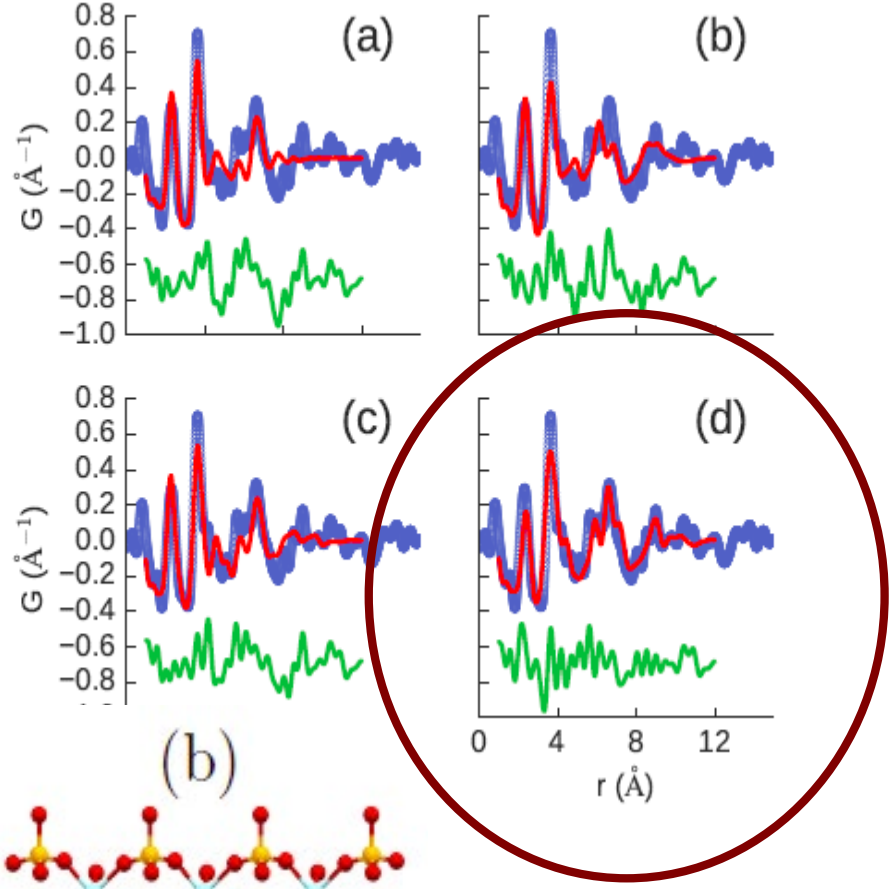
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

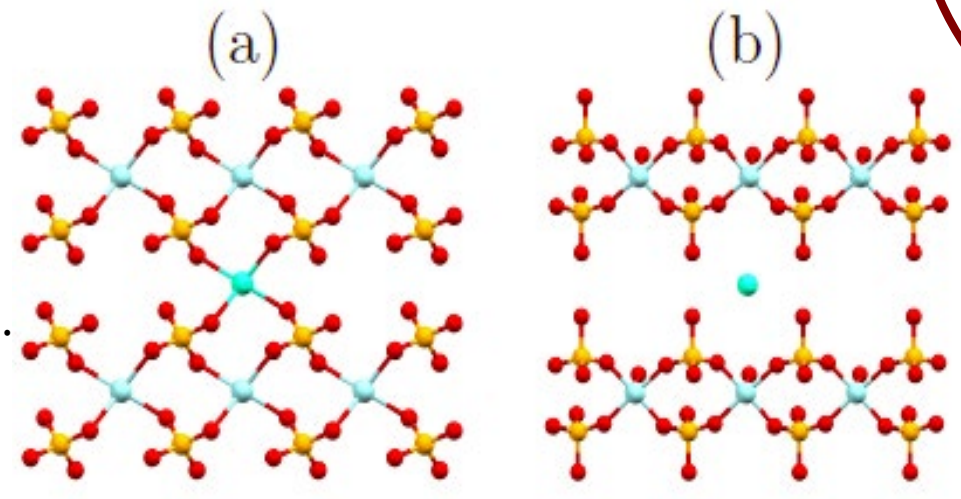
BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

Finding Tb intercalated into ZrPP

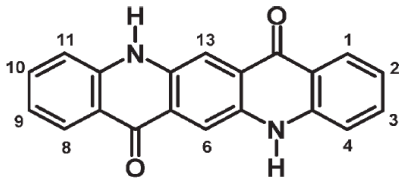


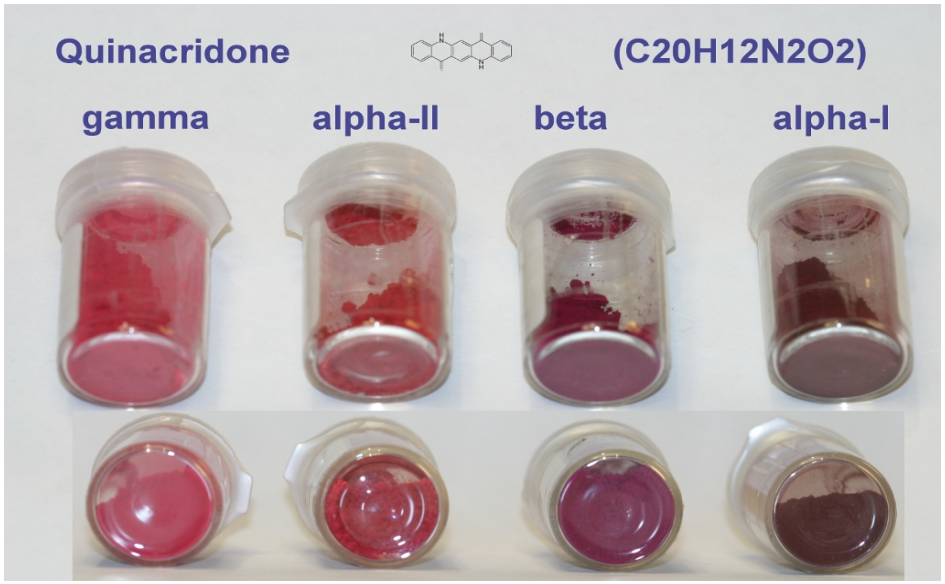




Maxwell W. Terban, Chenyang Shi, Rita Silbernagel, Abraham Clearfield and Simon J. L. Billinge, *Inorg. Chem.* **56**, 8837-8846 (2017).

X-ray PDF of weakly scattering organics





collaboration with Prof. Martin U. Schmidt and Dragica Prill, Inst. fur Anorg. Chemie, Goethe-Univ., Frankfurt am Main

quinacridone – $C_{20}H_{12}N_2O_2$

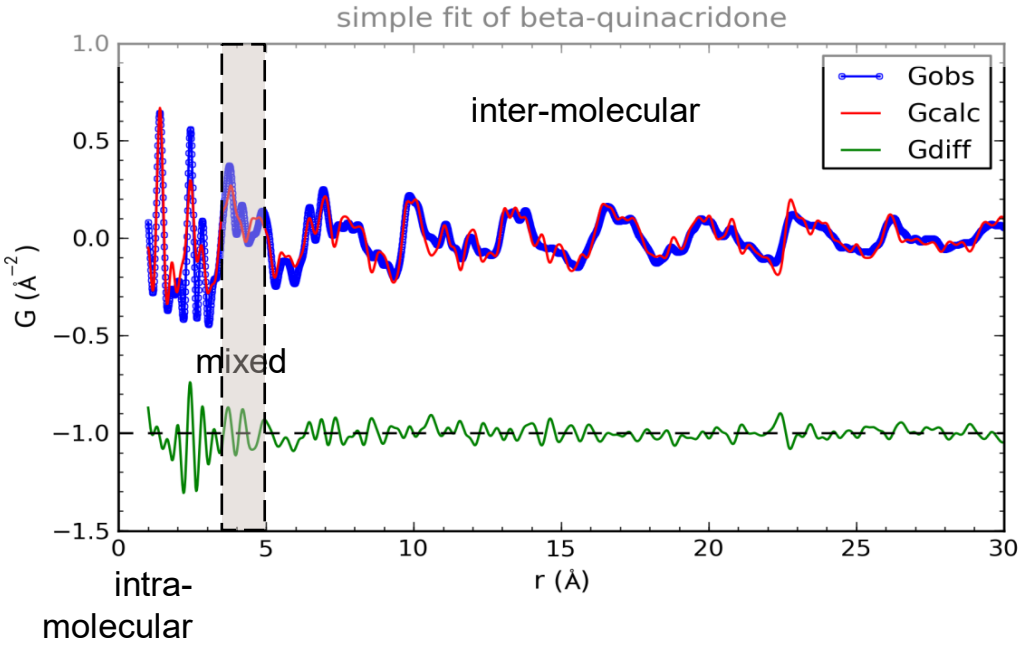
- industrially important pigments, red and violet paints
- can form many phases, some (alpha-II) do not crystallize and have unknown structure
- experimental PDFs measured at APS ANL beamline 11-ID-B
- standard refinement with PDFgui has poor quality even for the known beta phase

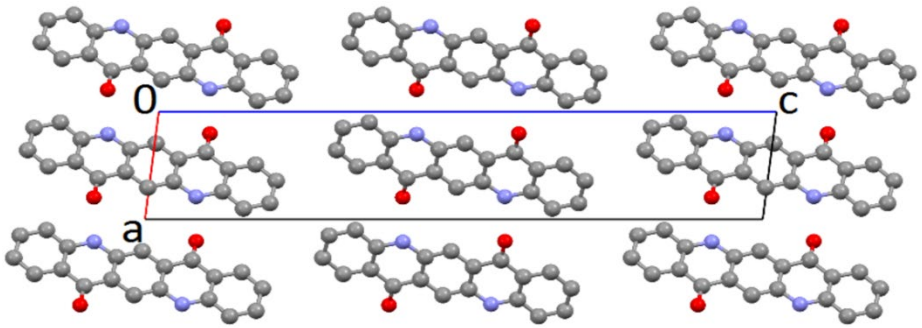
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

PDF modeling of β-quinacridone





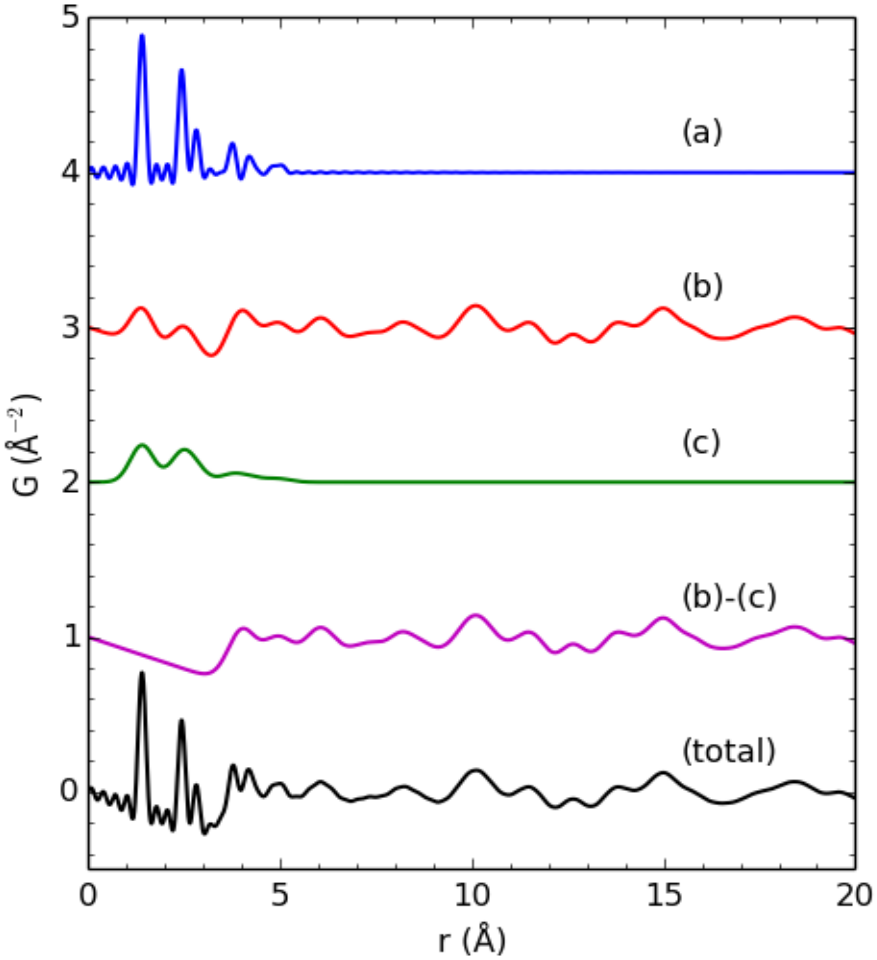
- monoclinic $P2_1/c$ with 2 molecules per unit cell
- refinement in PDFgui gives poor $\boldsymbol{R_w = 0.41}$
  - low-*r* peaks too wide
  - high-*r* peaks too sharp
- peak widths depend strongly on *r*
  - sharp peaks for strongly bonded intra-molecular atom pairs
  - broad peaks for inter-molecular correlations
- Debye-Waller model of independent atom displacements is grossly inaccurate
- PDF model has to account for different displacements for pairs in the same molecule and inter-molecular pairs

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

PDF modeling of β-quinacridone



PDF calculation with separate intra- and inter-molecular contributions

(a) PDF from a single-molecule, small atom displacements U_{intra}

(b) PDF from a crystal with large displacements U_{inter}

(c) PDF from a molecule with large displacements U_{inter}

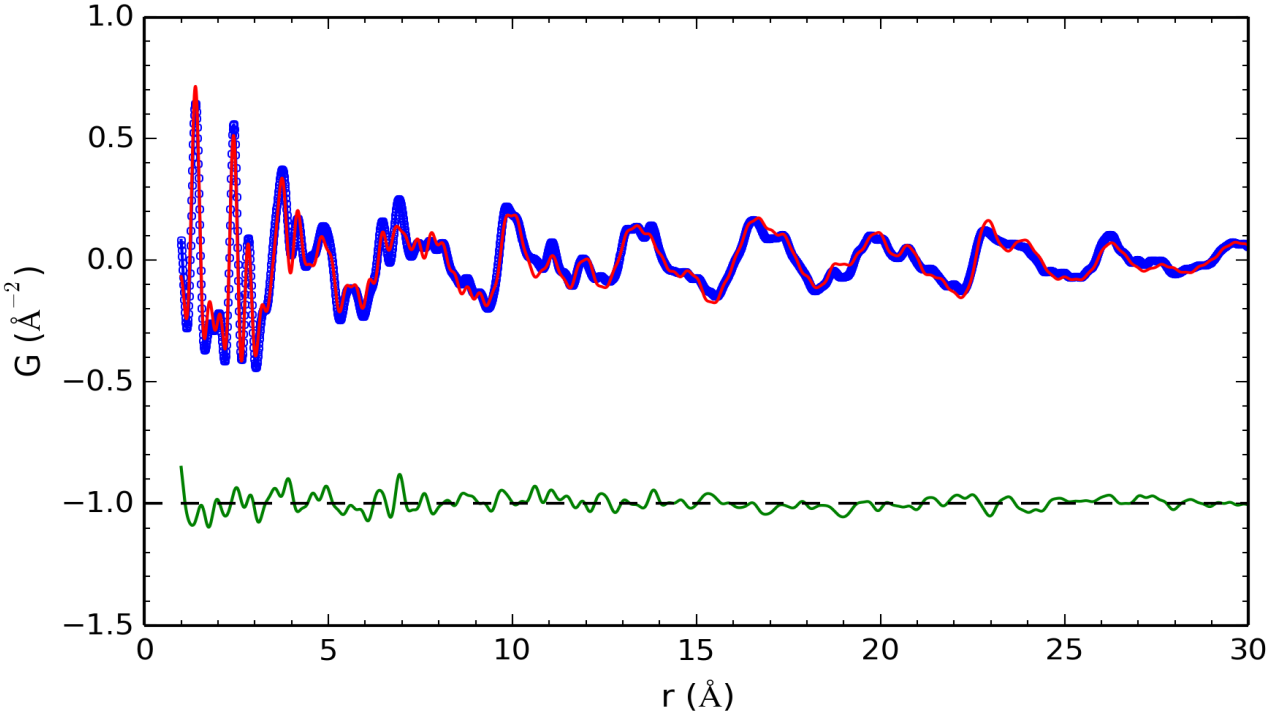
(b) – (c) PDF from inter-molecular interactions only

(a) + (b) – (c) total PDF reflecting both displacements U_{intra}, U_{inter}

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

PDF modeling of β-quinacridone



- refined unit cell, data scale and displacement factors
  U_{inter} = 0.0014(2) Å^2 U_{intra} = 0.023(2) Å^2
- significant fit improvement R_w = 0.41 → **R_w = 0.28**
- remnant fit difference due to anisotropic molecule displacements, displacement anisotropy can be studied with improved models

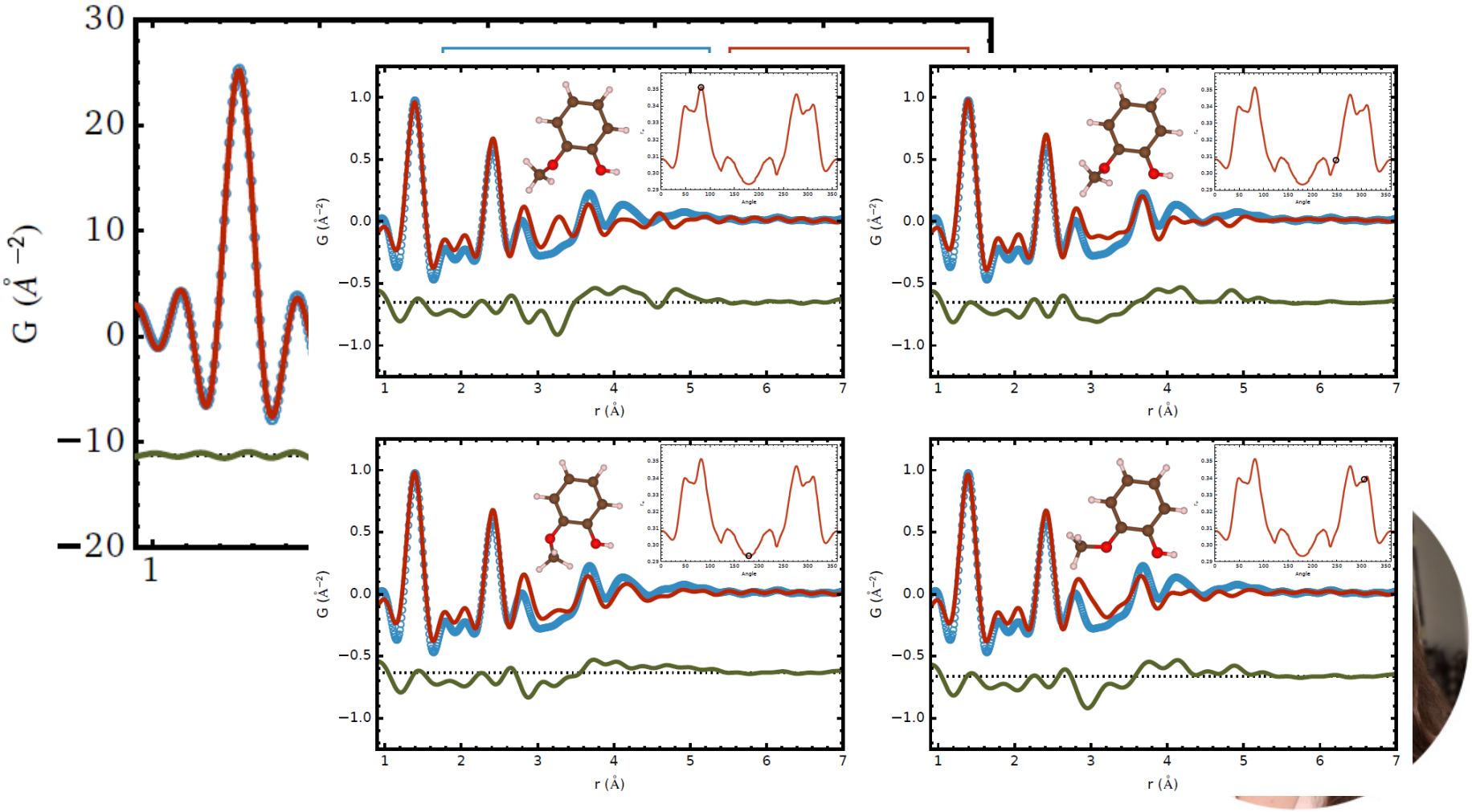
D. Prill *et al.*, J. Appl. Crystallogr. 48, 171-178 (2015)

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

Studying conformational degrees of Freedom

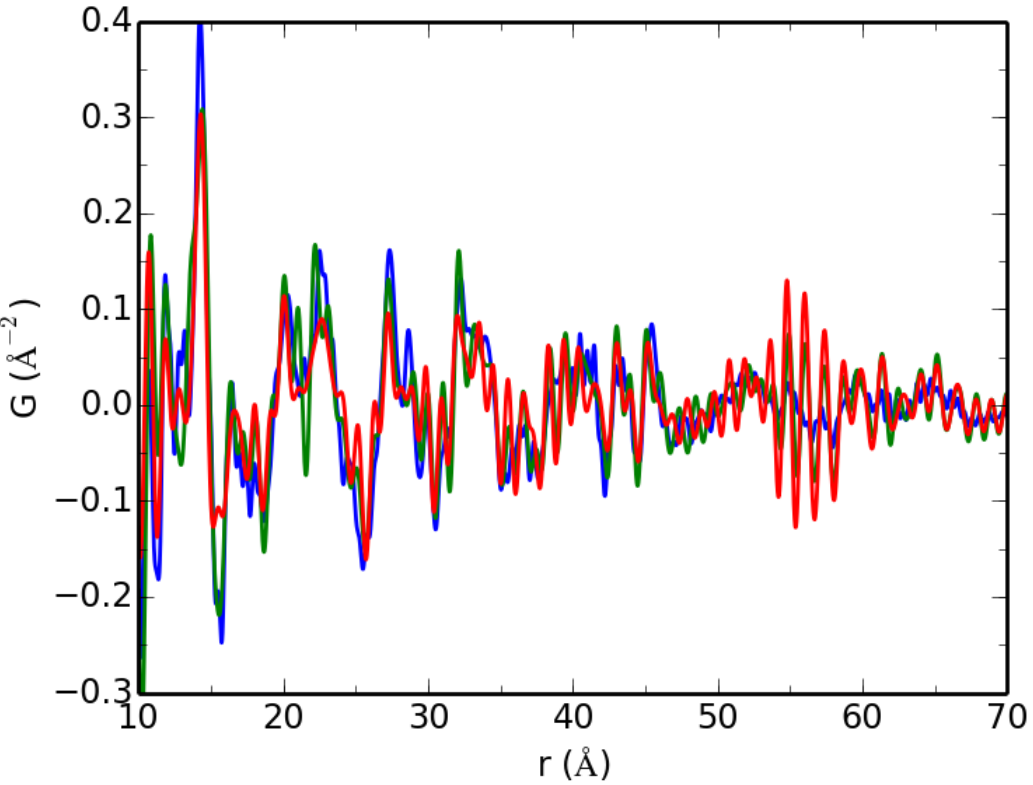


COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

Dilute signals from drugs in suspension



- Blue: 5wt%
- Green 0.66 wt%
- Red 0.25 wt%

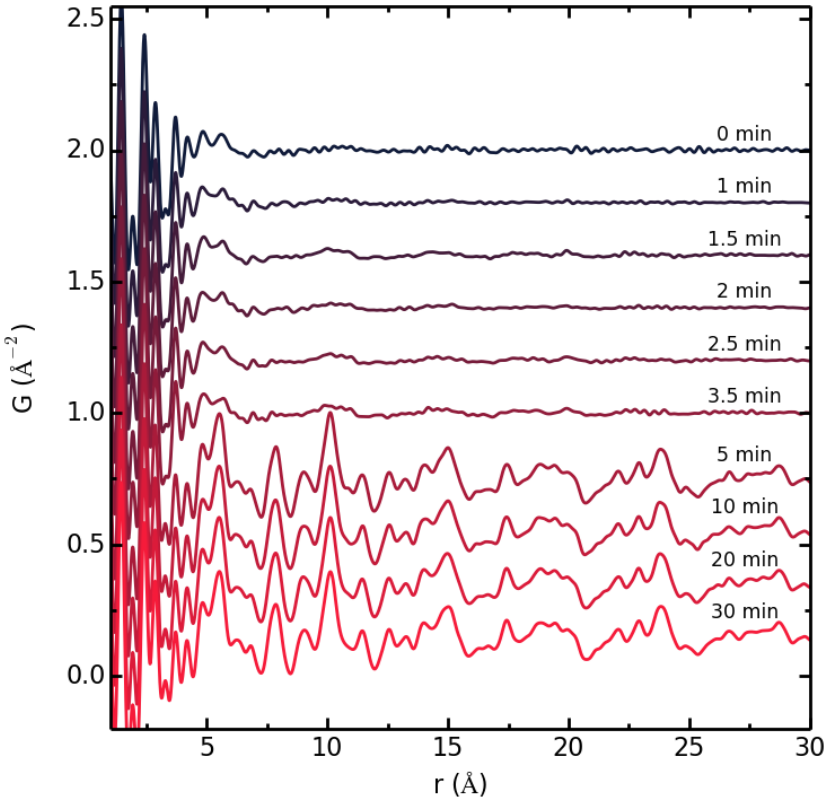
Can see nanocrystalline API even at the 0.25wt% level!

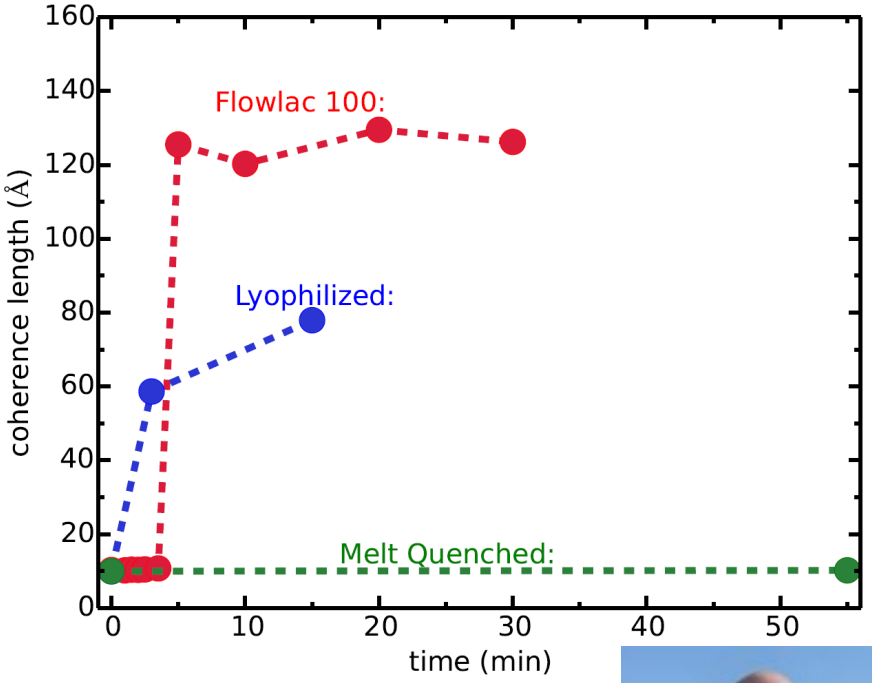
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

Recrystallization of lactose

- Various A-form samples aged at, 40°C/75% RH





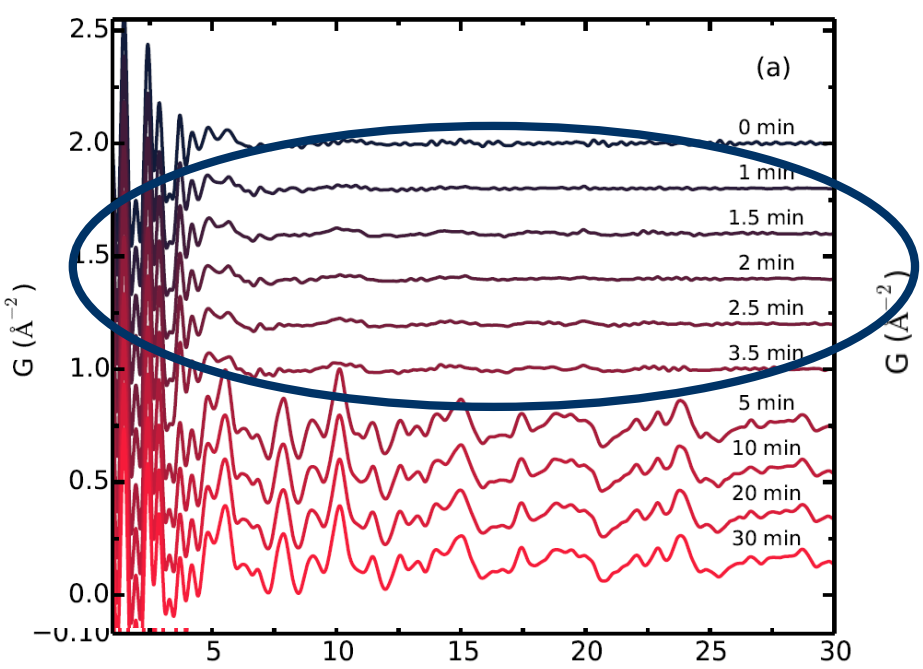
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

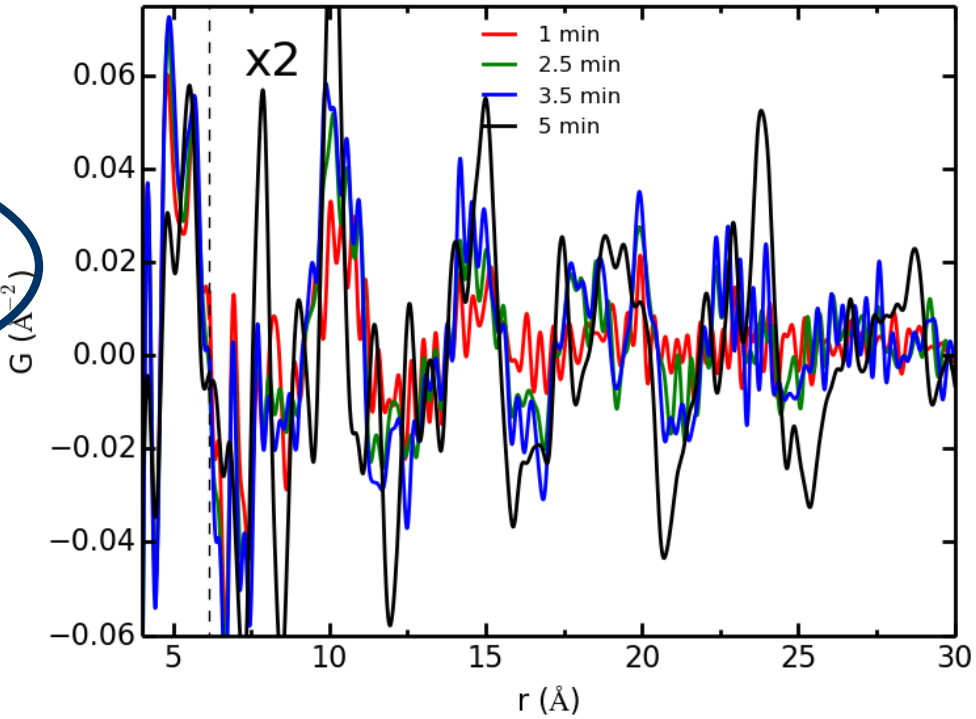
BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingeg

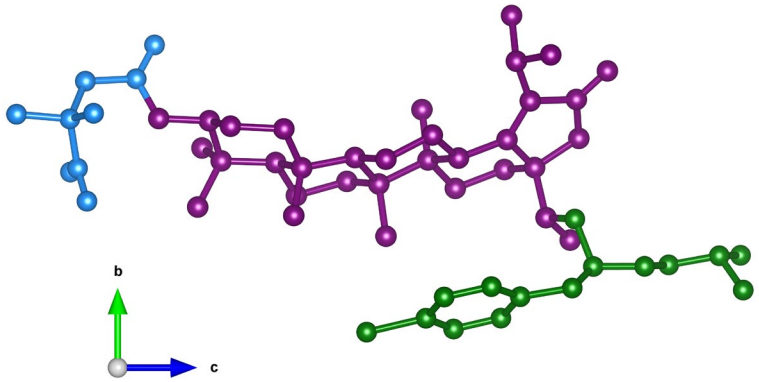
Understanding crystallization

- How does the crystallinity grow in?





Drug candidate: conformation changes in the amorphous state



Modeling of structure in the amorphous state

Findings:

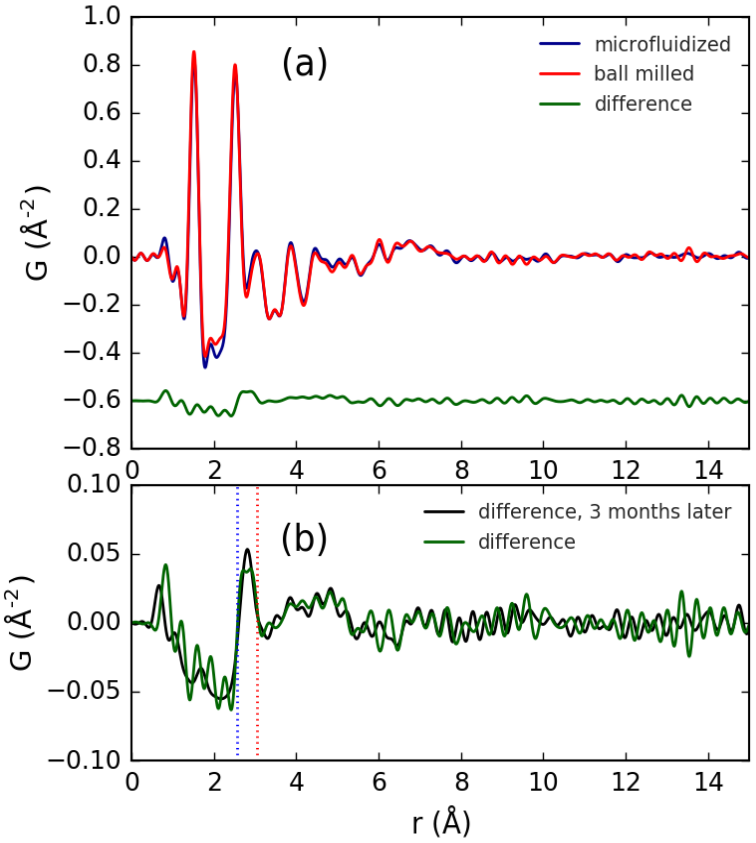
- Conformational disorder: Need to assign a separate thermal factor between different intra-regions of the molecule.
- Subtle differences between different micronization methods

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingeg

Hydrogen bonding interactions vary depending on how the material was amorphized



Distinct differences found between the PDFs of differently produced amorphous forms

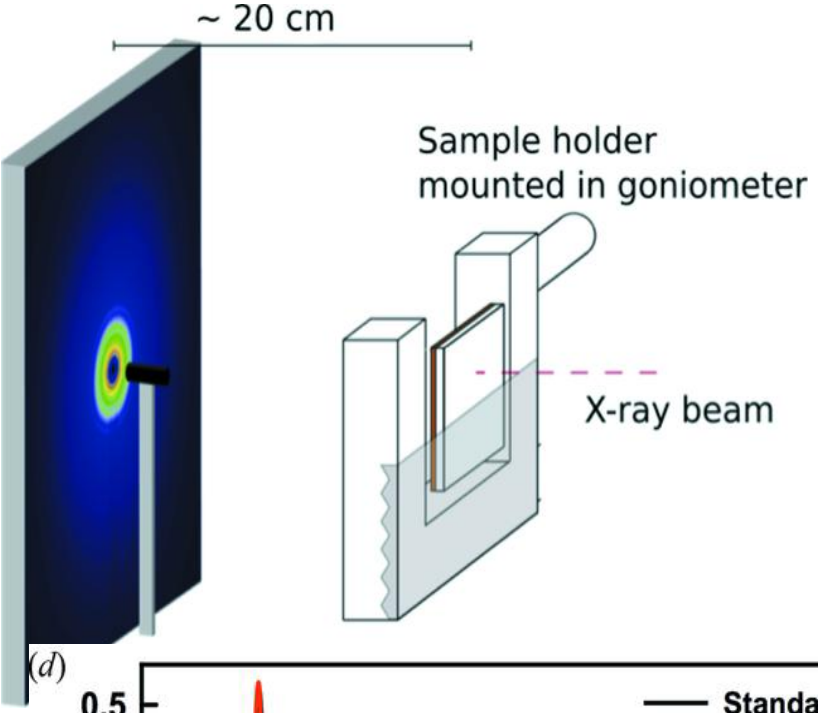
- prepared via mechanical versus solution based methods
- A sharp feature in the residual corresponds to the minimum and maximum hydrogen bonding distances, indicating a slight change in the intermolecular packing.

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

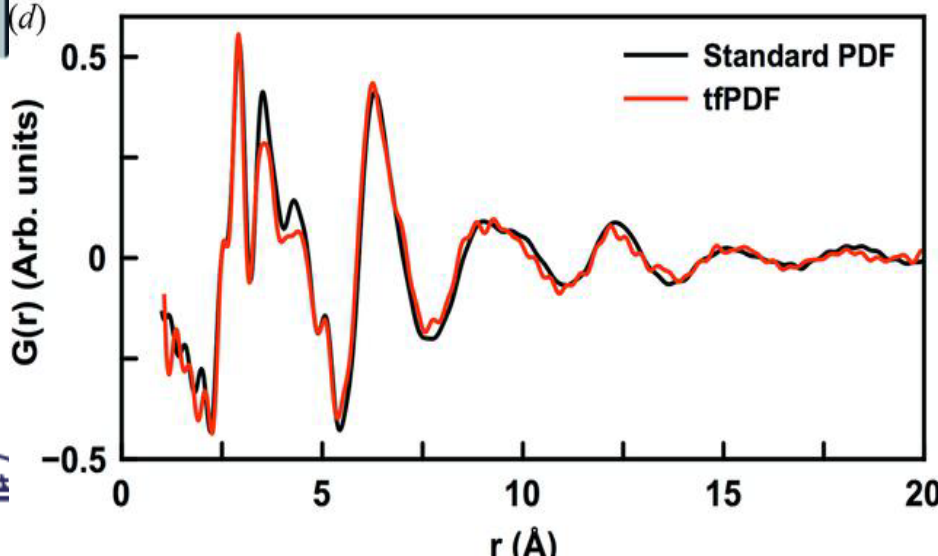
BROOKHAVEN NATIONAL LABORATORY

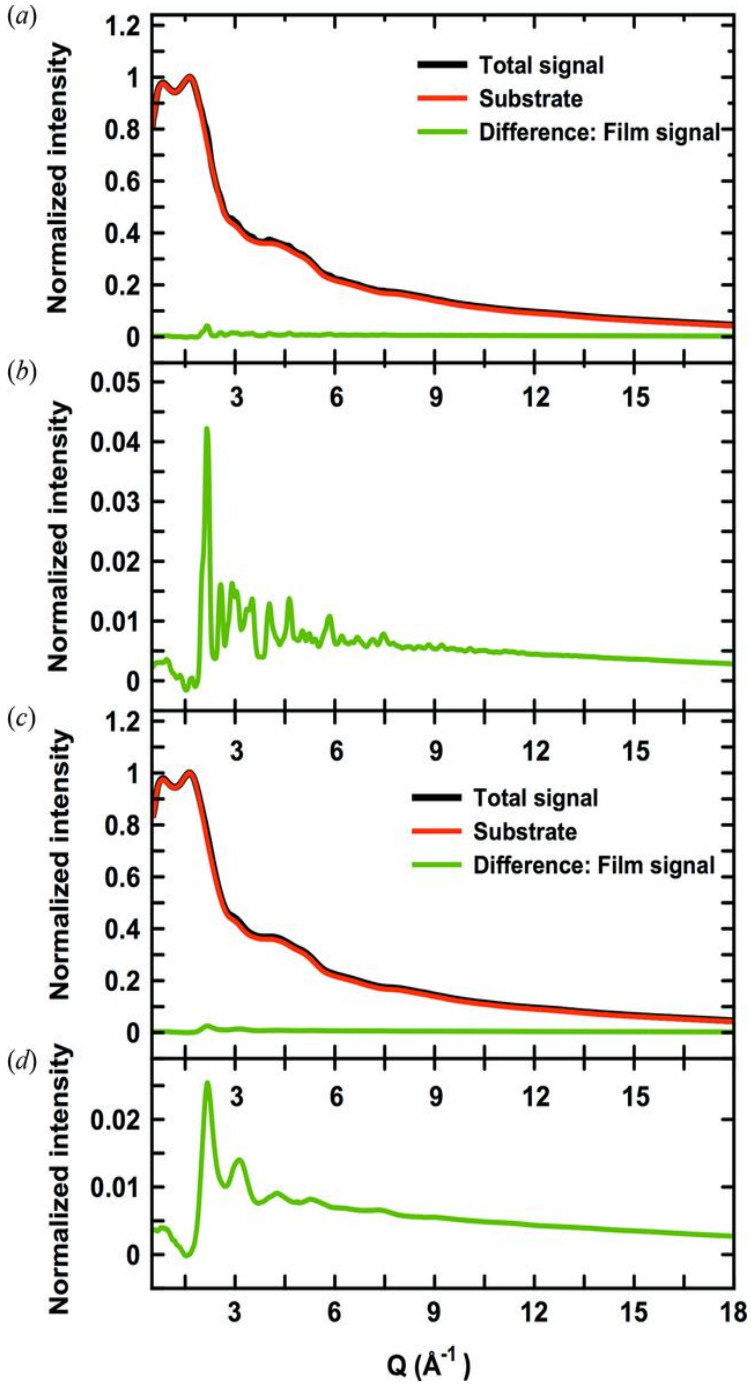
HTTP://thebillingegro

Dilute systems => Thin film PDF at normal incidence



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





PDFgui

- Used to be called "real-space Rietveld"
  - Starting point is a crystallographic description of the model
  - Small Unit cell, periodic boundary conditions
  - Fits to G(r) which is the FT of F(Q)

- PDFgui
  - Available from diffpy.org
  - Development funded by NSF-DANSE program

PbSe-480K.ddp (~\Documents\gl\16papers\16eb_pbse-emphanisis\new\20170412\Structure modeling\Recent data\PbSe-480K.ddp) - PDFgui

File Edit View Fits Phases Data Calculations Help

Fit Tree

- PbSe
  - PbSe.cif
  - PbSe.gr

Plot Control

X

step

Y

lat(1)
lat(2)
lat(3)
pscale
u11(1)
u11(2)

offset -5

Plot Reset

Configure | Constraints | Results

Phase Configuration

a	6.128	b	6.128	c	6.128
alpha	90.0	beta	90.0	gamma	90.0

Scale Factor	1.0				
delta1	0.0	delta2	1.0	spdiameter	0.0
sratio	1.0	rcut	0.0	stepcut	0.0

Included Pairs: all-all

	elem	x	y	z	u11	u22	u33	u12	u13	u23	occ
1	Pb	0.5	0.5	0.5	0.0186431	0.0186431	0.0186431	0.0	0.0	0.0	1.0
2	Pb	0.5	0.0	0.0	0.0186431	0.0186431	0.0186431	0.0	0.0	0.0	1.0
3	Pb	0.0	0.5	0.0	0.0186431	0.0186431	0.0186431	0.0	0.0	0.0	1.0
4	Pb	0.0	0.0	0.5	0.0186431	0.0186431	0.0186431	0.0	0.0	0.0	1.0
5	Se	0.0	0.0	0.0	0.0143116	0.0143116	0.0143116	0.0	0.0	0.0	1.0
6	Se	0.0	0.5	0.5	0.0143116	0.0143116	0.0143116	0.0	0.0	0.0	1.0
7	Se	0.5	0.0	0.5	0.0143116	0.0143116	0.0143116	0.0	0.0	0.0	1.0
8	Se	0.5	0.5	0.0	0.0143116	0.0143116	0.0143116	0.0	0.0	0.0	1.0

PDFfit2 Output

PbSe-480K.ddp (~\Documents\gl\16papers\16eb_pbse-emphanisis\new\20170412\Structure modeling\Recent data\PbSe-480K.ddp) - PDFgui

File Edit View Fits Phases Data Calculations Help

Fit Tree

- PbSe
  - PbSe.cif
  - PbSe.gr

Plot Control

X

step

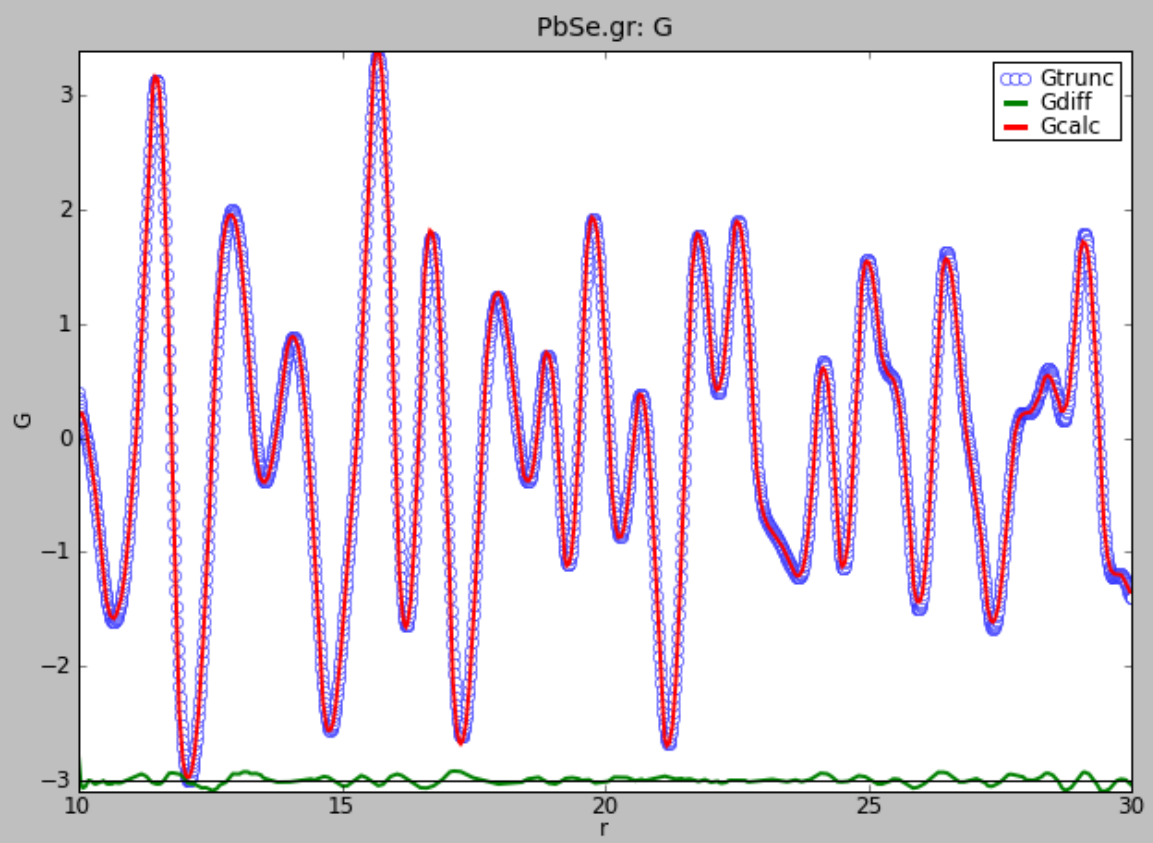
Y

Gcalc
Gdiff
Gobs
Gtrunc
crw
dGcalc

offset -3

Plot Reset

Plot [2] PbSe.gr: G



x = 22.619, y = 2.69891

PDFfit2 Output

CMI: complex modeling infrastructure

- CMI
  - Also available at diffpy.org
  - Very powerful
  - Very difficult to use
  - We are working on a more friendly user interface and better documentation, but for now it is advised to contact us and we can help you get started

Diffpy-Complex Modeling Infrastructure: Diffpy-CMI

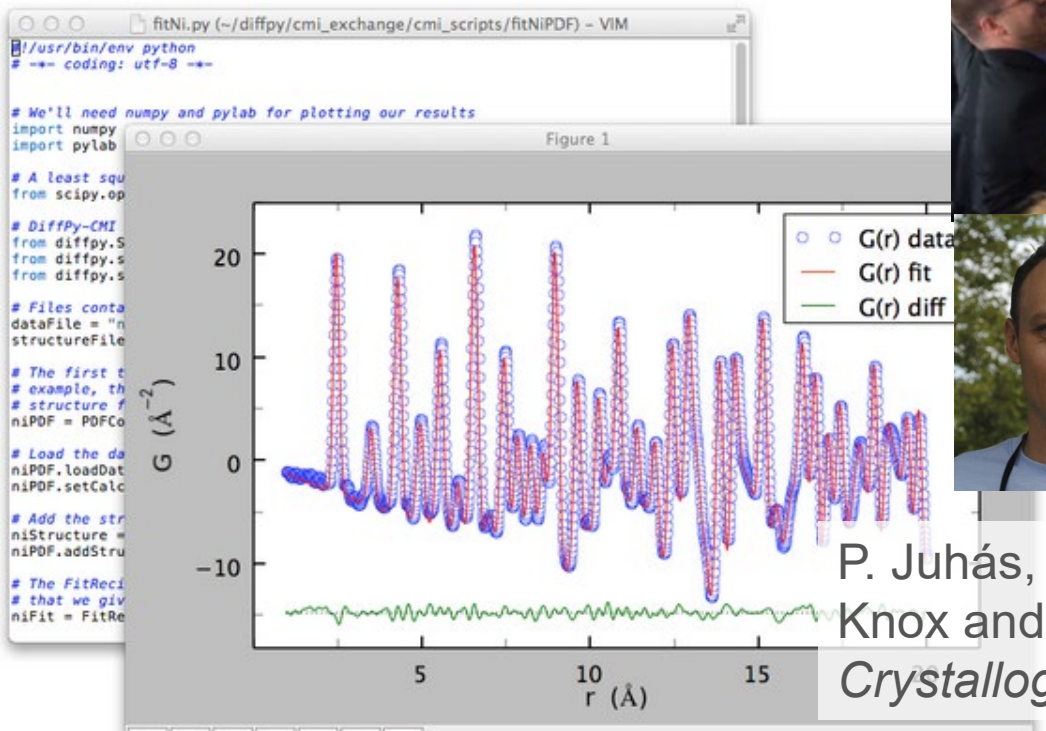
DiffPy Community Publications Products

DiffPy-CMI

DiffPy-CMI is our complex modeling framework. It is a highly flexible library of Python modules for robust modeling of nanostructures in crystals, materials.

The software provides functionality for storage and manipulation of structure data and calculation of structure-based quantities, such as PDF, SA overlaps, bond lengths, and coordinations. Most importantly the DiffPy-CMI package contains a fitting framework for combining multiple experime problem.

This is an early release of code that is under intense development, with support for installation on Unix, Linux, and Macintosh machines. The sco will evolve rapidly, but we want to make the code available at the earliest possible date. Please make use of the software and provide feedback a but please be patient and check back frequently for updates.



P. Juhás, C. L. Farrow, X. Yang, K. R. Knox and S. J. L. Billinge, *Acta Crystallogr. A*, (2015).

untitled • - Sublime Text

File Edit Selection Find View Goto Tools Project Preferences Help

untitled

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-


# We'll need numpy and pylab for plotting our results
import numpy as np
import pylab

# A least squares fitting algorithm from scipy
from scipy.optimize.minpack import leastsq

# DiffPy-CMI modules for building a fitting recipe
from diffpy.Structure import loadStructure
from diffpy.srfit.pdf import PDFContribution
from diffpy.srfit.fitbase import FitRecipe, FitResults

# Files containing our experimental data and structure file
dataFile = "ni-q27r100-neutron.gr"
structureFile = "ni.cif"
spaceGroup = "Fm-3m"

# The first thing to construct is a contribution. Since this is a simple
# example, the contribution will simply contain our PDF data and an associated
# structure file. We'll give it the name "nickel"
niPDF = PDFContribution("nickel")

# Load the data and set the r-range over which we'll fit
niPDF.loadData(dataFile)
niPDF.setCalculationRange(xmin=1, xmax=20, dx=0.01)

# Add the structure from our cif file to the contribution
niStructure = loadStructure(structureFile)
niPDF.addStructure("nickel", niStructure)

# The FitRecipe does the work of calculating the PDF with the fit variable
# that we give it.
niFit = FitRecipe()

# give the PDFContribution to the FitRecipe
niFit.addContribution(niPDF)

# Configure the fit variables and give them to the recipe.  We can use the
```

Line 107, Column 1

untitled • - Sublime Text

File Edit Selection Find View Goto Tools Project Preferences Help

untitled

```
66
67  # Turn off printout of iteration number.
68  niFit.clearFitHooks()
69
70  # We can now execute the fit using scipy's least square optimizer.
71  print "Refine PDF using scipy's least-squares optimizer:"
72  print "  variables:", niFit.names
73  print "  initial values:", niFit.values
74  leastsq(niFit.residual, niFit.values)
75  print "  final values:", niFit.values
76  print
77
78  # Obtain and display the fit results.
79  niResults = FitResults(niFit)
80  print "FIT RESULTS\n"
81  print niResults
82
83  # Plot the observed and refined PDF.
84
85  # Get the experimental data from the recipe
86  r = niFit.nickel.profile.x
87  gobs = niFit.nickel.profile.y
88
89  # Get the calculated PDF and compute the difference between the calculated and
90  # measured PDF
91  gcalc = niFit.nickel.evaluate()
92  baseline = 1.1 * gobs.min()
93  gdiff = gobs - gcalc
94
95  # Plot!
96  pylab.figure()
97  pylab.plot(r, gobs, 'bo', label="G(r) data",
98          markerfacecolor='none', markeredgecolor='b')
99  pylab.plot(r, gcalc, 'r-', label="G(r) fit")
100 pylab.plot(r, gdiff + baseline, 'g-', label="G(r) diff")
101 pylab.plot(r, np.zeros_like(r) + baseline, 'k:')
102 pylab.xlabel(r"r ($\AA$)")
103 pylab.ylabel(r"G ($\AA^{-2}$)")
104 pylab.legend()
105
106 pylab.show()
107
```

Line 107, Column 1

Diffpy project (BNL LDRD)
Complex Modeling infrastructure: Diffpy-CMI

Official release of Diffpy-CMI v2.0 (also Python 3 version available)

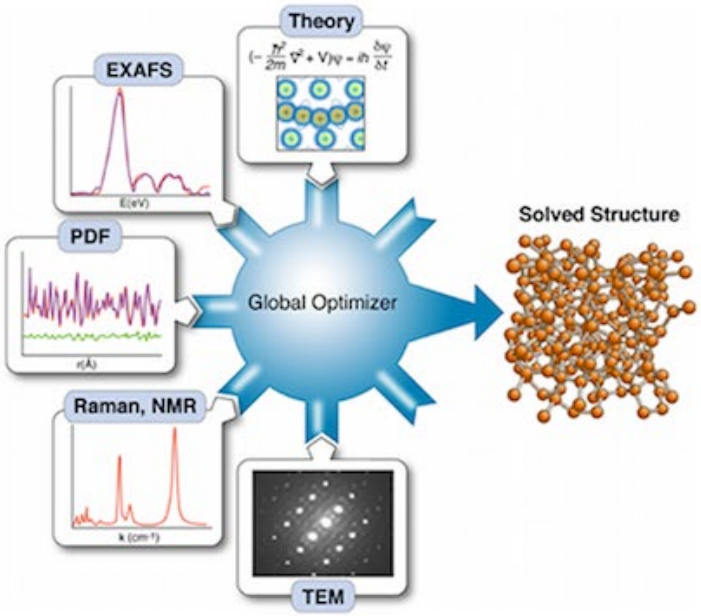
DiffPy Community Publications Products Search

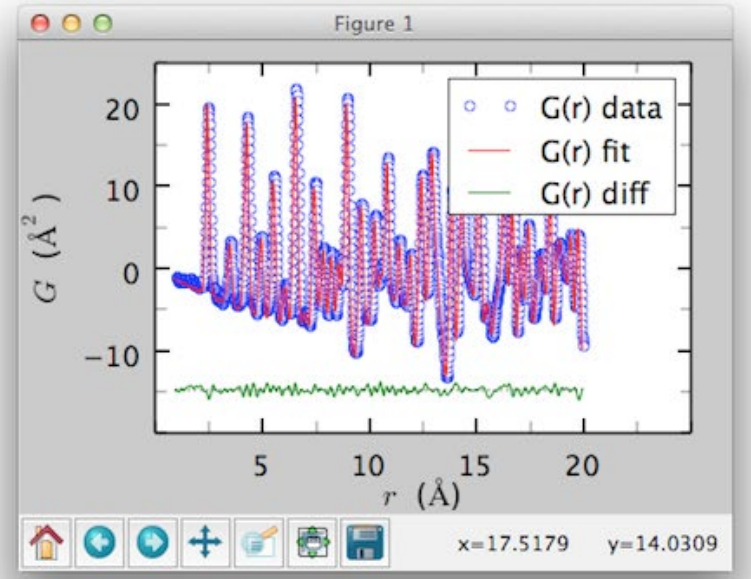
DiffPy-CMI is now available! Get DiffPy-CMI Credits

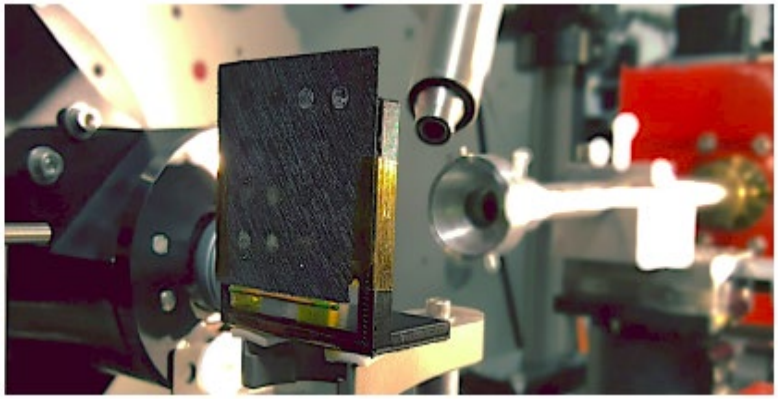
www.diffpy.org

DiffPy - Atomic Structure Analysis in Python

A free and open source software project to provide python software for diffraction analysis and the study of the atomic structure of materials.

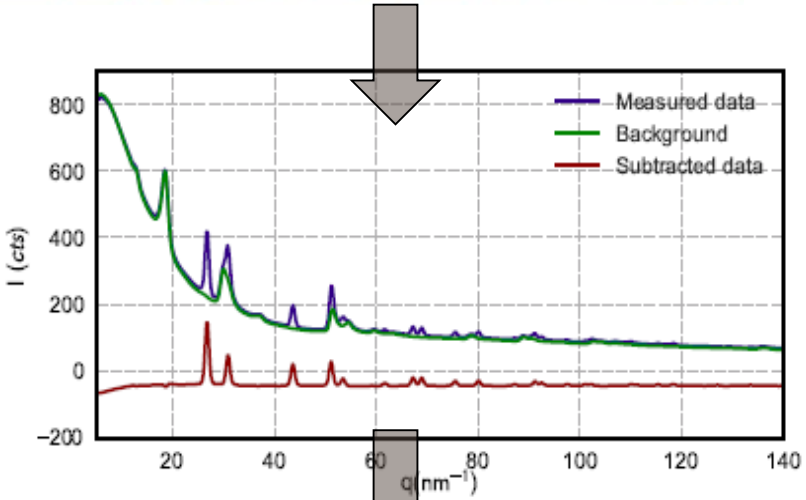


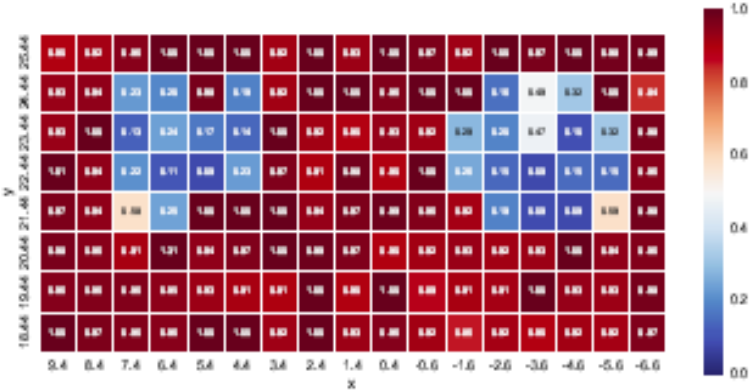


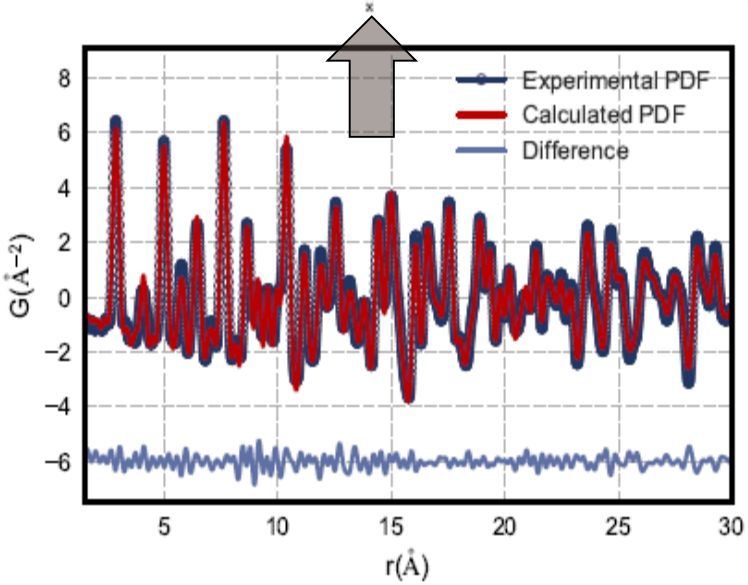


=> Spatially Resolved PDFs

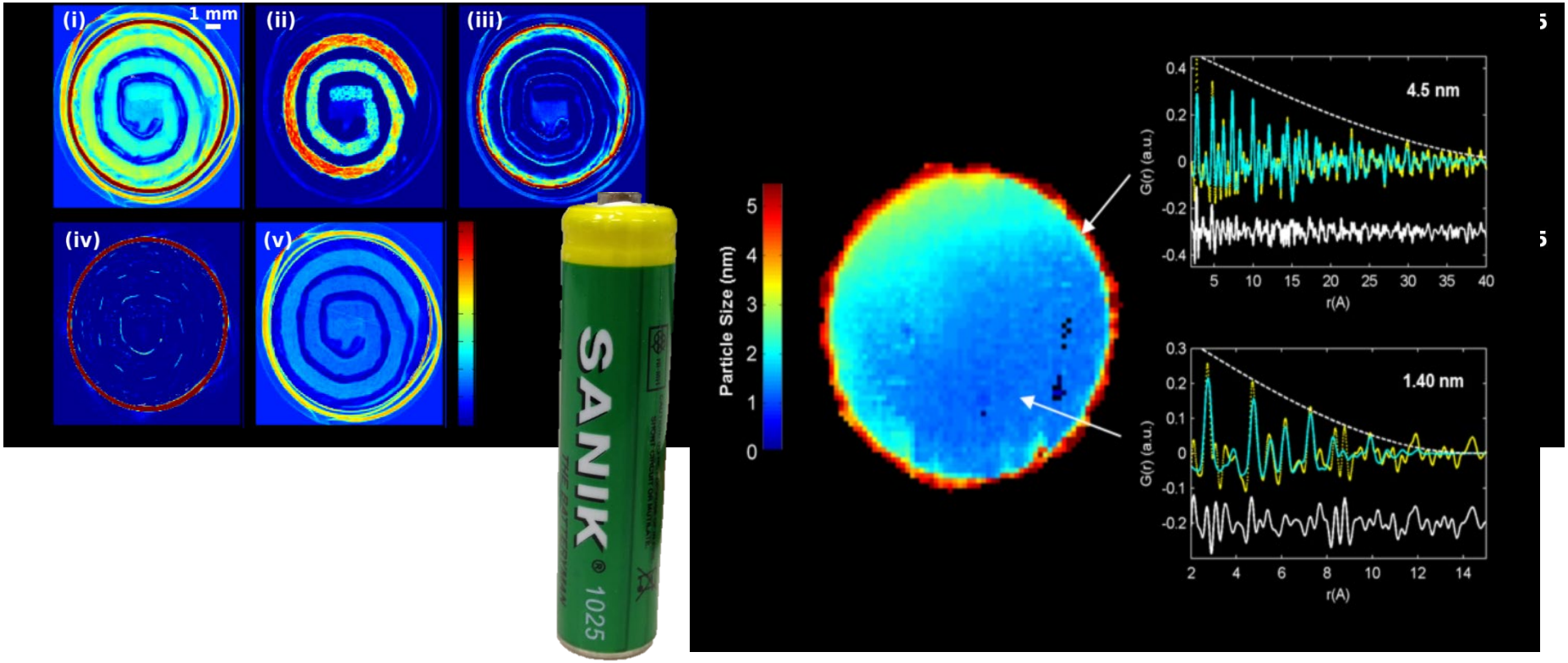
- Anton Kovyakh, Soham Banerjee, Chia Hao Liu







=> Combine PDF and tomography (ctPDF)



ctPDF developed with Simon Jacques, Marco DiMichiel, Andy Beal and Bob Cernik

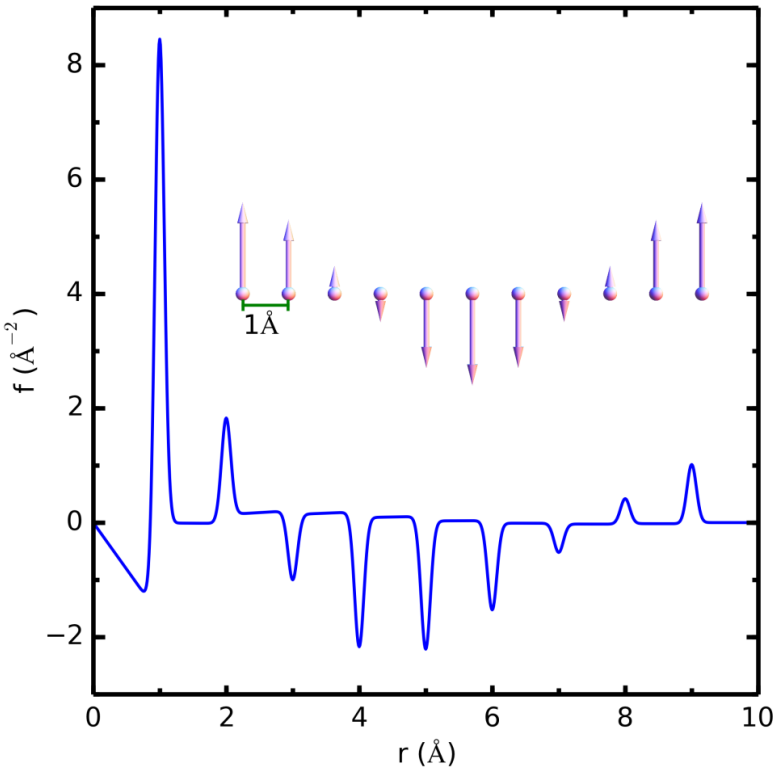
10,000 2D datasets per image, 30 mins per image ~10Tb/day

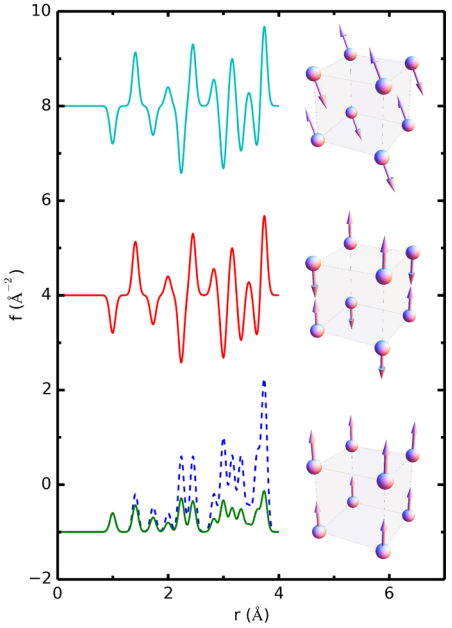
Simon D. M. Jacques, Marco Di Michiel, Simon A. J. Kimber, Xiaohao Yang, Robert J. Cernik, Andrew M. Beale and Simon J. L. Billinge, *Nat. Commun.* **4**, 2536 (2013).

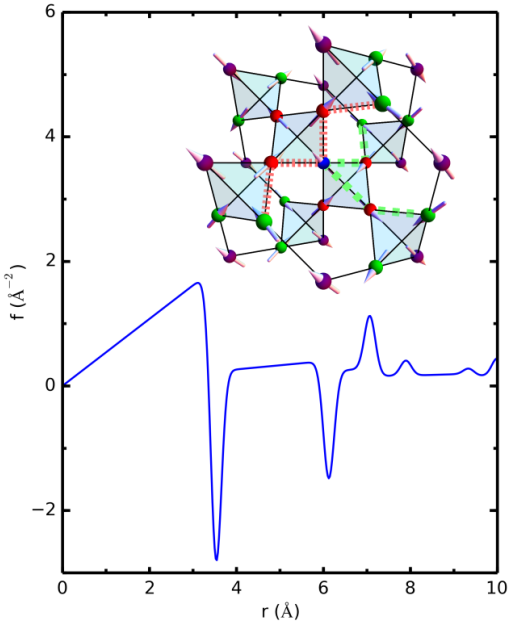
Jensen, Corr, Di Michiel, SJLB et al., *J. Electrochem. Soc.* (2015)

http://thebillingegroup.com

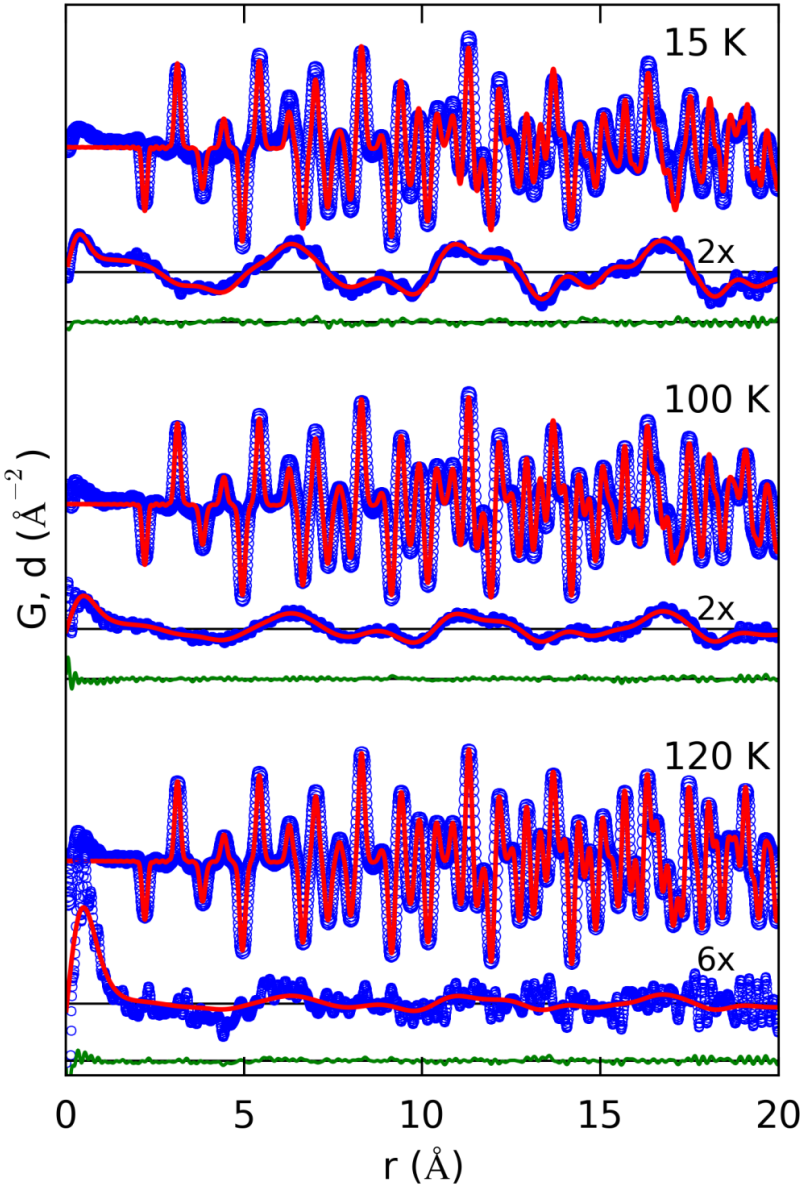
mPDF: PDF of short-range magnetic correlations







- Benjamin A. Frandsen, Xiaohao Yang and Simon J.L. Billinge, *Acta Crystallogr. A* **70**, 3-11 (2014).



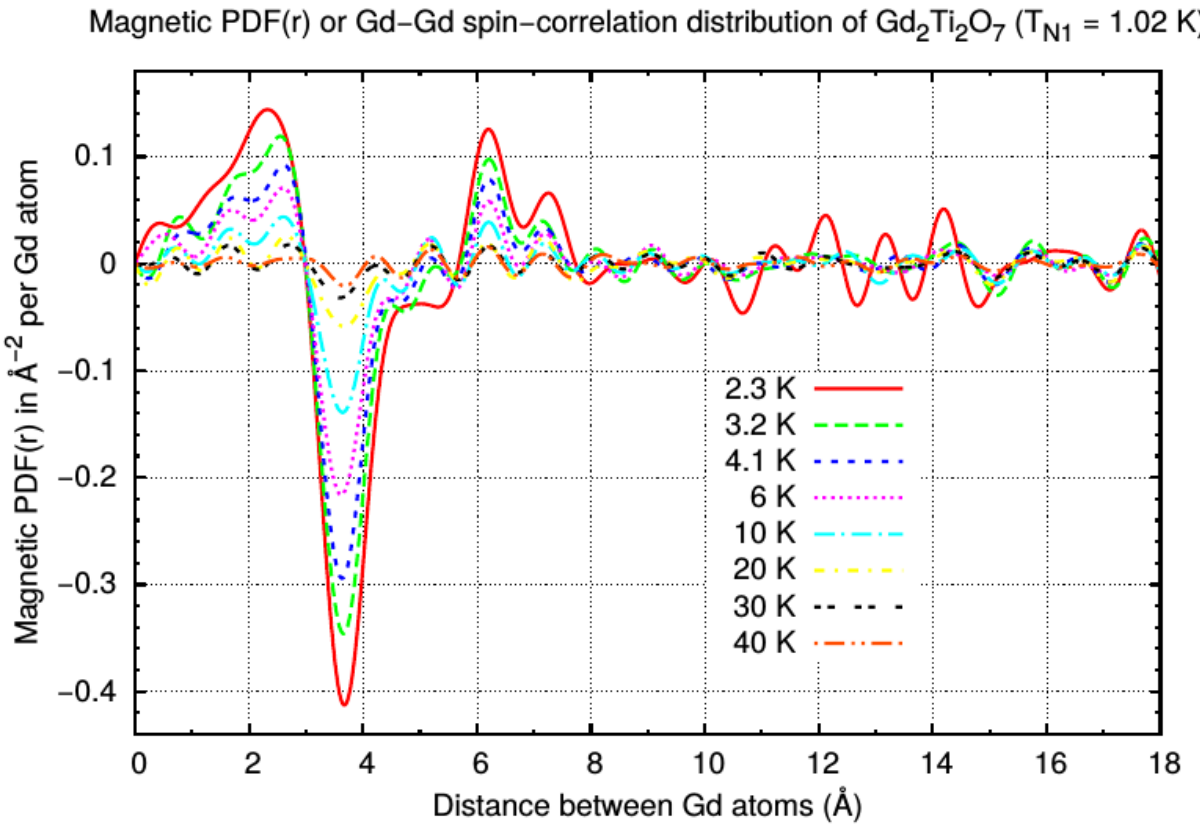
MnO

Top: *Blue*: measured nPDF signal
Red (top) calculated structural PDF
Btm: *Blue*: difference between nPDFsignal and calculated structural PDF
Red: calculated mPDF for AF MnO

Benjamin A. Frandsen, M. Brunelli, K. Page, Y. J. Uemura, Julie B. Staunton, SJLB, arXiv:1512.06270.

Frandsen, SJLB, Acta Crystallogr. A 71, 325-334 (2015)

Intuitive way to resolve interesting questions

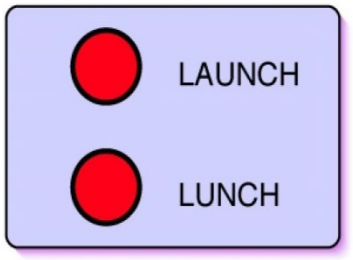


- Henry Fischer poster!

Software Projects

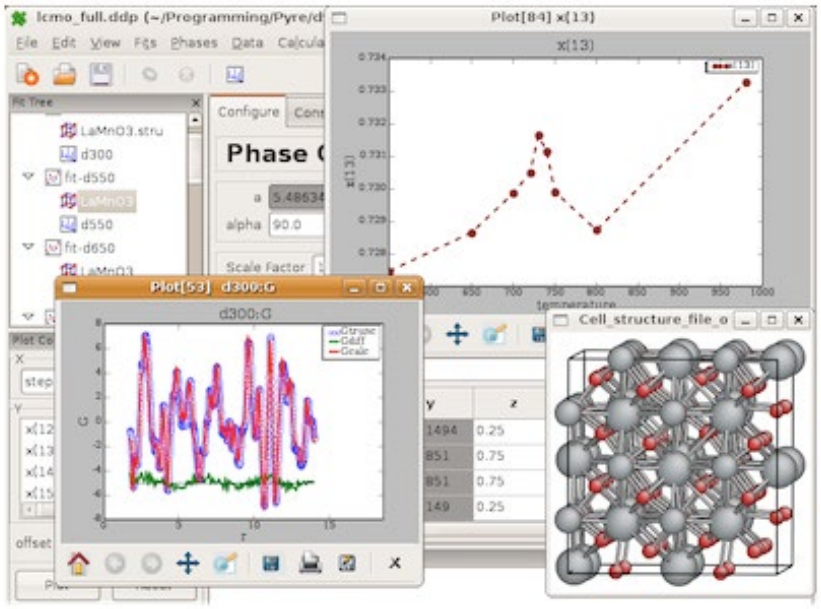
Software Projects that go wrong

- Who: US department of homeland security
- What: Develop a GUI for the US president to undertake his most important functions
- Functional Requirements: Must be very simple and easy to use
- Solution:



PDFgui2.0

- PDFFIT2 + gui = PDFgui



- Diffpy-CMI + gui = PDFgui2.0

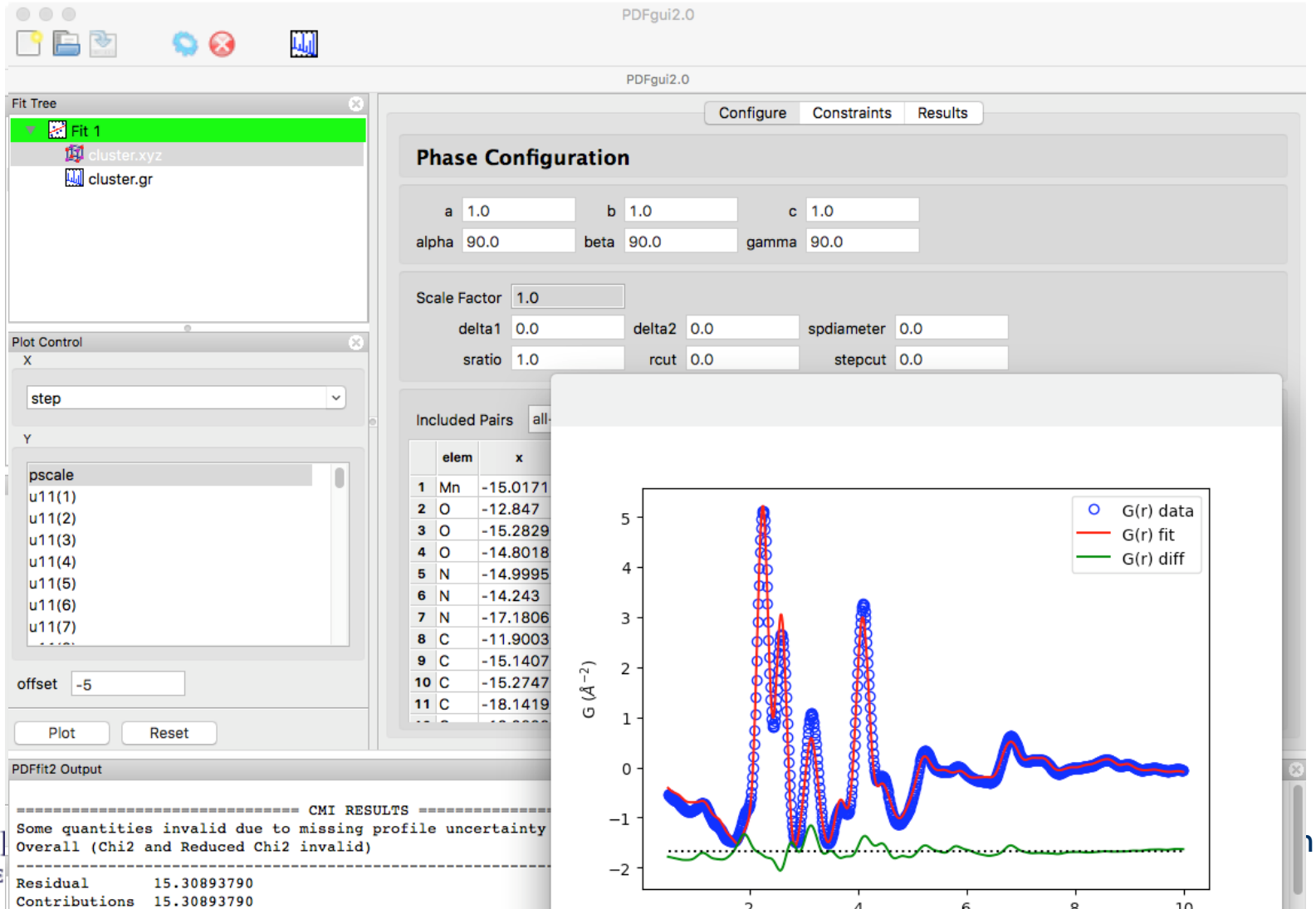
PDFgui2.0 is indistinguishable from PDFgui

PDFgui

C. L. Farrow, P. Juhás, J. W. Liu, D. Bryndin, E. S. Božin, J. Bloch, Th. Proffen and S. J. L. Billinge, PDFfit2 and PDFgui: computer programs for studying nanostructures in crystals, *J. Phys.: Condens. Matter* **19**, 335219 (2007)

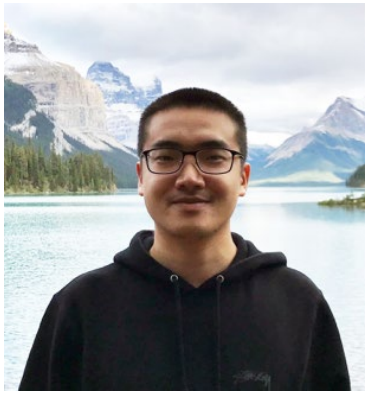
But PDFgui2.1….

- Cluster modeling in a gui……



PDFgui2.X

- PDFgui2.2: Log normal particle distributions
- PDFgui2.3: mPDF integration
- PDFgui2.4: organics
- PDFgui2.5: Your dream here!



Status: Work of Long Yang, collaboration with Matt Tucker @ORNL. We have working prototypes of PDFgui2.0 and PDFgui2.1

It is an open-source community coding project. If you are interested in contributing, please let us know!

High throughput + penetrating x-rays + dilute systems => in-situ synthesis

- We can see precursor species in solution
- We can measure Nanoparticle structural parameters
- High energy x-rays can penetrate relatively thick walled vessels

=> Let's do in-situ studies of synthesis

- Rich collaboration with the group of Bo Iversen (Aarhus)

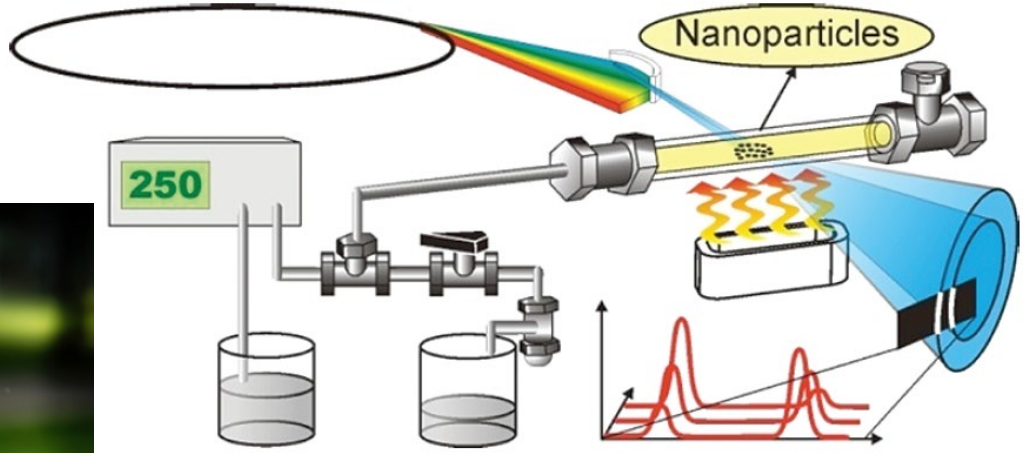
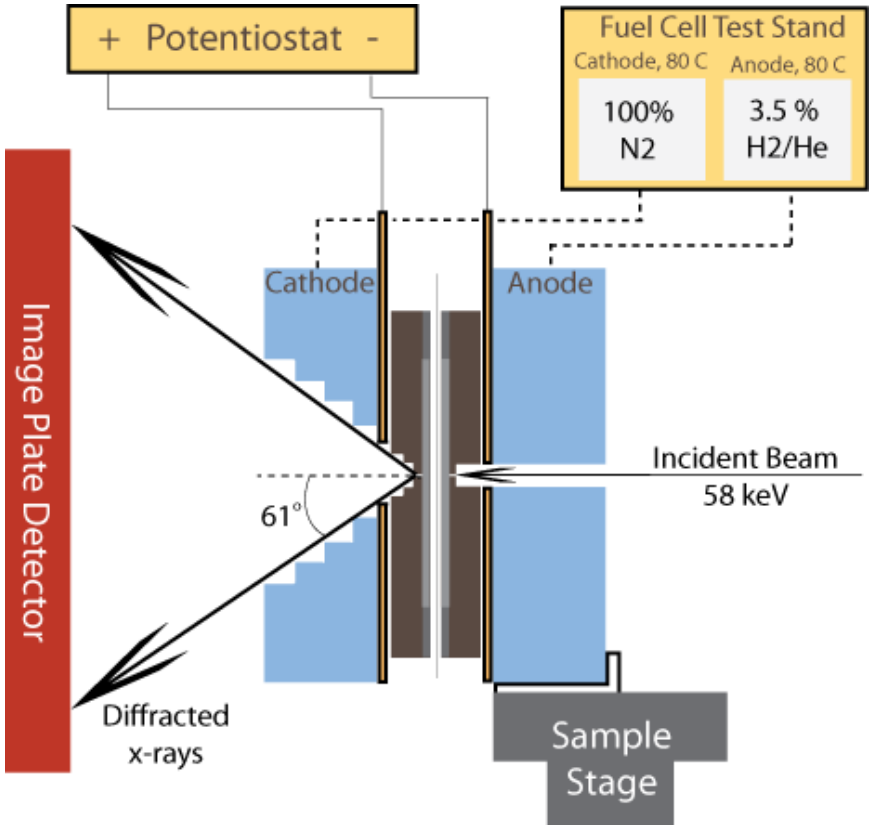
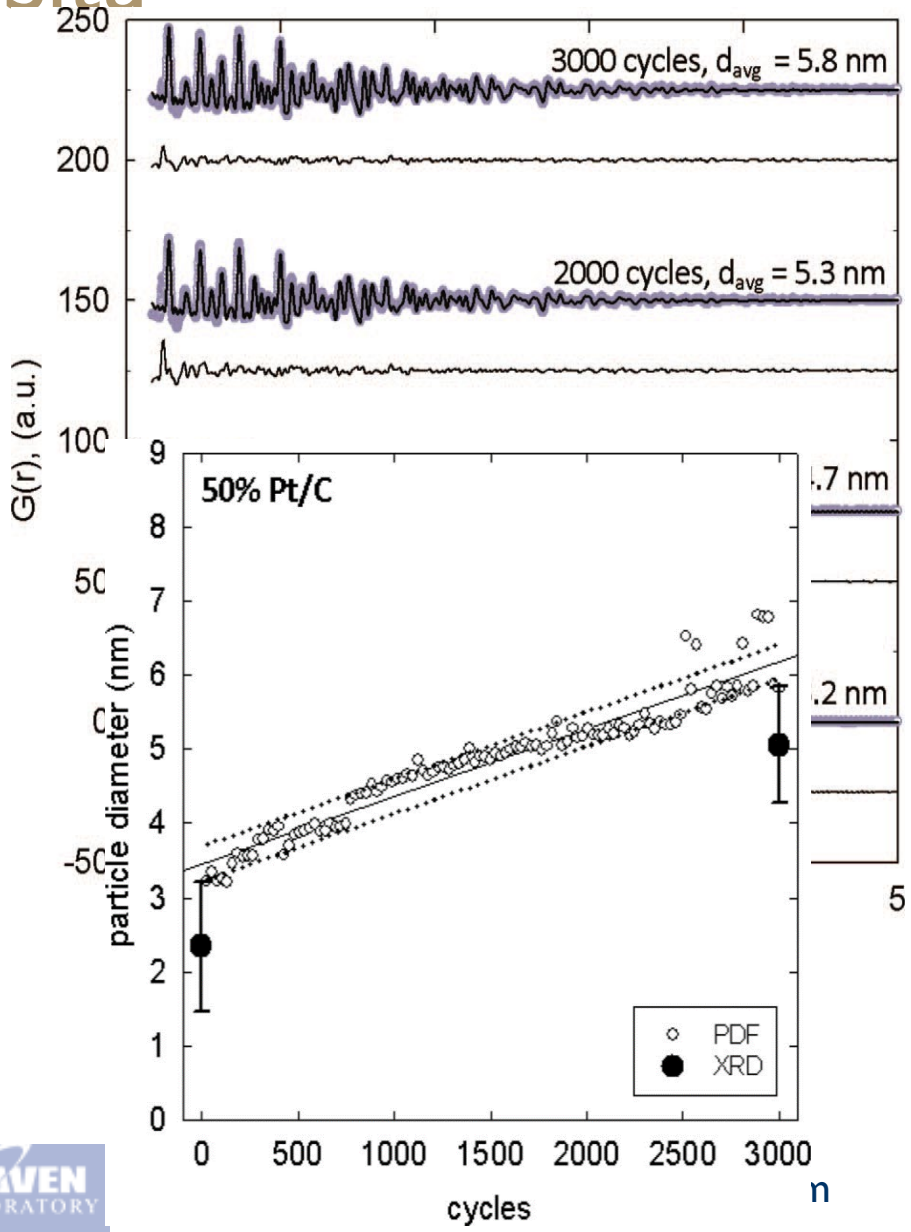


Image credit
Christoffer Tyrsted

lingegroup.com

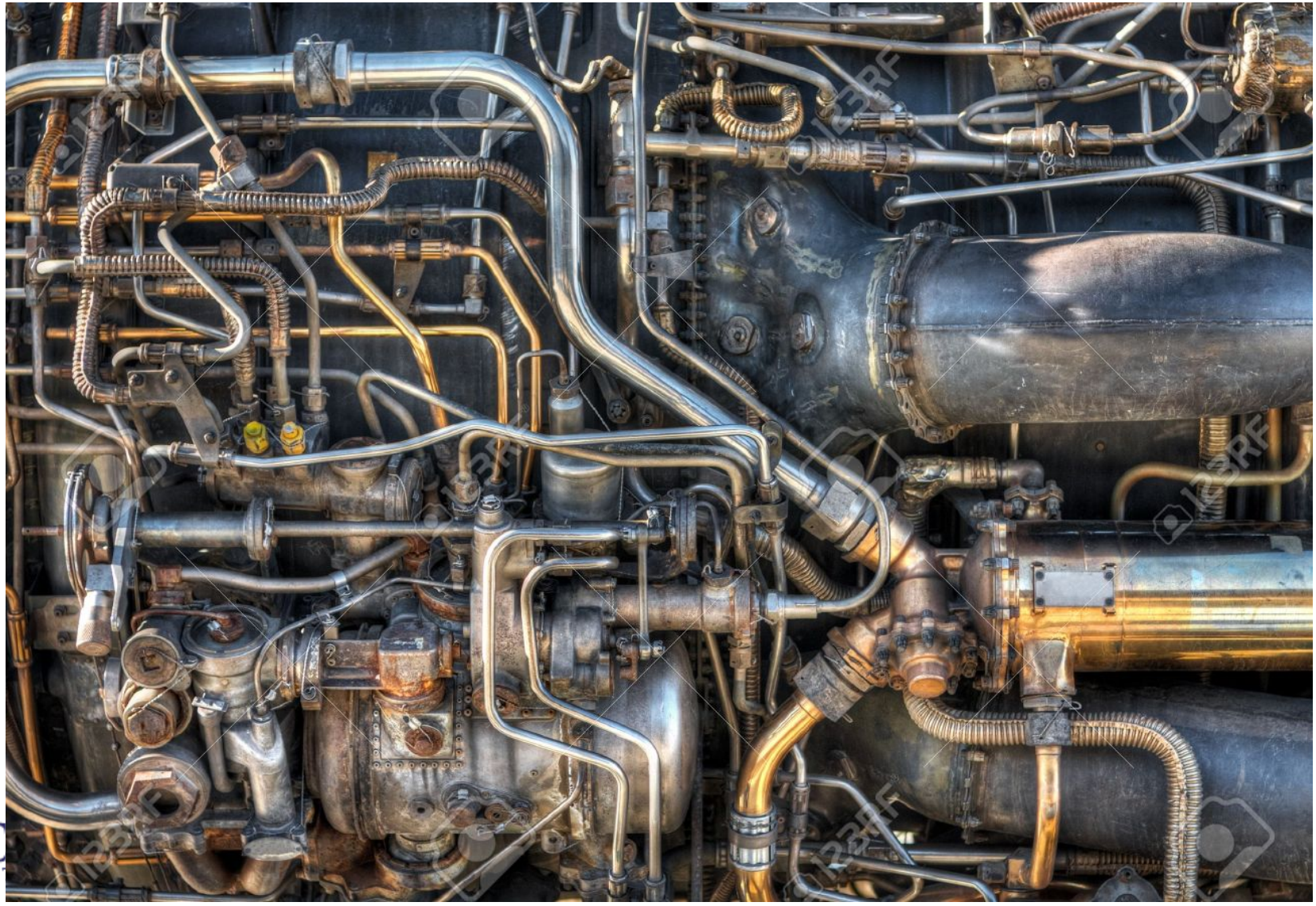
High-throughput, penetrating x-rays => In Situ



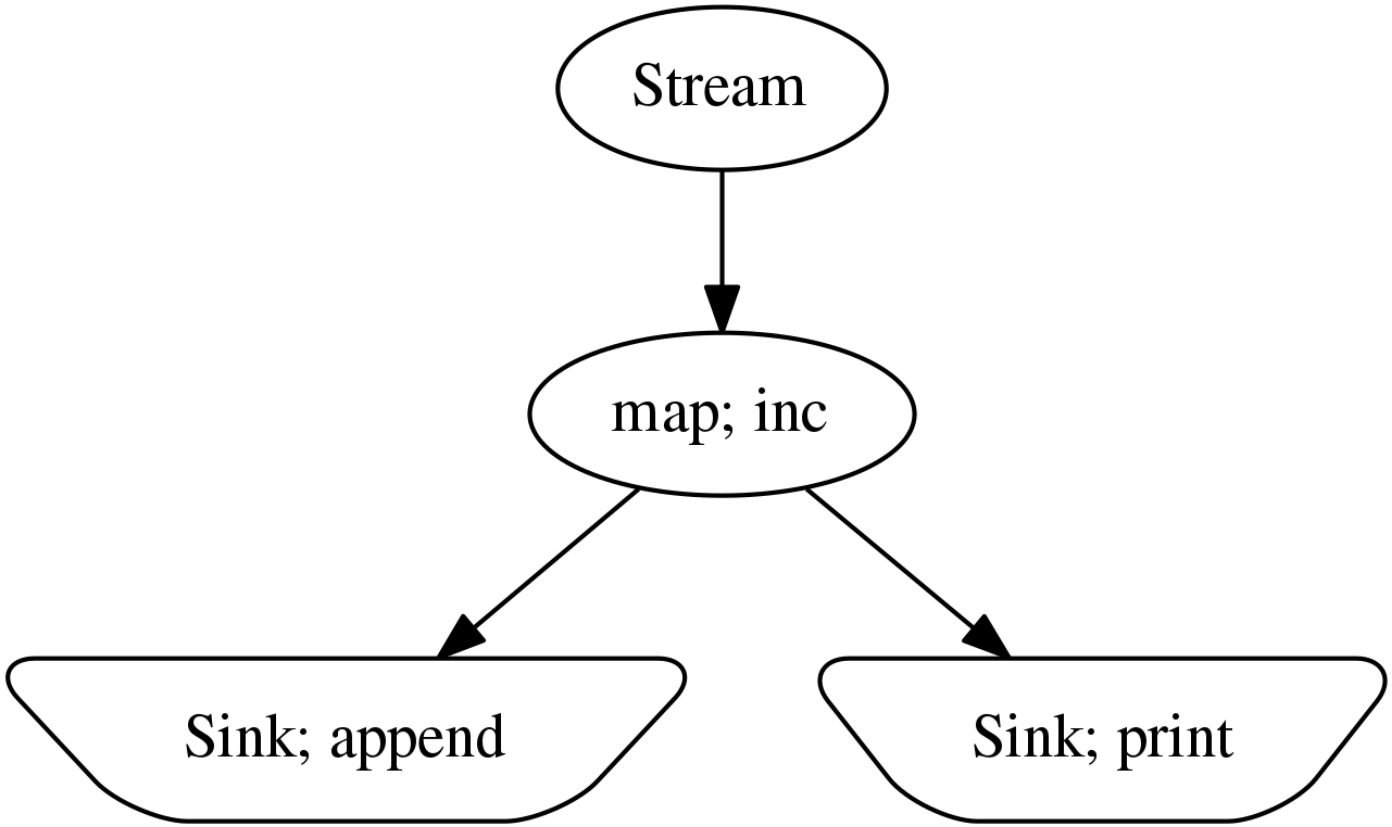


- Redmond, et al., *Electrochem. Solid St., 15* (5), B72–B74.

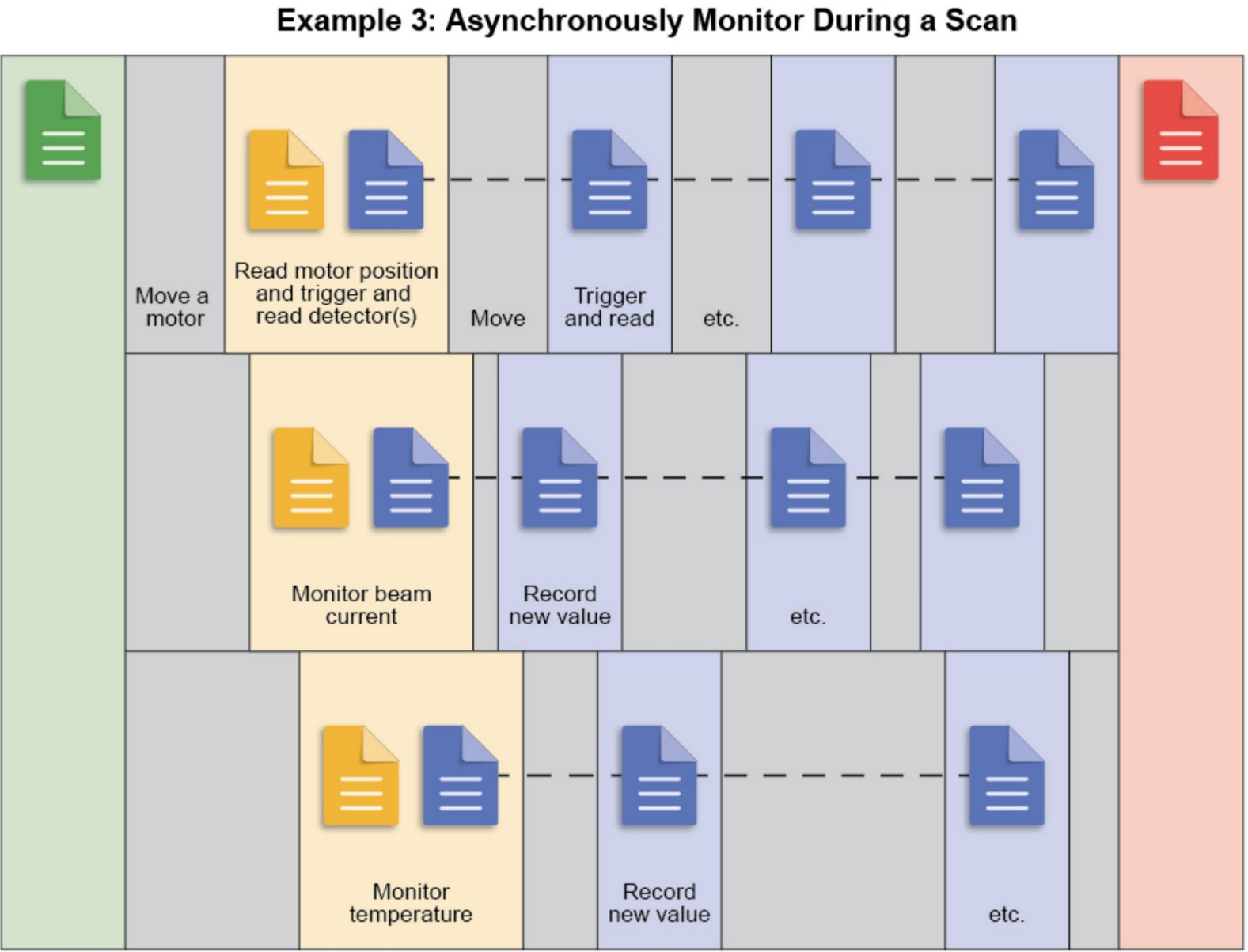
Data analysis/modeling as pipes



Data as streams through those pipes



Experiments as sets of asynchronous streams

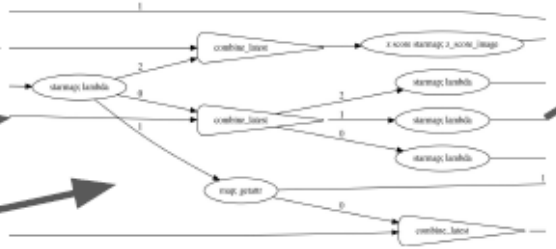


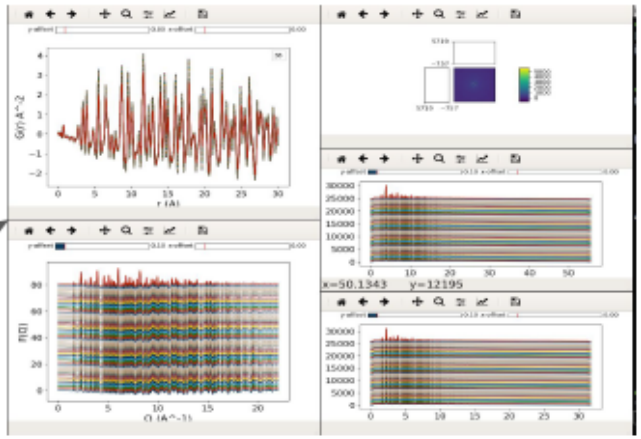
- DAMA group @BNL

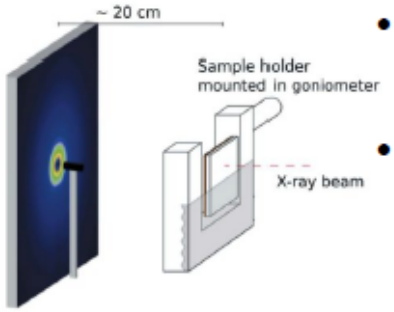
group.com

xpdAn/xpdtools

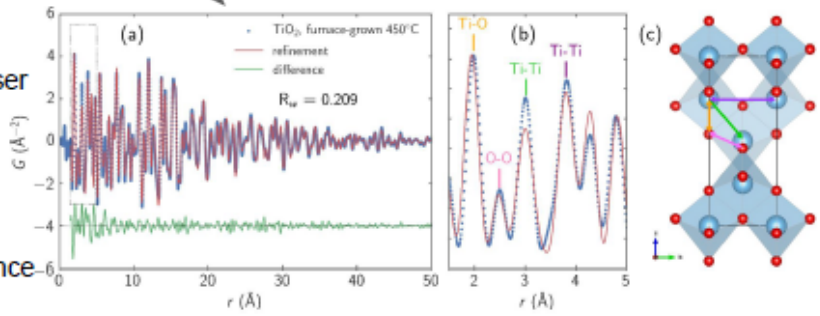
Data and metadata capture, combined with streaming data analysis pipelines, produces live data visualization and structure refinement.







- High quality data and metadata, are required for data mining, machine learning and automation
- The software stack is designed to be
  - Easy to use and automated
  - Capture rich and accurate metadata with minimal user overhead
- Software stack (GitHub and conda) includes:
  - **xpdAcq** for data/metadata acquisition
  - **xpdAn** for live data analysis
  - **xpdtools** for analysis tools
  - **SHED** for streaming the data analysis with provenance



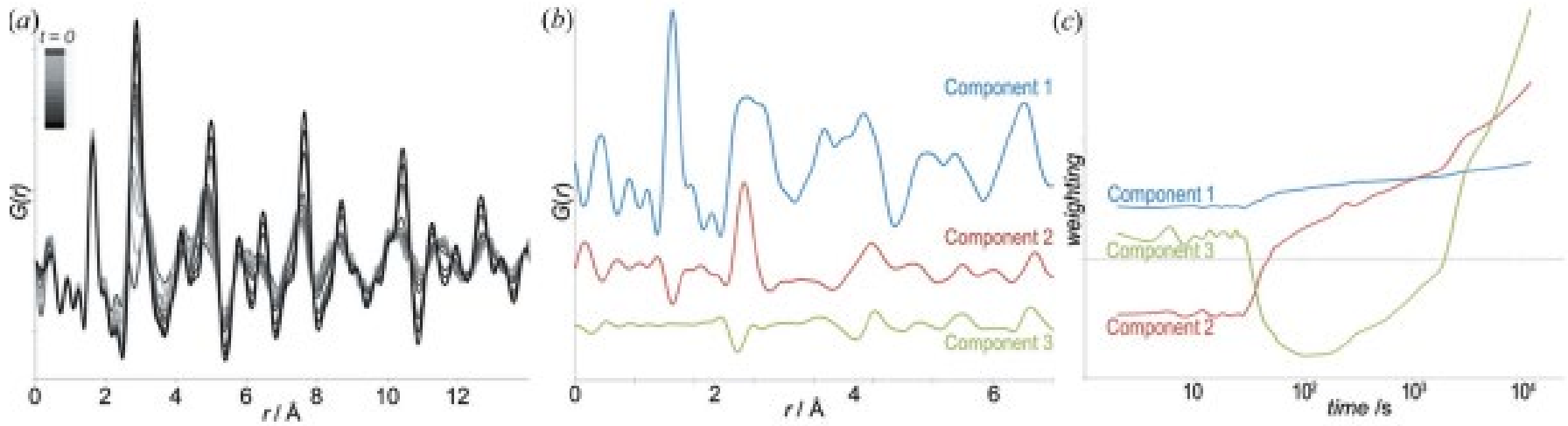
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP://thebillingegroup.com

Monitor data as it comes

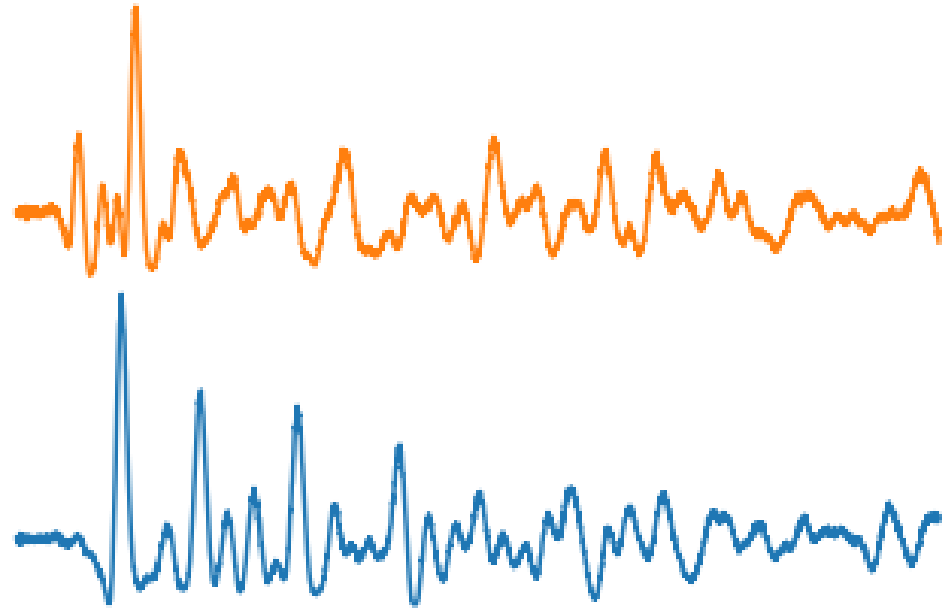
- Define Quantities of Interest (QOI's)
- Monitor the QOI's
- E.g., projection coefficients onto principle components

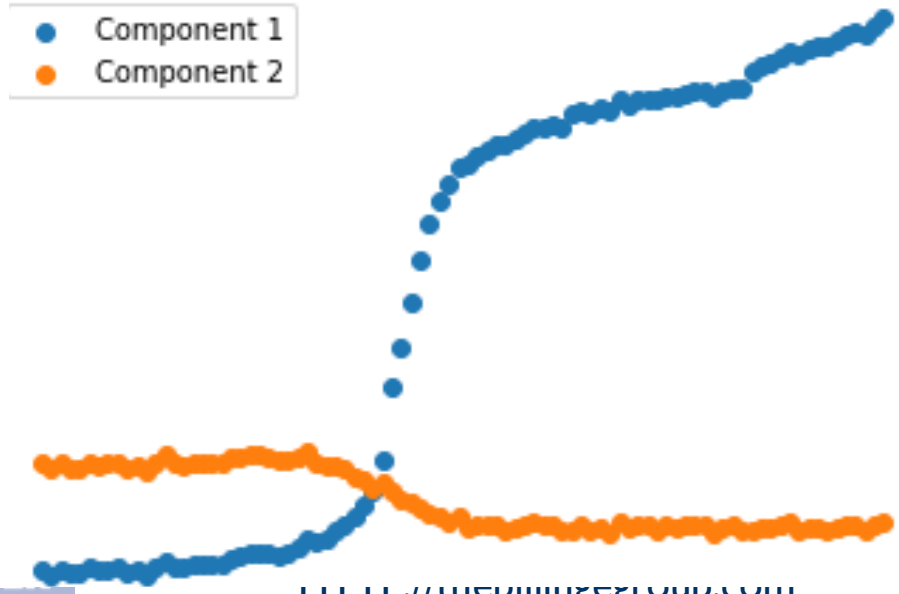


- Chapman et al., JAC 2015, PCA on gold nanoparticle formation

Other physical measures

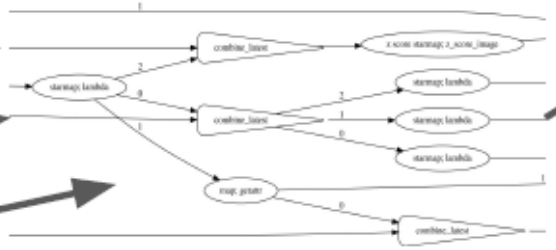
- It is straightforward to incorporate other statistical measures into the pipelines.
- E.g., if you want a more physical matrix decomposition, use Non-negative matrix factorization (NMF)
- Slower and less convergent, but more physical

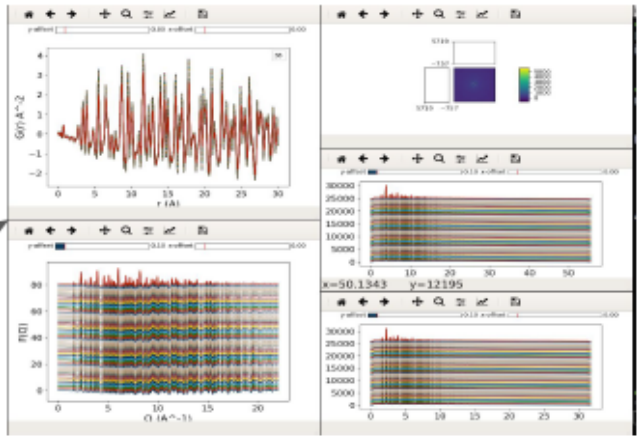




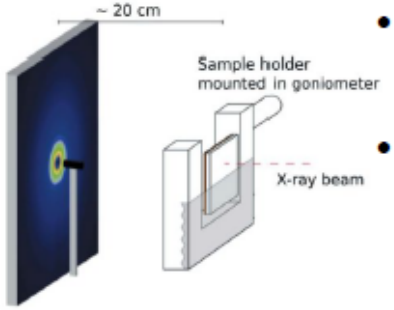
xpdtools

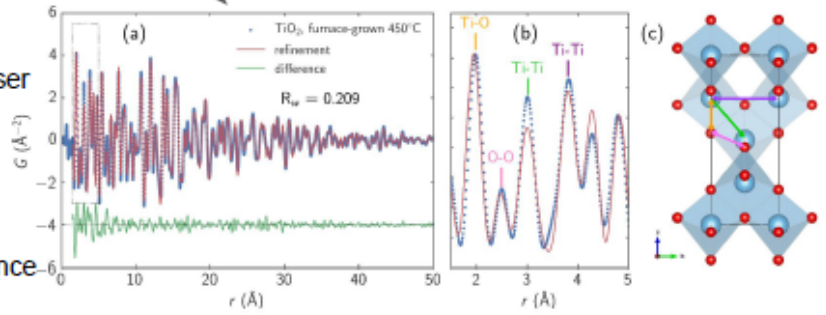
Data and metadata capture, combined with streaming data analysis pipelines, produces live data visualization and structure refinement.





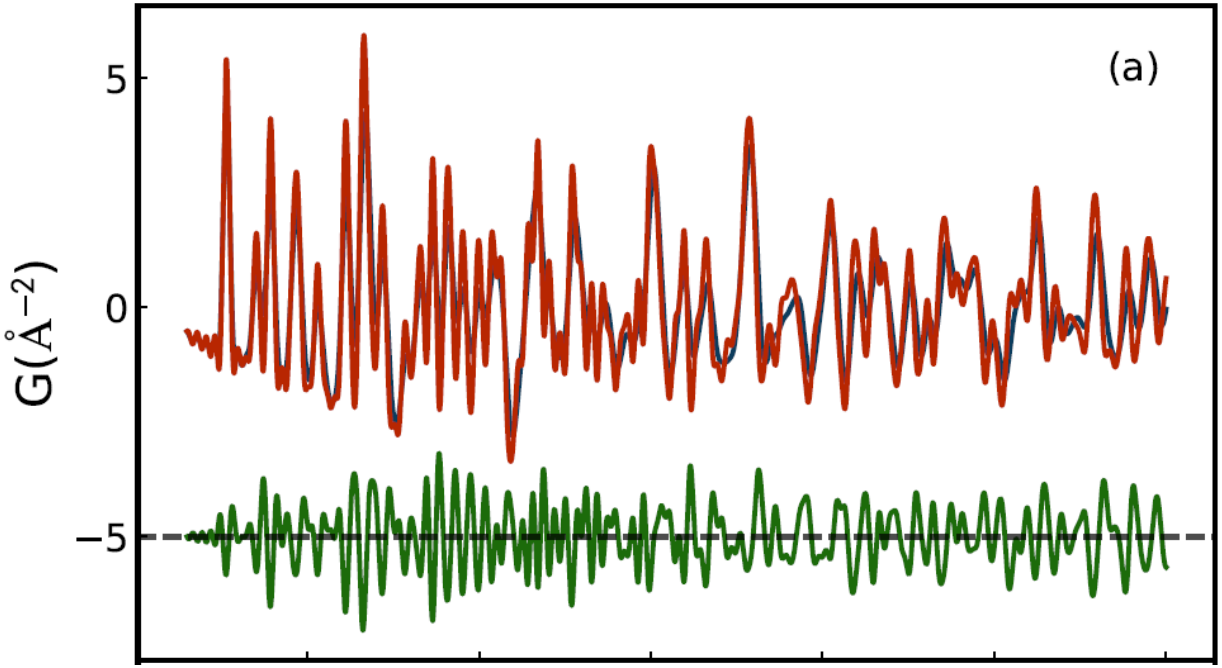
- High quality data and metadata, are required for data mining, machine learning and automation
- The software stack is designed to be
  - Easy to use and automated
  - Capture rich and accurate metadata with minimal user overhead
- Software stack (GitHub and conda) includes:
  - **xpdAcq** for data/metadata acquisition
  - **xpdAn** for live data analysis
  - **xpdtools** for analysis tools
  - **SHED** for streaming the data analysis with provenance





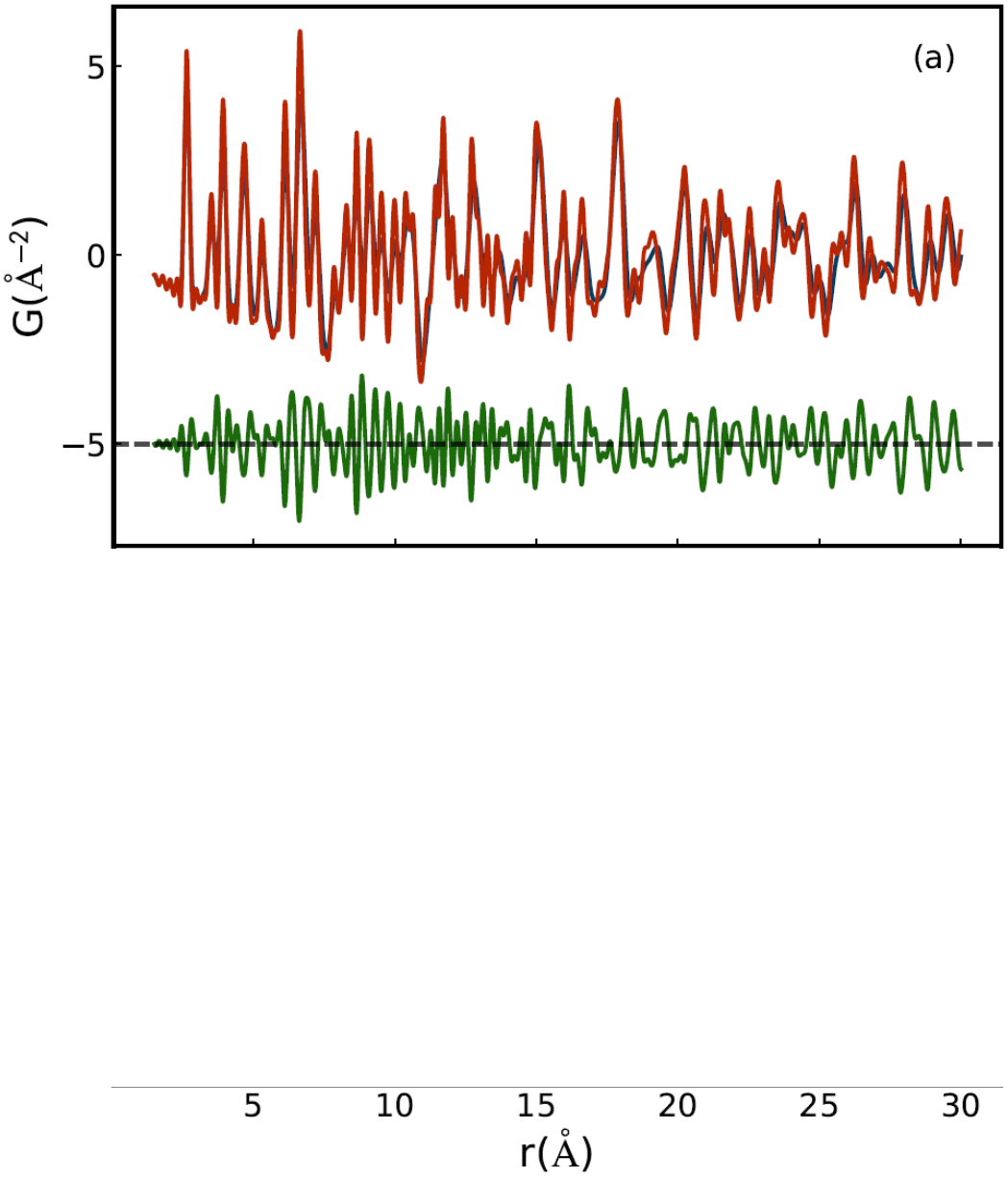
PDFmorph

- Has my sample undergone a phase transition?
- Plot high and low-T data on top of each other, are they different?



- Yes, um no, um, not sure

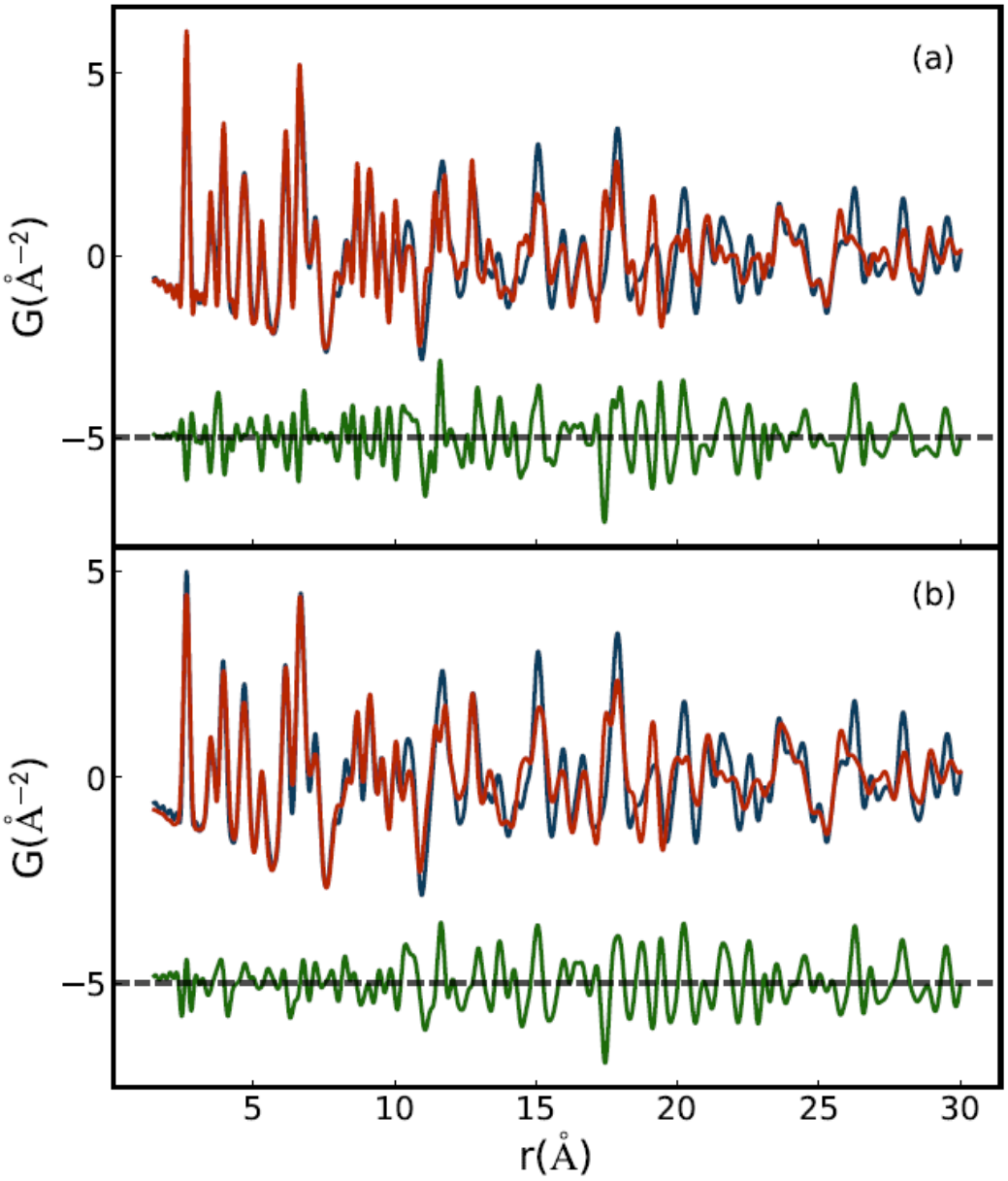
PDFmorph



- Apply reasonable transformations to the data in a regression loop:
  - Stretch
  - Smear
  - Scale
- Low-cost model independent way to account for "uninteresting" effects such as thermal expansion and thermal broadening

HTTP://thebillingegroup.com

PDFmorph



- If there is a phase transformation, stretching, smearing and scaling won't cut it!
- Been using PDFmorph in the group for some time but we are getting close to a public release

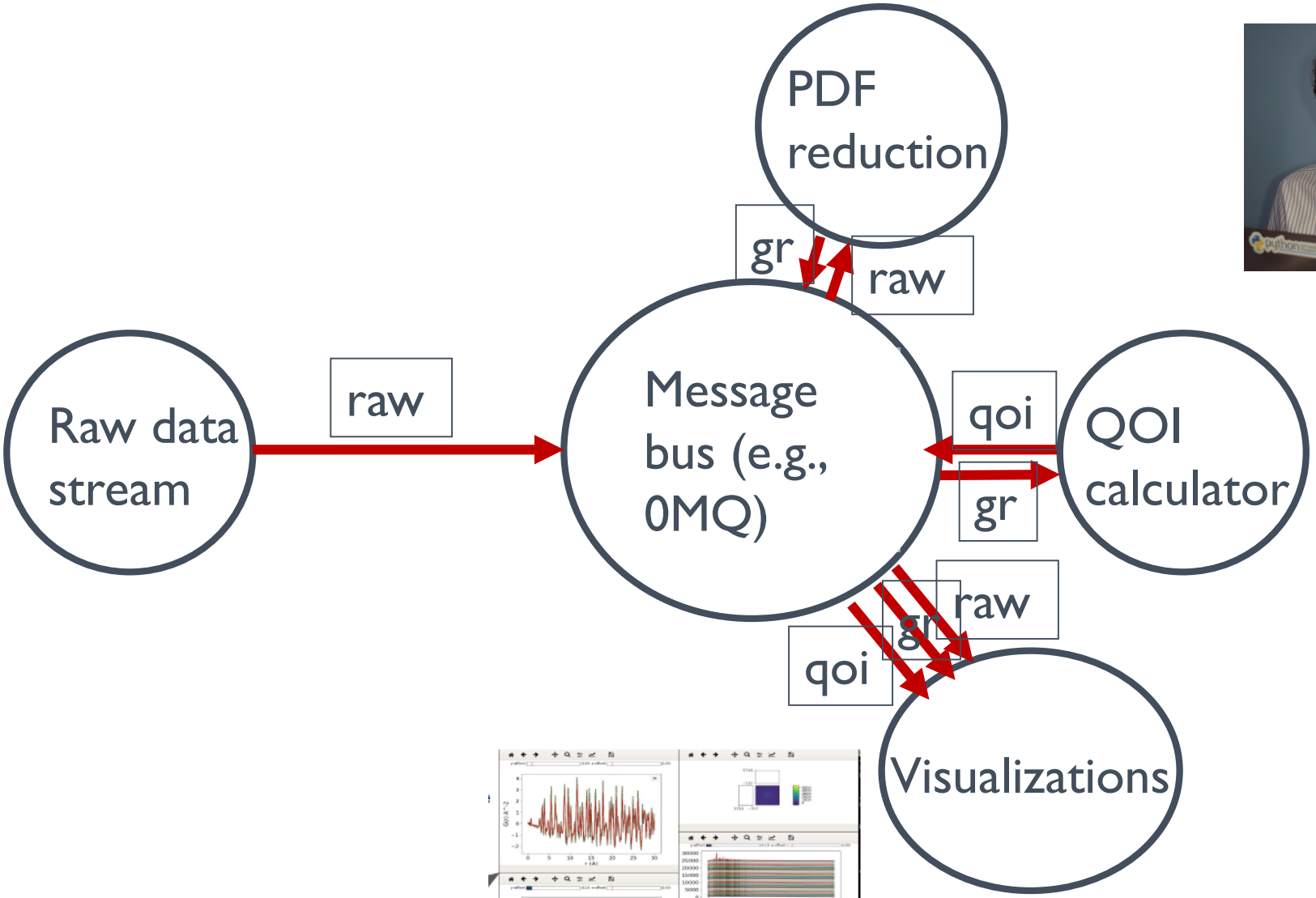
HTTP://thebillingegroup.com

Put PDFmorph into the pipeline

1. Define something as a scalar "feature"
    1. E.g., the Rw or the Pearson correlation coefficient of the current PDF morphed onto either the previous or the first PDF in the stream
2. Incorporate this as a node in the data analysis pipeline graph
3. Visualize the QOI vs. index

Pipelines can be run at data collection time (xpdAn@NSLSII) but using xpdtools, it can be run on data that you take home from any experiment.

Computational infrastructure

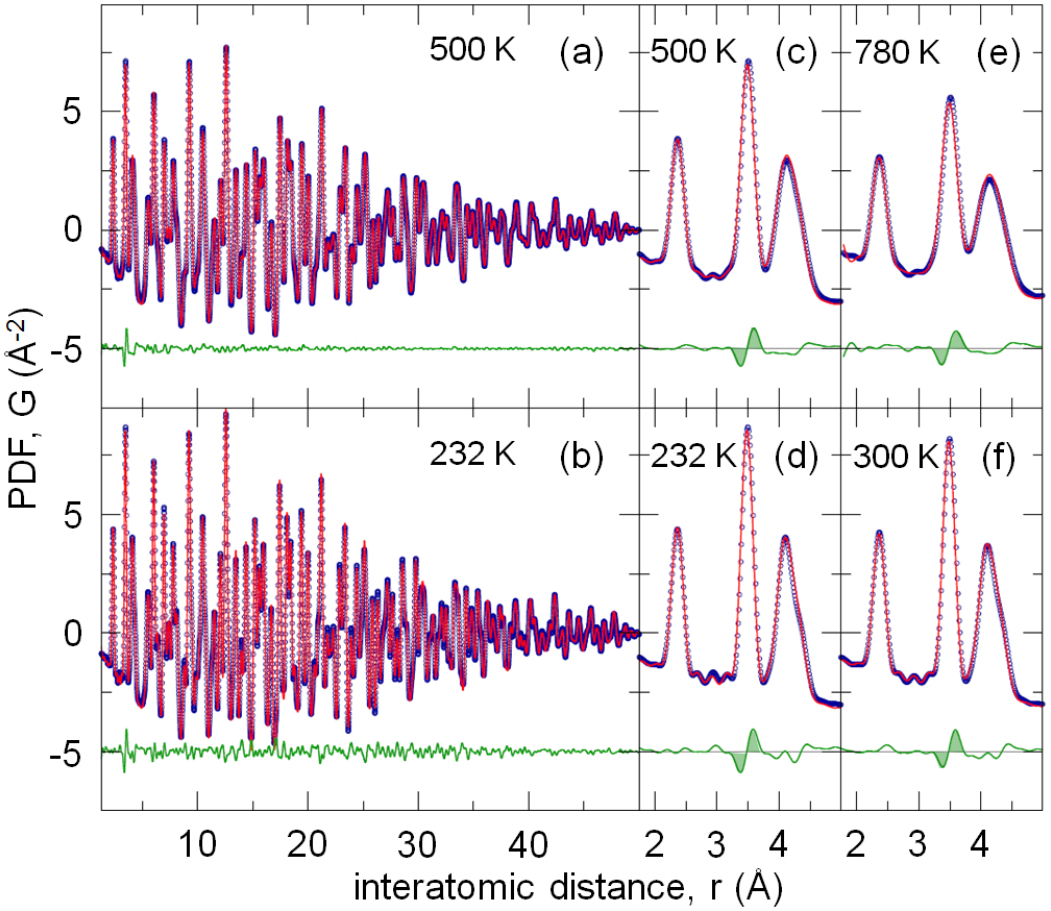


Automating modeling

- Can we think about how to make modeling more automated, and possibly introducing it into the pipeline?

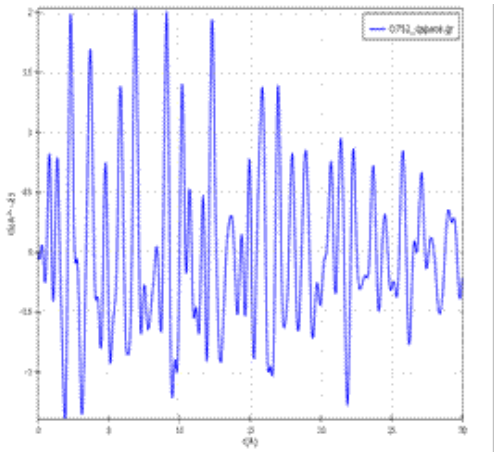
Standard approach

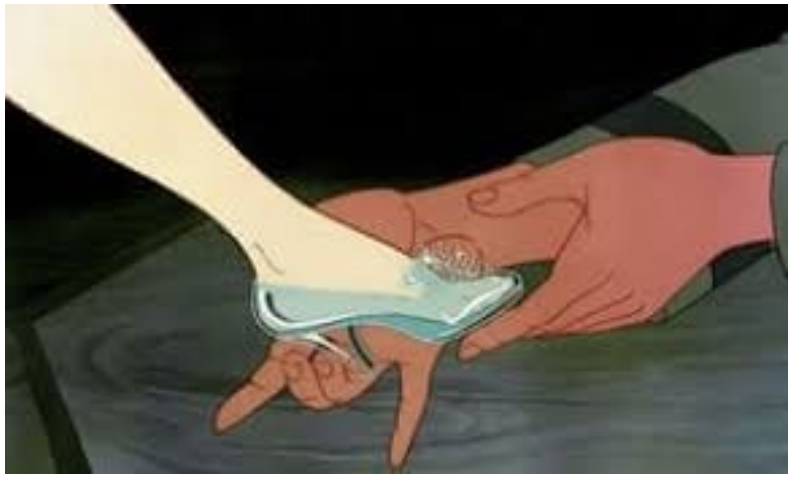
- 1 PDF
- 1 model
- 20 parameters
- Vary the parameters until the model agrees as well as possible with the data.
- Emphasis on parameter estimation, not on model selection
- Challenge is finding the right model



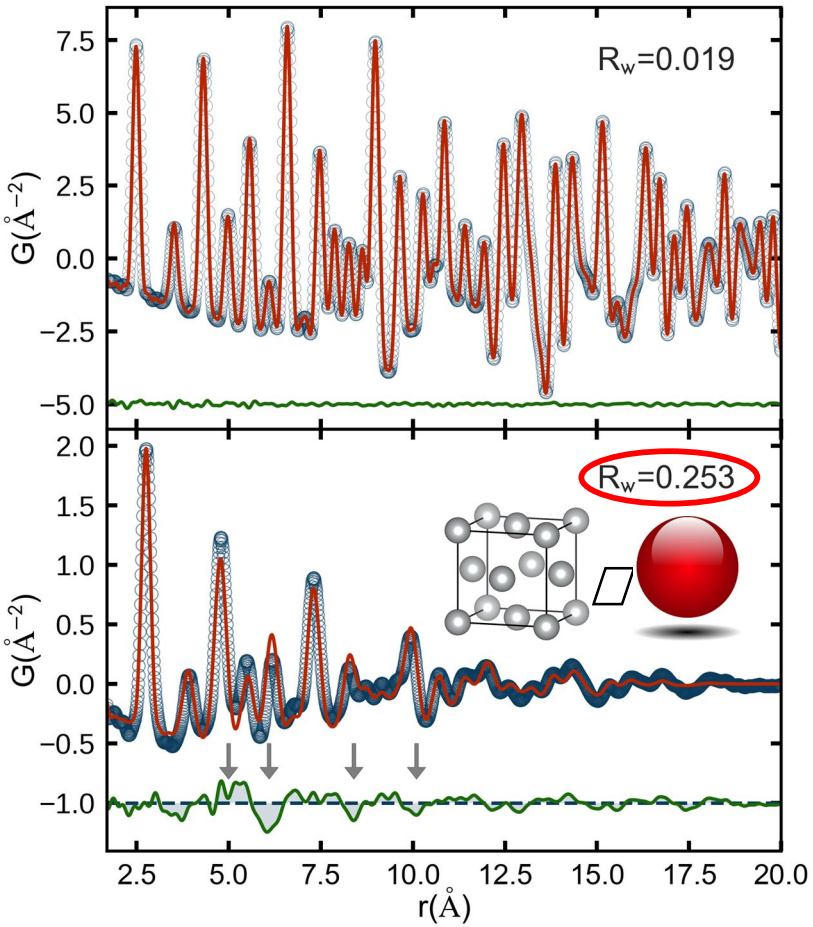
New approach: Structure-Mining

- 1 PDF
- many models
- few parameters
- Iterate over large numbers of models
- Emphasis on model selection
- Advantage: find multiple nearby models!
- Challenge: structure must be in the structure-mine

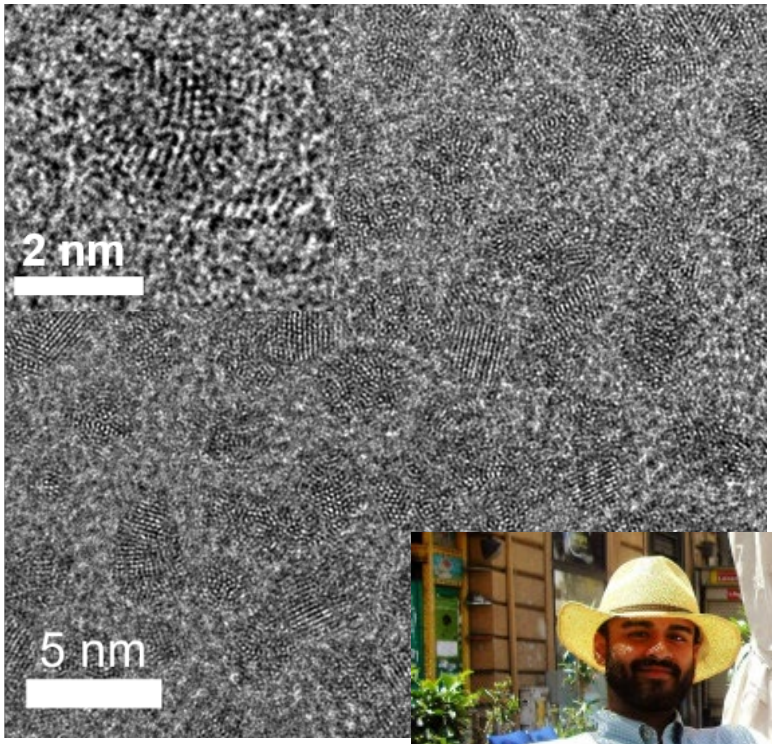




Structure-Mining 1: Metallic nanoparticles



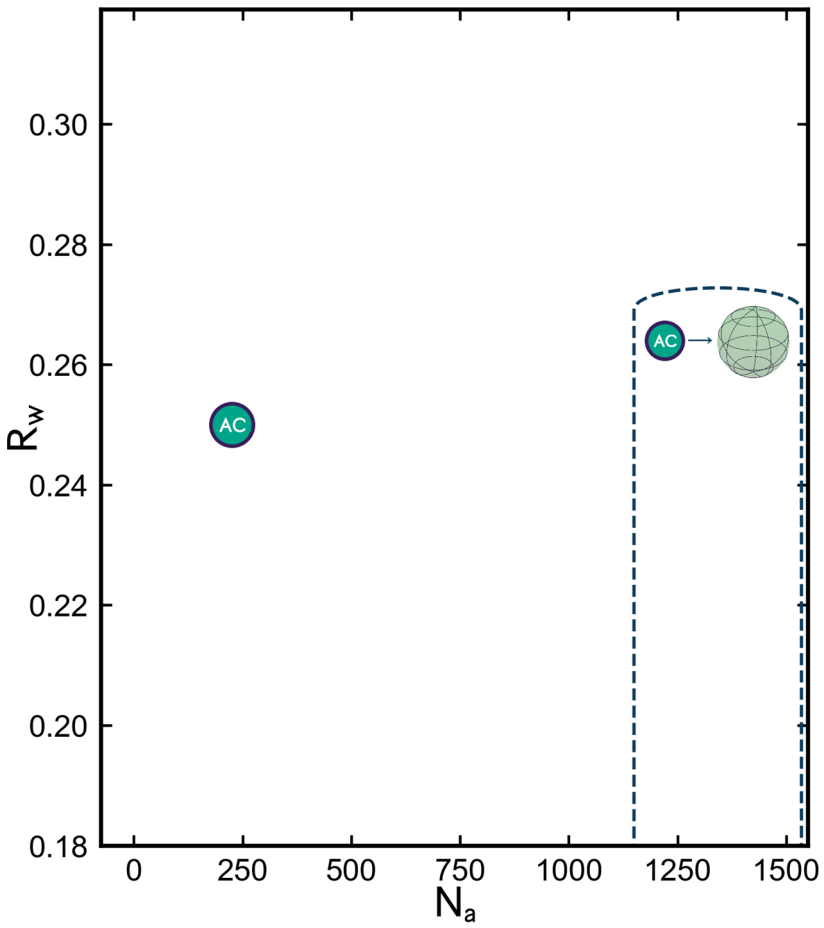
~3 nm Pd nanoparticles

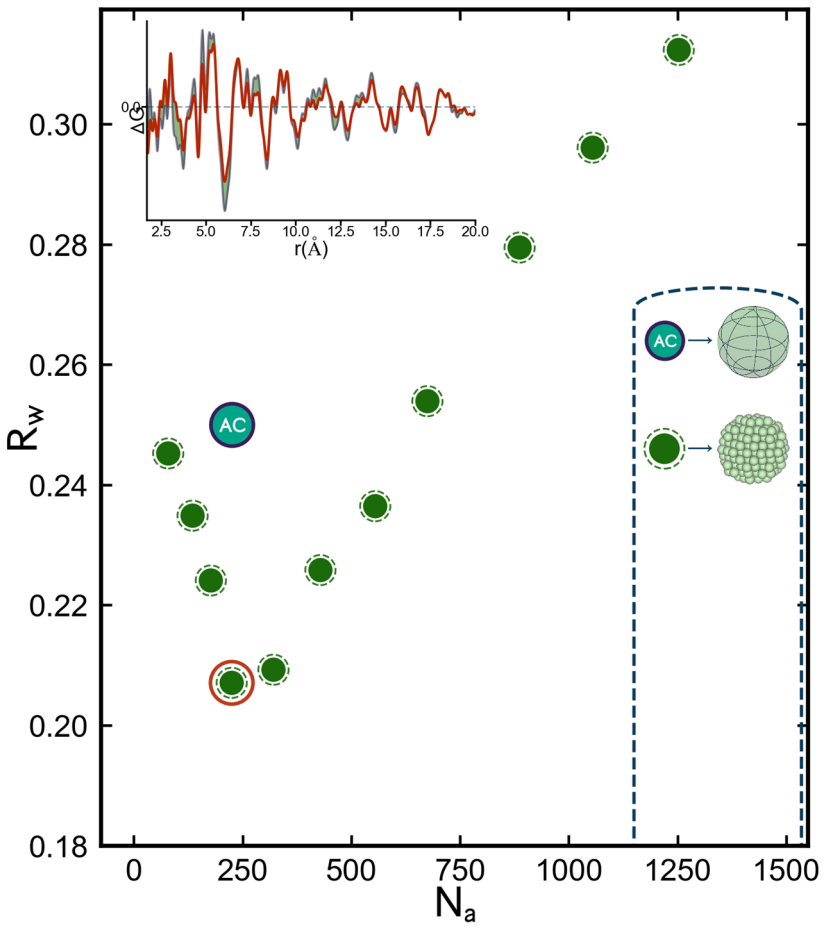


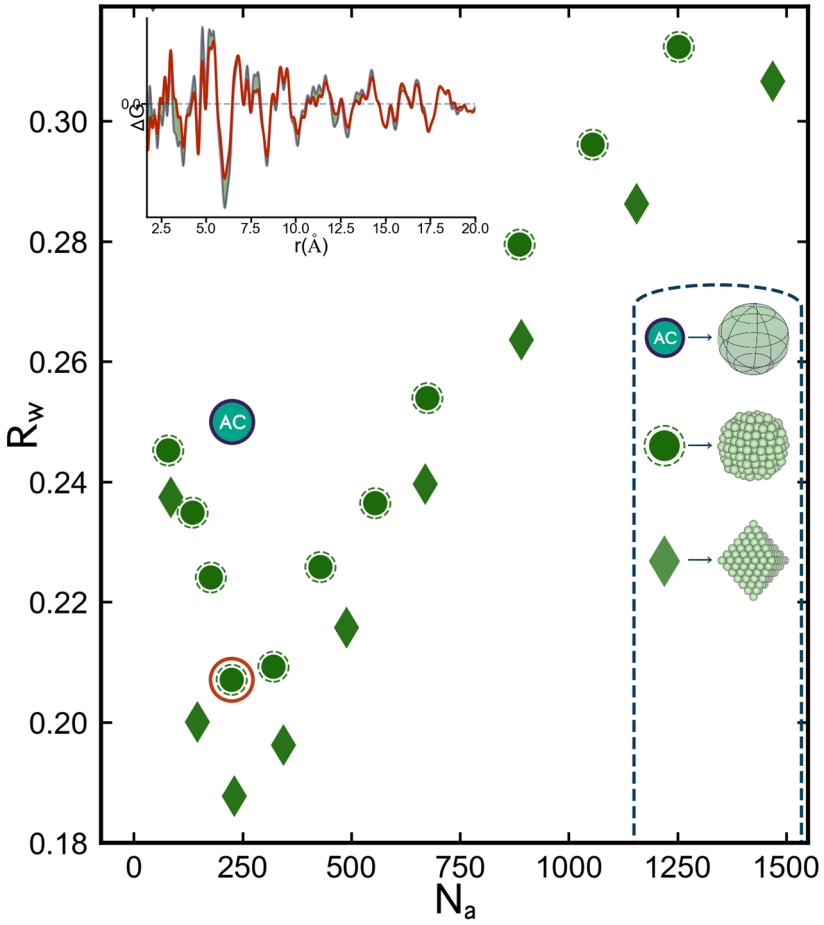
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

HTTP:/


0.30
0.28
0.26
0.24
0.22
0.20
0.18
R_w
AC
AC
0
250
500
750
1000
1250
1500
N_a

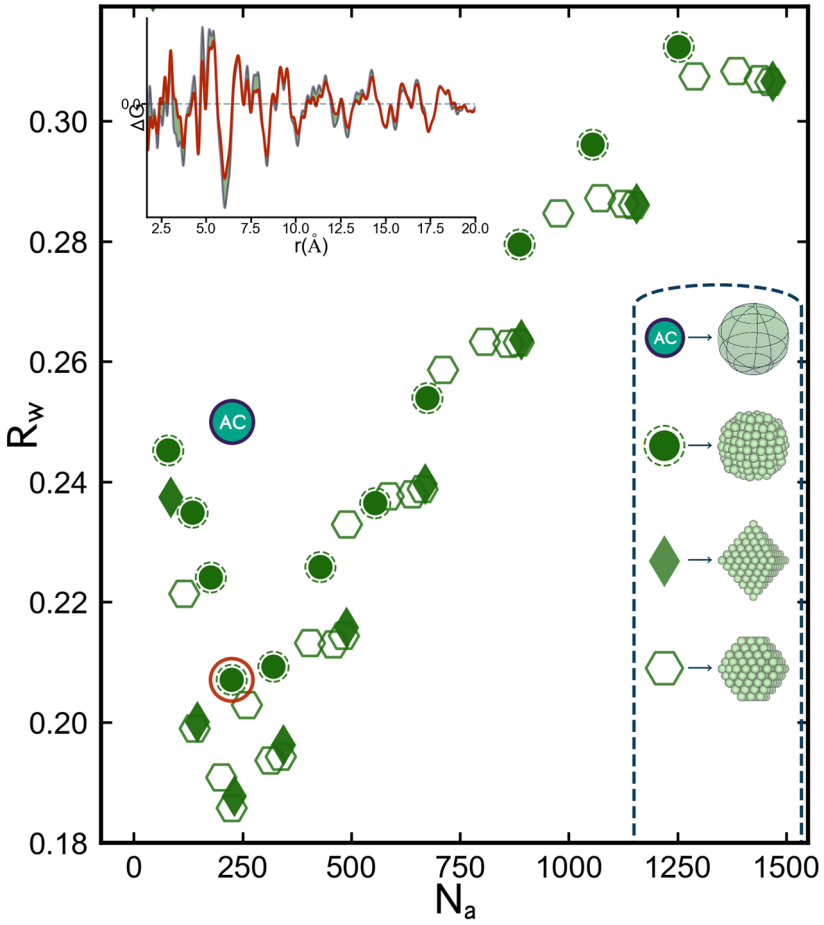

0.30
0.28
0.26
0.24
0.22
0.20
0.18
R_w
0
250
500
750
1000
1250
1500
N_a
AC
AC
0.0
ΔG
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
r(Å)



COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

BROOKHAVEN NATIONAL LABORATORY

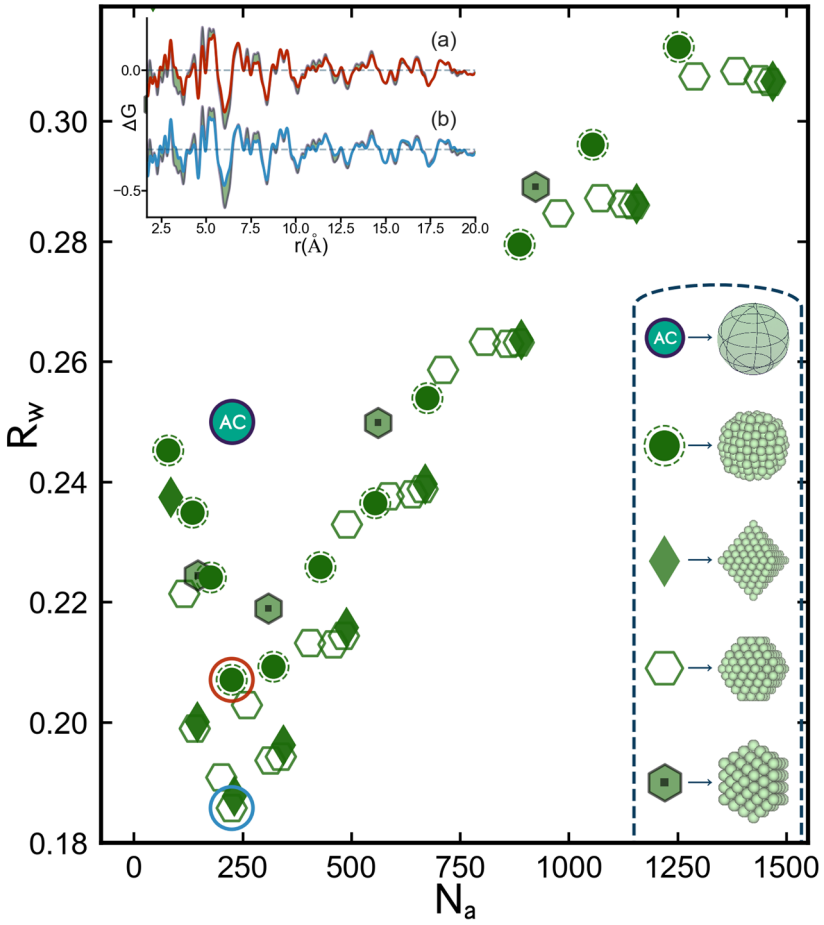
HTTP://thebillingegroup.com

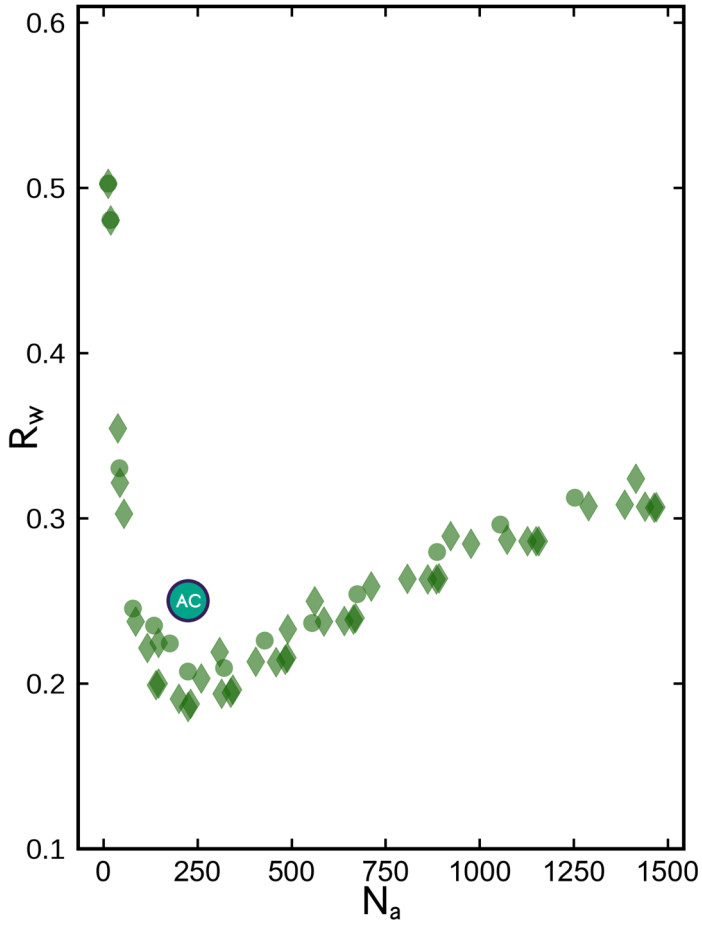


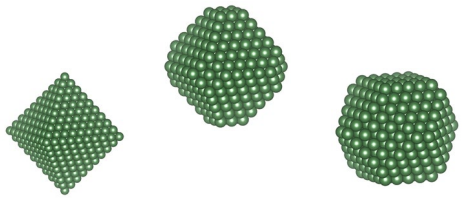
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

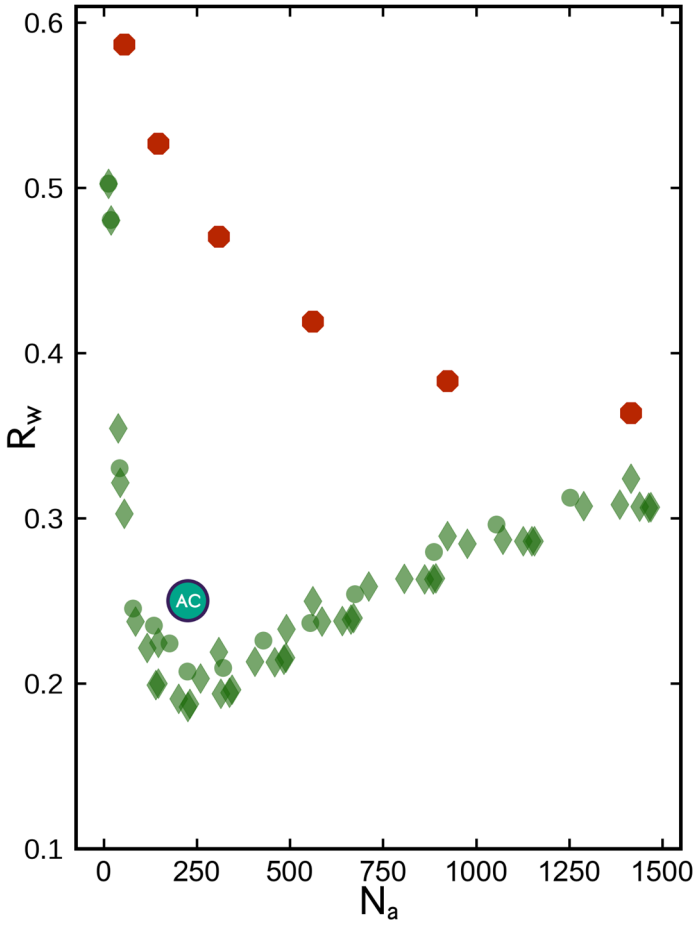
BROOKHAVEN NATIONAL LABORATORY

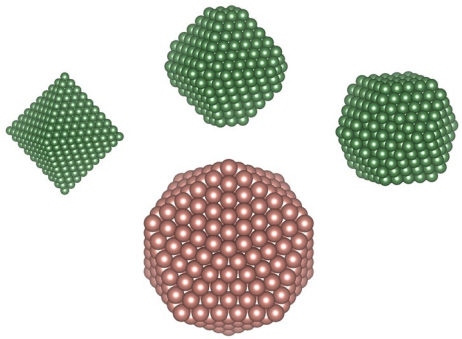
HTTP://thebillingegroup.com

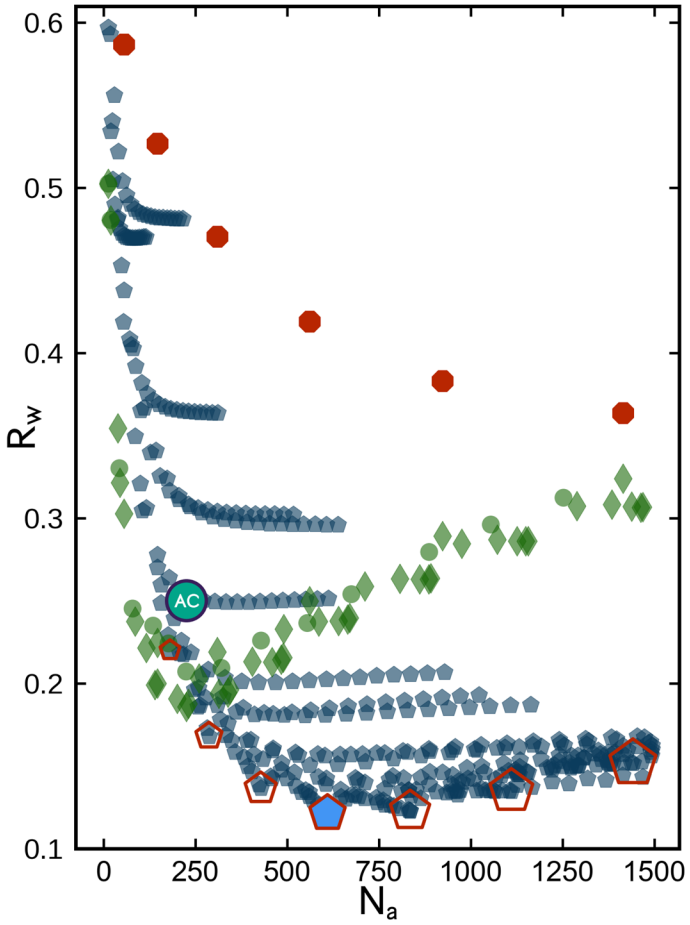

(a)
(b)
ΔG
r(Å)
AC
R_w
N_a

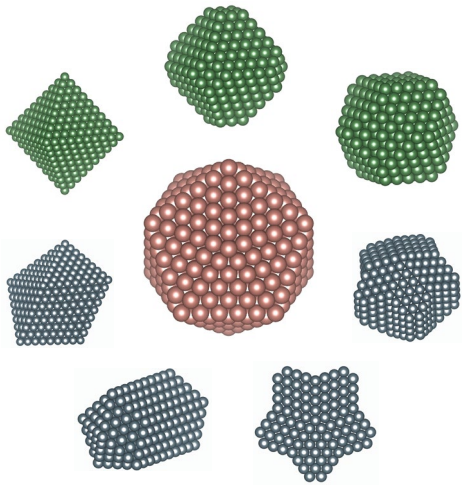

0.6
0.5
0.4
0.3
0.2
0.1
R_w
AC
0
250
500
750
1000
1250
1500
N_a

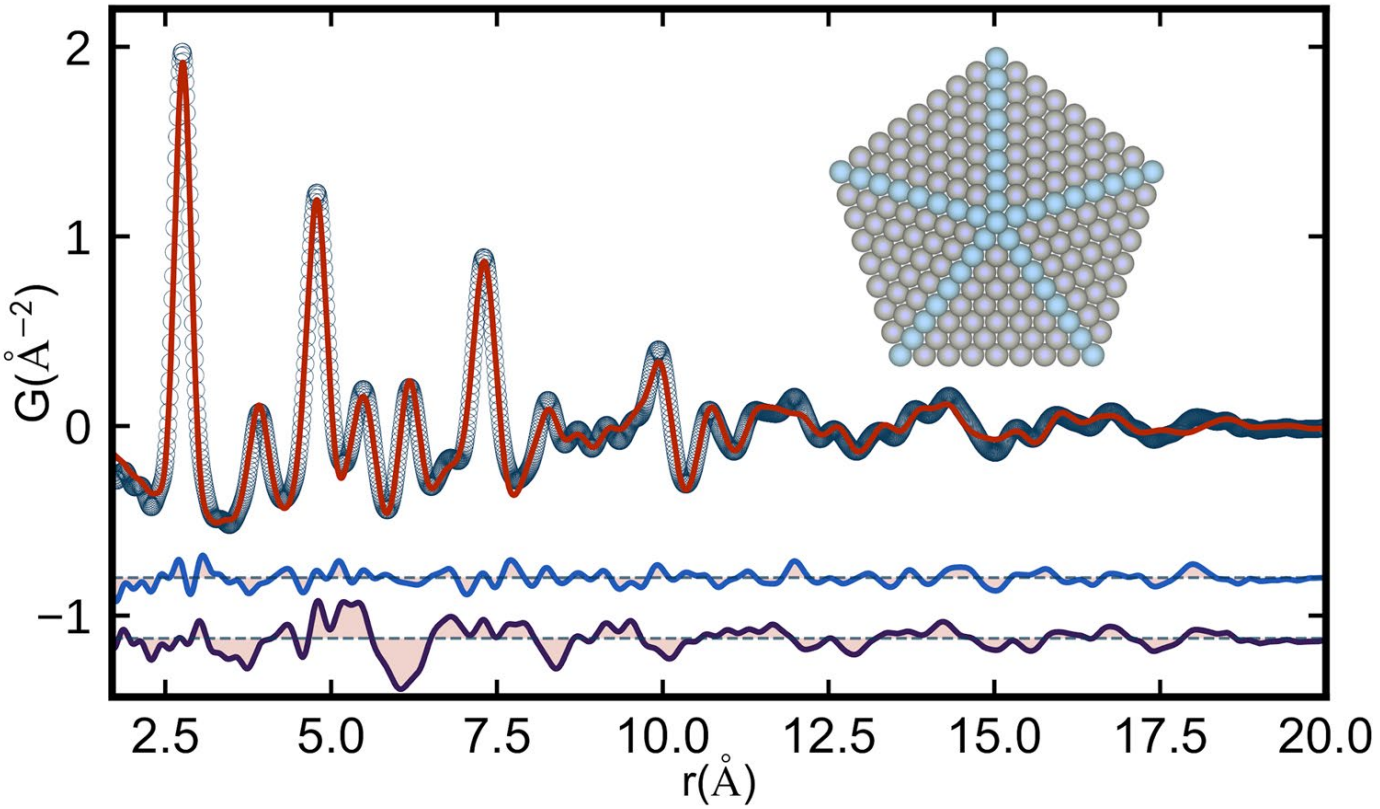


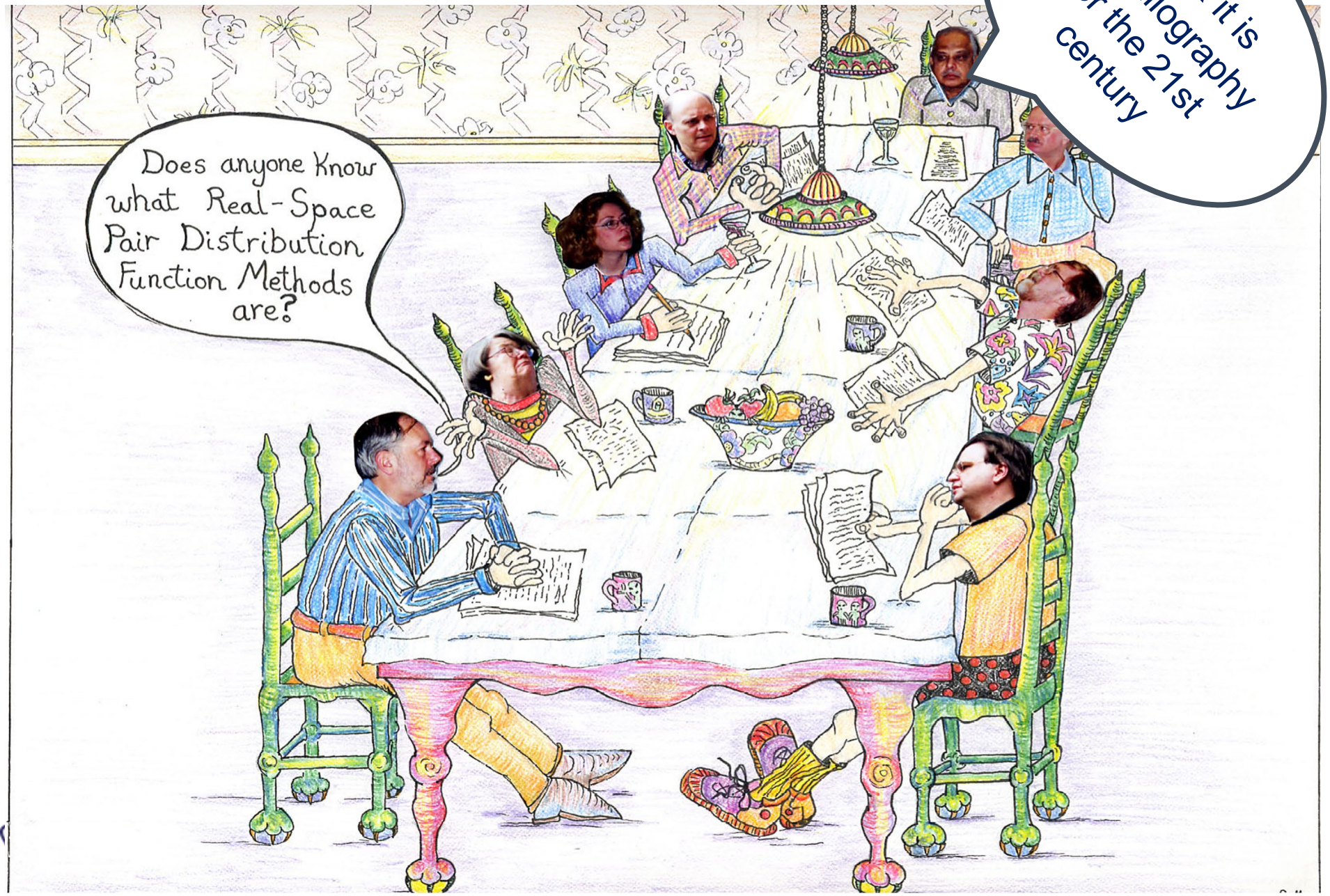

0.6
0.5
0.4
0.3
0.2
0.1
R_w
AC
0
250
500
750
1000
1250
1500
N_a




0.6
0.5
0.4
0.3
0.2
0.1
R_w
AC
0
250
500
750
1000
1250
1500
N_a




2
1
0
−1
G(Å^{-2})
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
r(Å)


Does anyone know what Real-Space Pair Distribution Function Methods are?
I think it is crystallography for the 21st century

Summary

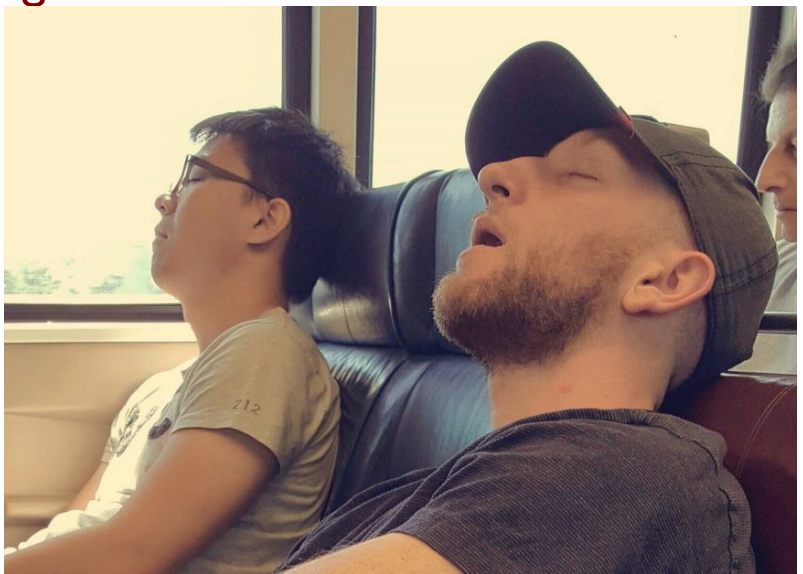
- Modern materials are complex
  - => Real Materials "crystallography"
- But modern synchrotrons/neutron sources are bright
- And modern computers are fast

Use PDF/total scattering

Avoid Fake News

Acknowledgements

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



Questions?

NEWS IN BRIEF

Seagull With Diarrhea Barely Makes It To Crowded Beach In Time

2/19/13 9:45am • SEE MORE: NATURE

NAPLES, FL—Describing it as a "real close call," a local seagull suffering from an acute case of diarrhea told reporters that he was barely able to make it to a crowded public beach in time to relieve himself Monday. "Oh,