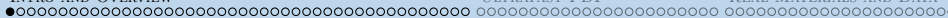


RECENT AND FUTURE DEVELOPMENTS IN PDF-LAND

Simon J. L. Billinge

¹Materials Department,
University of California, Santa Barbara

January 13, 2026



COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK





COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK



UC SANTA BARBARA



UC SANTA BARBARA



An aerial photograph of a coastal area. The foreground shows dark, rocky outcrops in the greenish-blue water. A yellow circle highlights a section of the shoreline where the water meets a sandy beach and a line of trees and buildings. In the background, there are mountains under a clear blue sky.





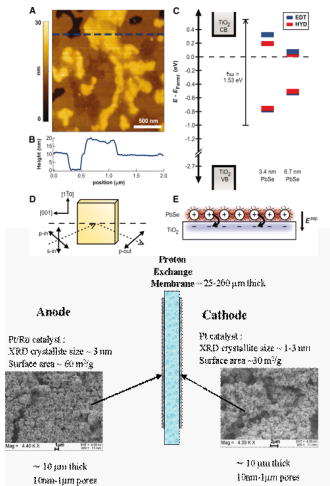
REAL MATERIALS

- 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

-

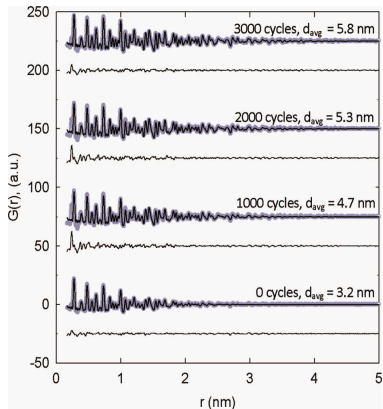
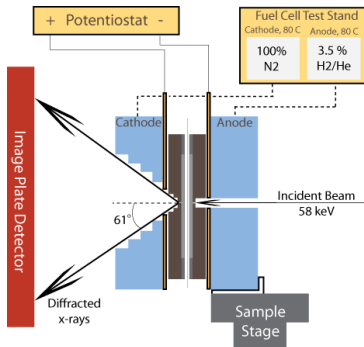
REAL MATERIALS



- 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

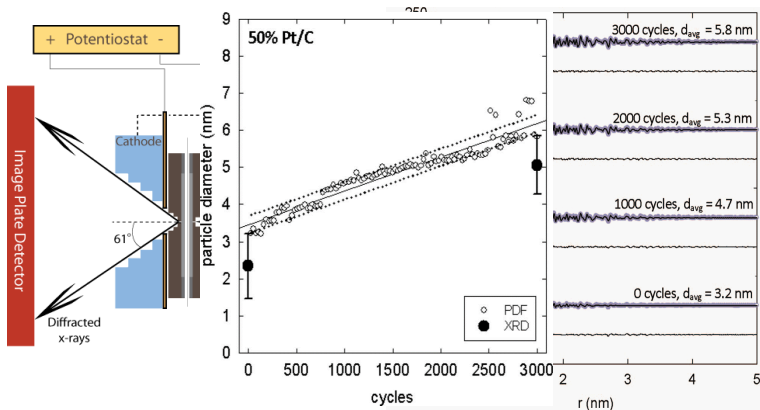


REAL MATERIALS IN ACTION: OPERANDO MEASUREMENTS

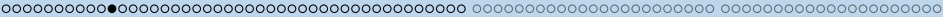


Redmond, *et al.*, *Electrochem. Solid St.*, 15 (5), B72-B74.

REAL MATERIALS IN ACTION: OPERANDO MEASUREMENTS



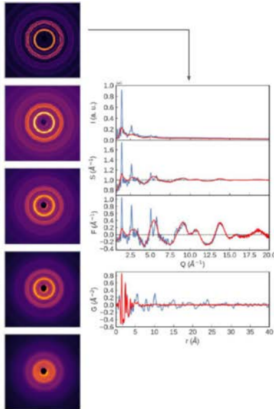
Redmond, *et al.*, *Electrochem. Solid St.*, 15 (5), B72-B74.



GETTING THE PDF FROM DATA



EXPERIMENTS



DATA ANALYSIS

$$S(Q) - 1 = \frac{I(Q)}{N \langle f \rangle^2} - \frac{\langle f^2 \rangle}{\langle f \rangle^2}$$

Normalize by scattering cross section

$$F(Q) = Q[S(Q) - 1]$$

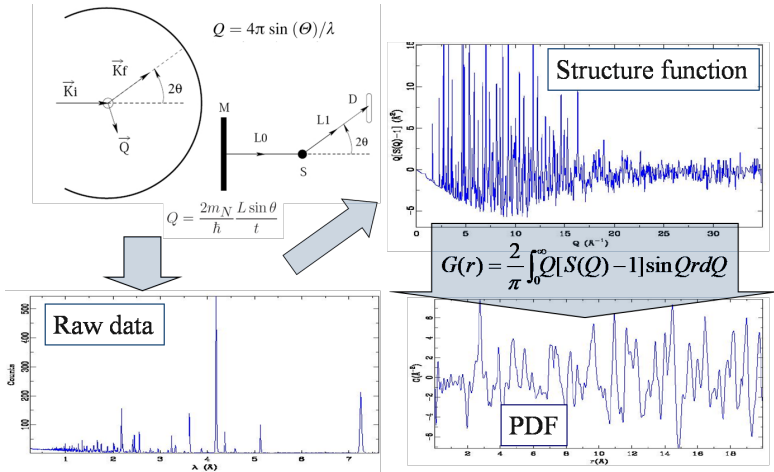
Transform to physical units

$$G(r) = \frac{2}{\pi} \int_0^\infty F(Q) \sin(Qr) dQ$$

Fourier transform

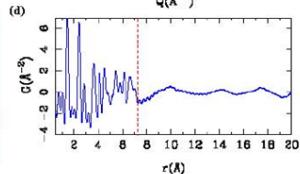
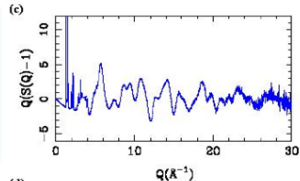
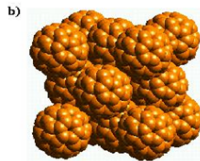
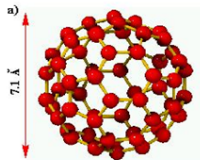
AI/ML/ALGORITHM
DEVELOPMENT

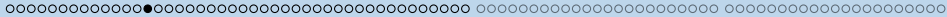
GETTING THE PDF FROM DATA



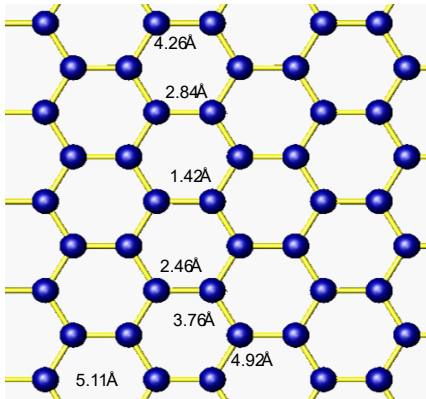
PDF INTUITION

- Sit on an atom and look at your neighborhood
- $G(r)$ gives the probability of finding a neighbor at a distance r
- PDF is experimentally accessible
- PDF gives the local structure

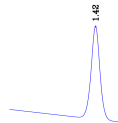
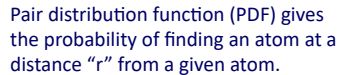




GETTING THE PDF FROM A MODEL

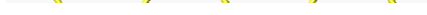


Pair distribution function (PDF) gives the probability of finding an atom at a distance “ r ” from a given atom.













STRUCTURE OF SMALL GOLD CLUSTERS

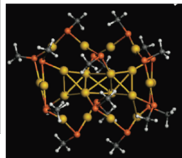
J|A|C|S
COMMUNICATIONS

Published on Web 09/02/2009

Thiolate-Protected $\text{Au}_{20}(\text{SR})_{16}$ Cluster: Prolate Au_8 Core with New $[\text{Au}_3(\text{SR})_4]$ Staple Motif

Yong Pei, Yi Gao, Nan Shao, and Xiao Cheng Zeng*

Department of Chemistry and Nebraska Center for Materials and Nanoscience, University of Nebraska-Lincoln, Lincoln, Nebraska 68588



20

J|A|C|S
COMMUNICATIONS

Published on Web 04/12/2008

Correlating the Crystal Structure of A Thiol-Protected Au_{25} Cluster and Optical Properties

Manzhou Zhu,[†] Christine M. Aikens,[‡] Frederick J. Hollander,[§] George C. Schatz,[¶] and Rongchao Jin^{*,†}

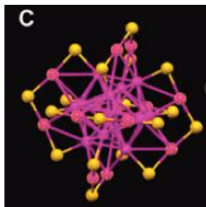
J|A|C|S
COMMUNICATIONS

Published on Web 06/01/2010

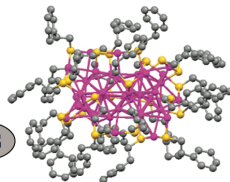
Total Structure Determination of Thiolate-Protected Au_{38} Nanoparticles

Huifeng Qian,[†] William T. Eckenhoff,[‡] Yan Zhu,[†] Tomislav Pintauer,[§] and Rongchao Jin^{*,†}

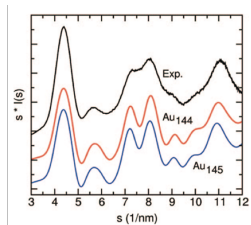
Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 and Department of Chemistry and Biochemistry, Duquesne University, Pittsburgh, Pennsylvania 15282



25



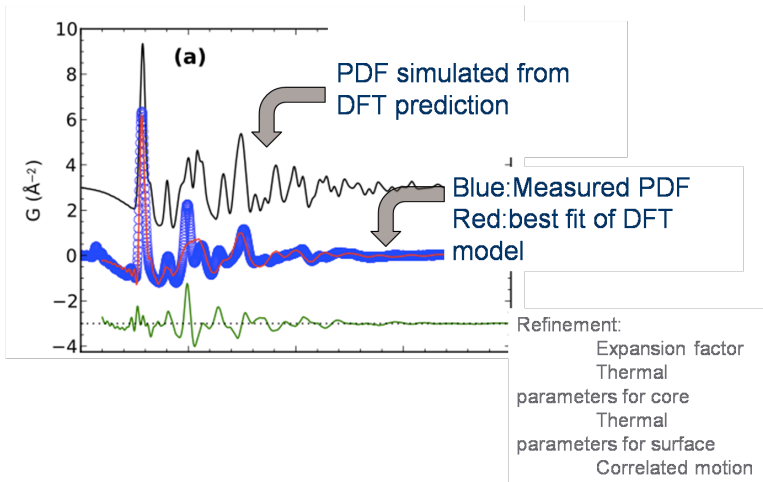
38

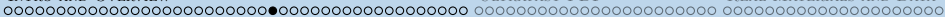


Olga Lopez-Acevedo,[†] Jaakko Akola,[†] Robert L. Whetten,[‡] Henrik Grönbeck,[§] and Hannu Häkkinen^{*,†,||}

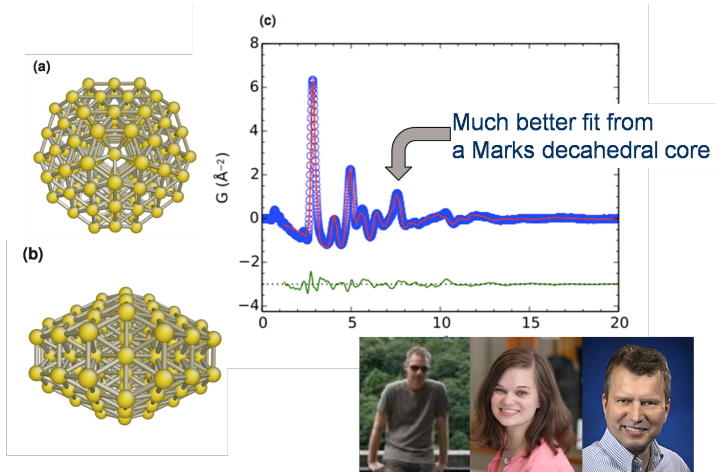
Olga Lopez-Acevedo,[†] Jaakko Akola,[†] Robert L. Whetten,[‡] Henrik Grönbeck,[§] and Hannu Häkkinen^{*,†,||}

STRUCTURE OF SMALL GOLD CLUSTERS





THE MD6441 STRUCTURE: 144 GOLD ATOMS



STRUCTURE OF SMALL GOLD CLUSTERS

NATURE COMMUNICATIONS | 7:11859 | DOI: 10.1038/ncomms11859

ARTICLE

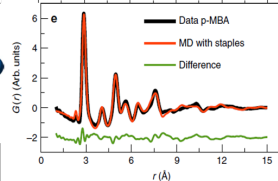
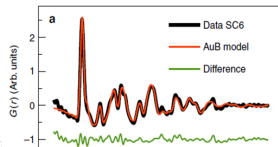
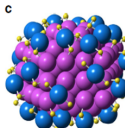
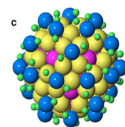
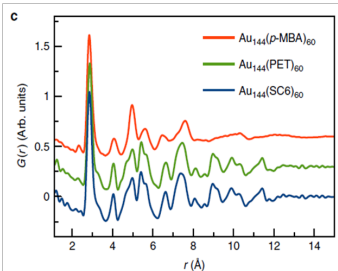
Received 22 Aug 2015 | Accepted 6 May 2016 | Published 14 Jun 2016

DOI: 10.1038/ncomms11859

OPEN

Polymorphism in magic-sized $\text{Au}_{144}(\text{SR})_{60}$ clusters

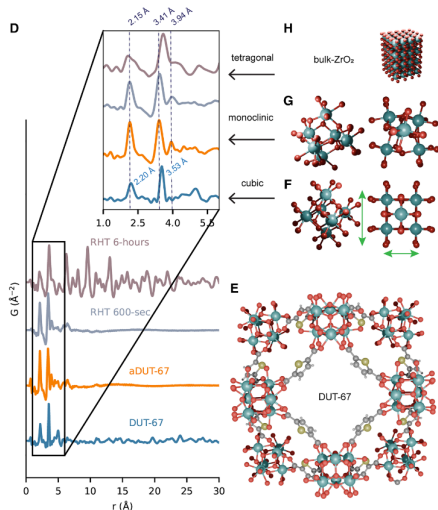
Kirsten M.Ø. Jensen^{1,*}, Pavol Juhas^{2,*}, Marcus A. Tofanelli³, Christine L. Heinecke³, Gavin Vaughan⁴, Christopher J. Ackerson³ & Simon J.L. Billinge^{1,2}



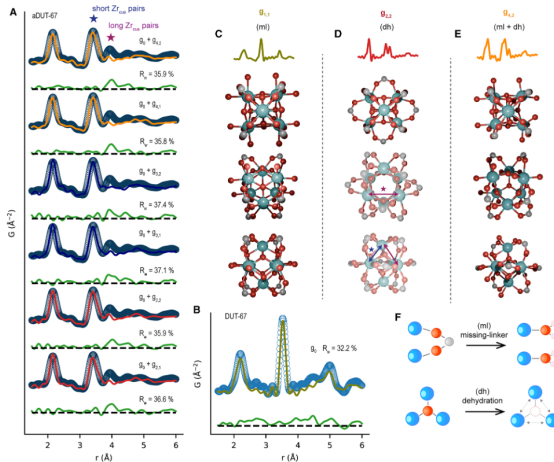
PDF ON MOFs

Distortions to the metal cluster nodes in a catalytic MOF

- Till Schertenleib
- collaboration with Wendy Queen and Mehrdad Asgari
- In Chem 11 (2025),
- doi:
10.1016/j.chempr.2025.102619

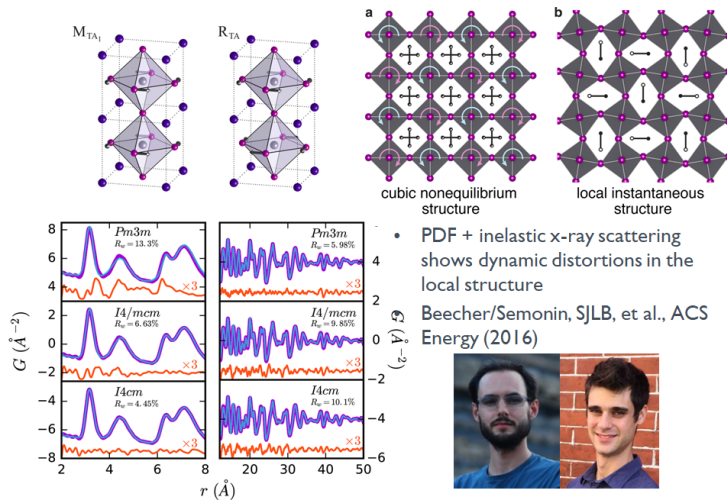


PDF ON MOFs

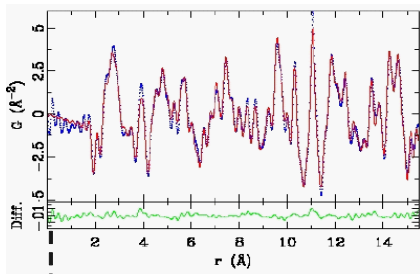




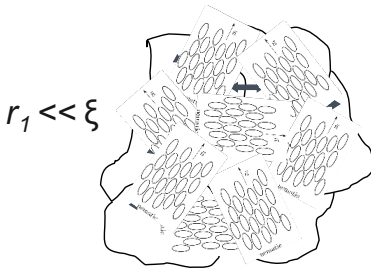
BROKEN LOCAL SYMMETRY DOMAINS IN MAPI

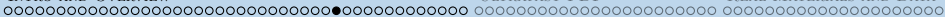


HOW CAN WE SEE LOCAL BROKEN SYMMETRIES?

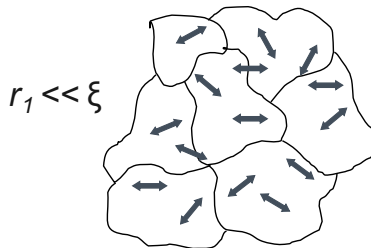
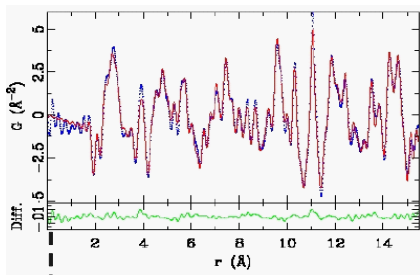


Inter-domain structure





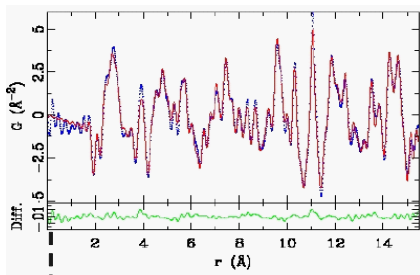
HOW CAN WE SEE LOCAL BROKEN SYMMETRIES?



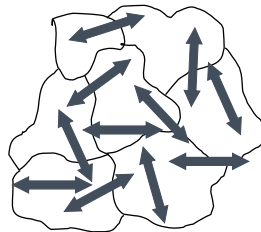
r_1
 r_2
 Intra-domain structure Inter-domain structure



HOW CAN WE SEE LOCAL BROKEN SYMMETRIES?



$$r_2 \sim \xi/2$$

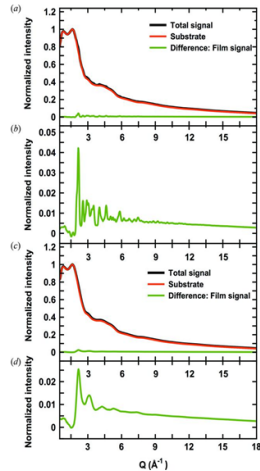
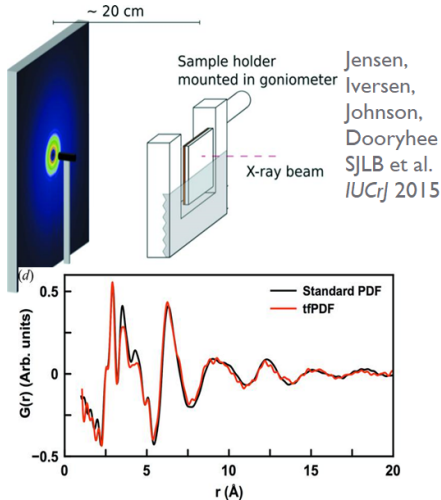

 r_1
 r_2

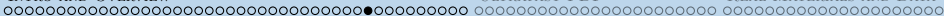
Intra-domain structure

Inter-domain structure

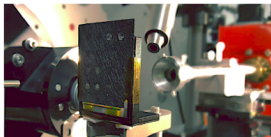


THIN FILM PDF (AT NORMAL INCIDENCE) TFPDF

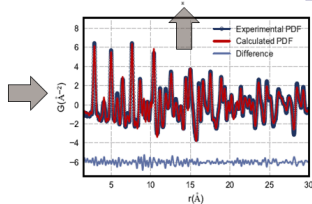
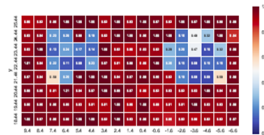
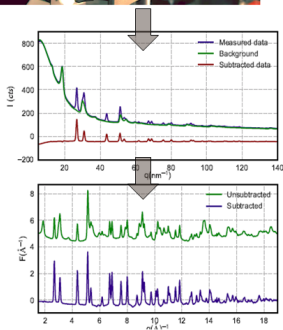




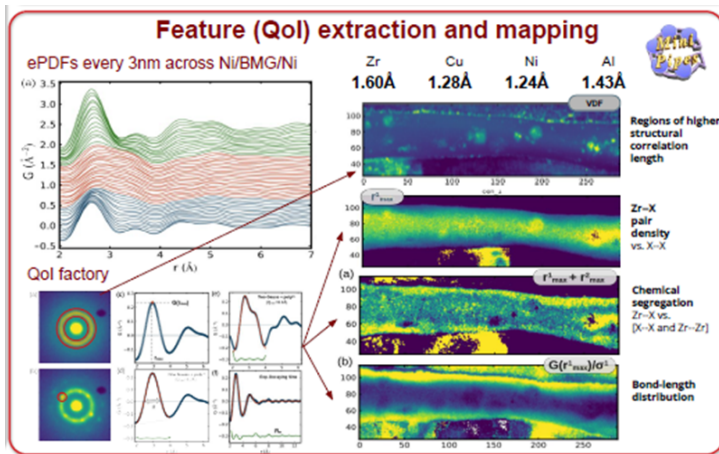
SPATIALLY RESOLVED PDFs (LAB ON A CHIP)



- Kovyakh, Banerjee, Liu, et al., arXiv:2110.01656 [cond-mat]
- Every pixel contains a complete PDF



SCANNING NANOSTRUCTURE EM (SNEM)

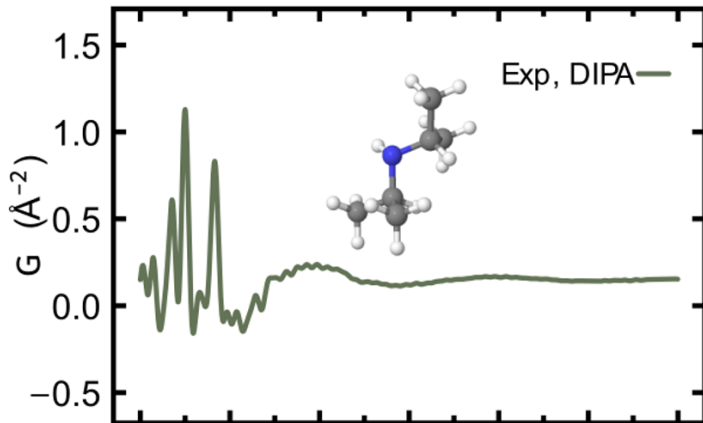


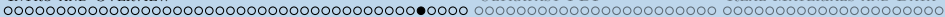
Rakita, *et al.* Acta Materiala 242 (2023), 118426. 10.1016/j.actamat.2022.118426



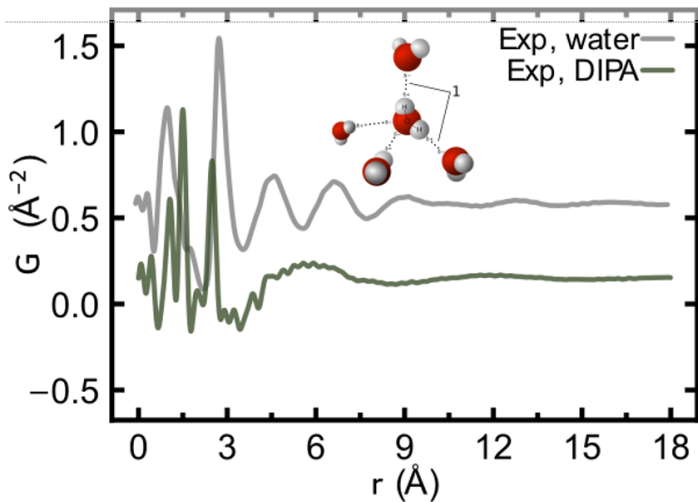
- Collaboration with Ngai Yin Yip (Columbia University)
- Work of students Ian Billinge, Songsheng Tao
- I. Billinge *et al.* , Matter (2024) doi:
10.1016/j.matt.2024.09.023

PDF of amine



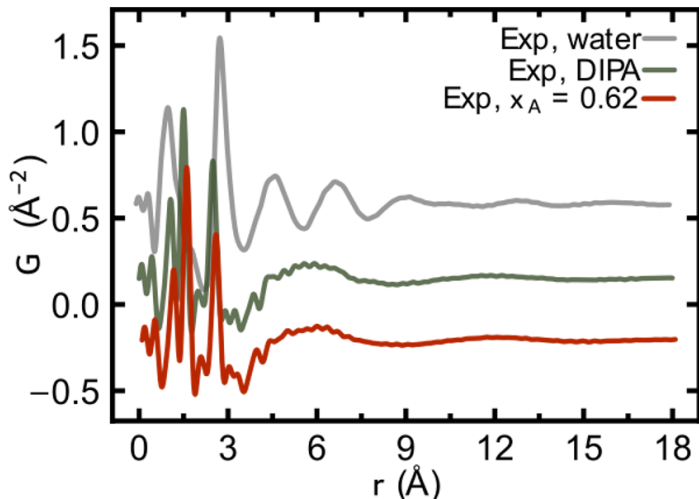


STRUCTURED LIQUIDS: AMINE - WATER MIXTURES





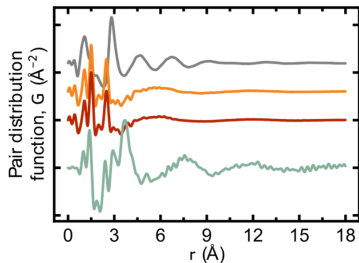
STRUCTURED LIQUIDS: AMINE - WATER MIXTURES





STRATEGY - FIND THE AMINE-WATER CORRELATIONS

PDFs and the ddG



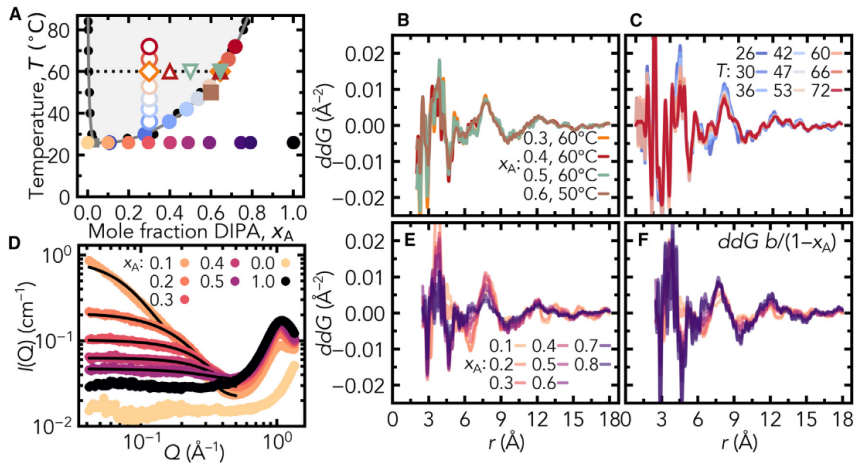
How to get the way that Amine and Water molecules coordinate each other?

- $G_{ww}(r)$ - water - water correlations
- $G_{aa}(r)$ - amine - amine correlations
- $G_{aw}(r)$ - amine - water correlations
- $G_{mixture}, G_{water}, G_{amine}$ - measured PDFs

$$ddG = G_{mixture} - \alpha G_{water} - \beta G_{amine}$$

$$G_{aw} \approx ddG$$

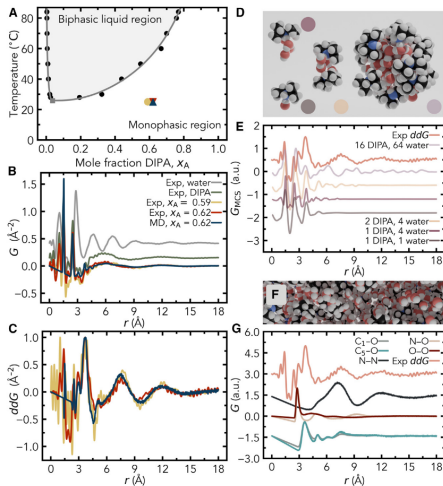
RESULTS

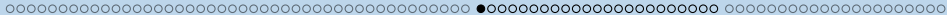


RESULTS

In the mixture:

- Amine hydrogen bonds with water
- Water hydrogen bonds with water
- Results in formation of stable inverse micelle objects
- They do not change size but change quantity





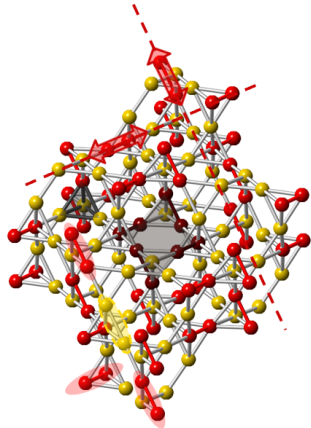
ODL AND THE METAL-INSULATOR TRANSITION IN CuIr_2S_4

Emil Bozin

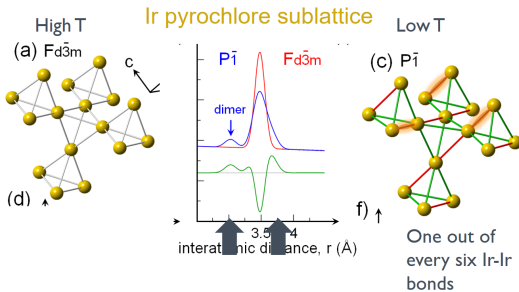
w/ JF Mitchell, ANL

M. Abeykoon, BNL

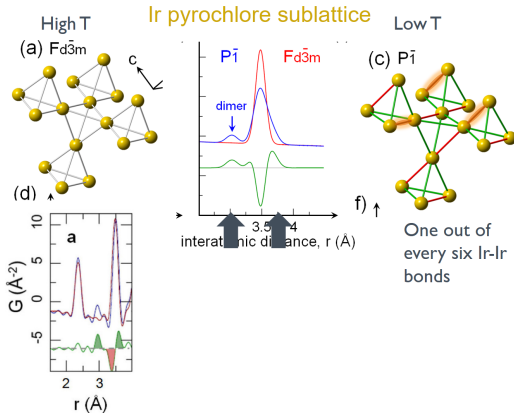
Data from 28-ID-2 NSLS-II



ODL AND THE METAL-INSULATOR TRANSITION IN CuIr_2S_4

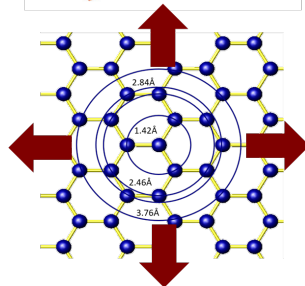
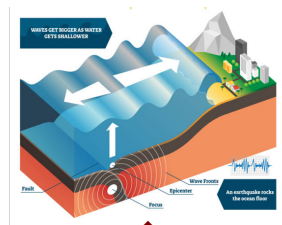


ODL AND THE METAL-INSULATOR TRANSITION IN CuIr_2S_4

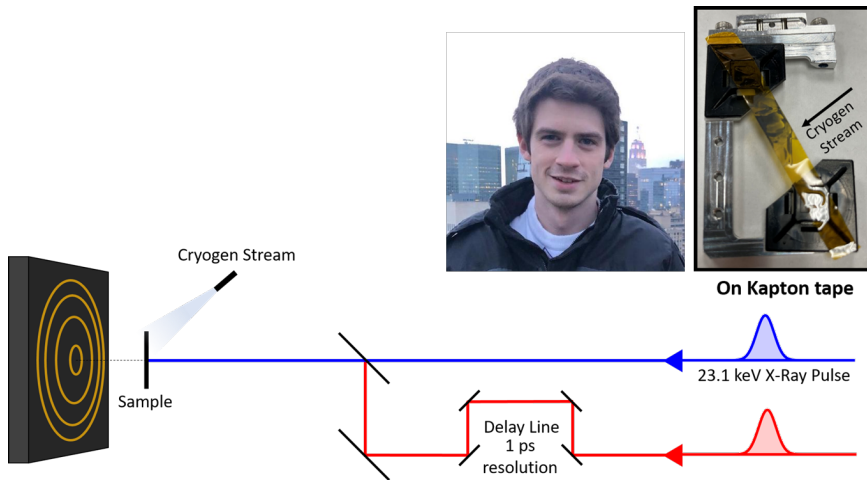


GOAL OF THE EXPERIMENT

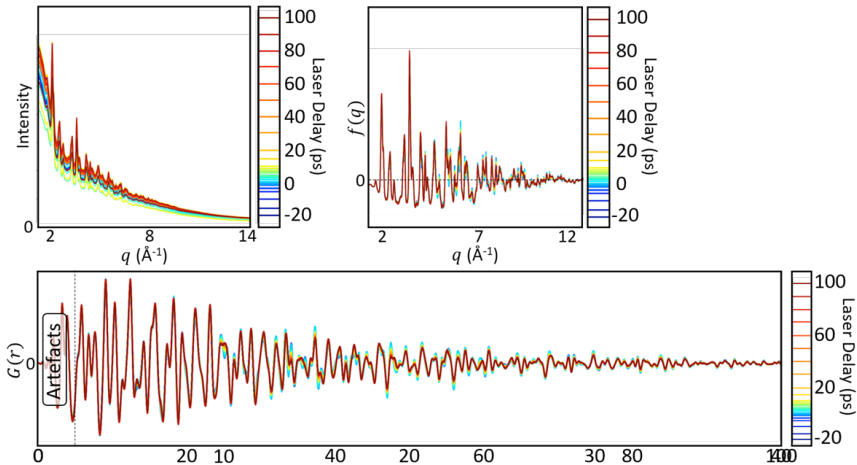
- Cool sample so it is in the dimer state
- Photo-excite with fs laser pulse to “excite” the system
- Does the dimer go away?
- What is the resulting state?
- How quickly is the dimer state destroyed?
- How does the effect propagate through the lattice

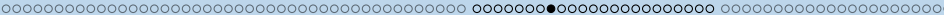


EXPERIMENTAL SETUP

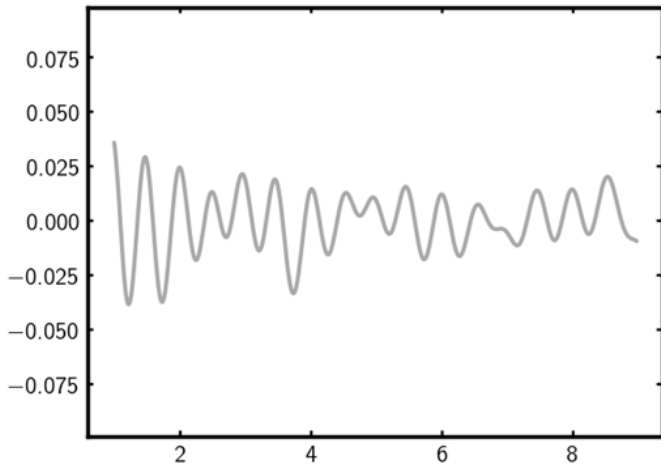


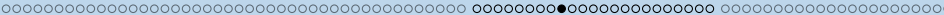
DATA REDUCTION



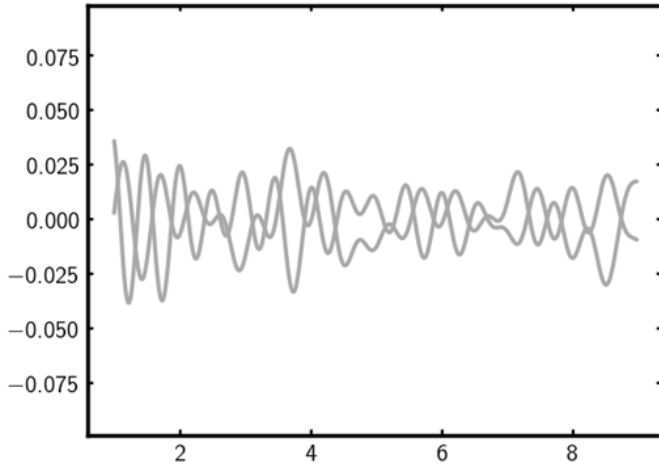


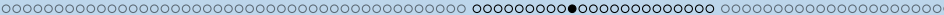
$$\Delta G = G(t = -20\text{ps}) - \langle G(t < 0) \rangle$$



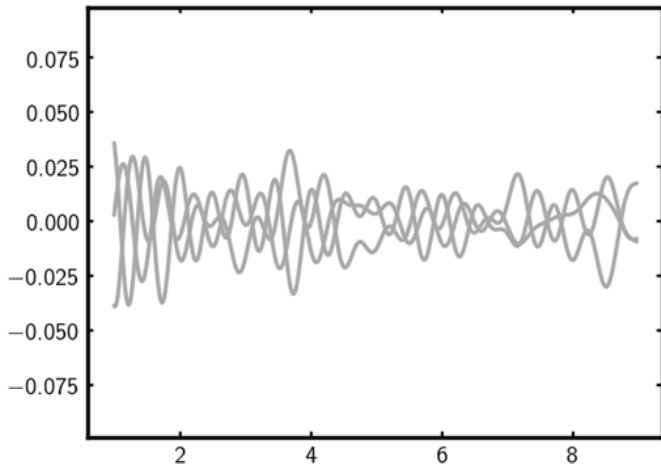


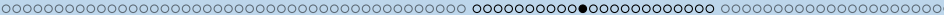
$$\Delta G = G(t = -15\text{ps}) - \langle G(t < 0) \rangle$$



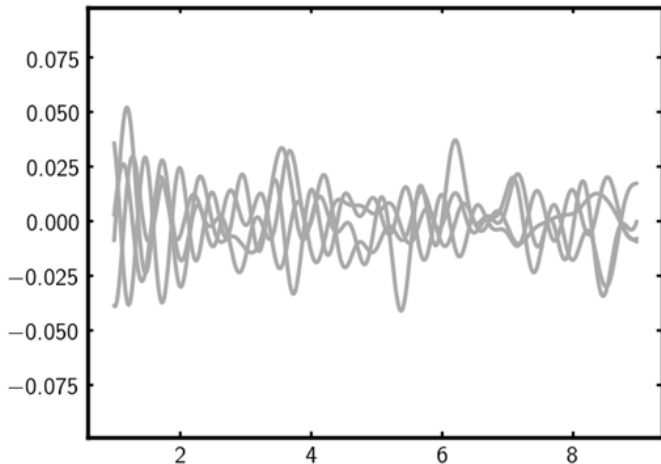


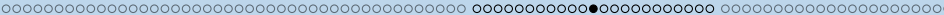
$$\Delta G = G(t = -10\text{ps}) - \langle G(t < 0) \rangle$$



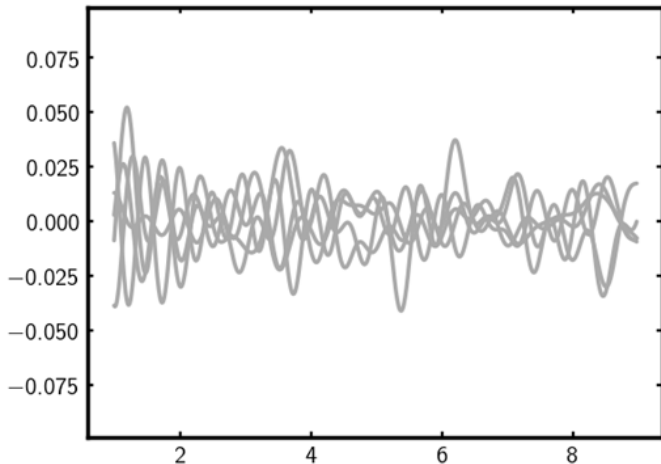


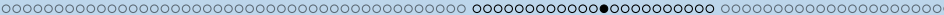
$$\Delta G = G(t = -5\text{ps}) - \langle G(t < 0) \rangle$$



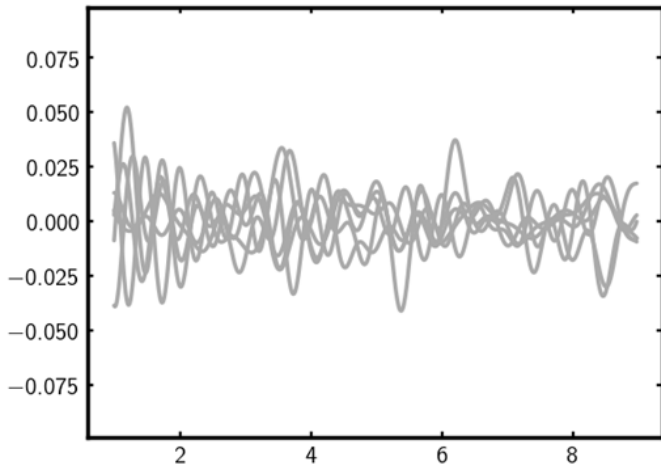


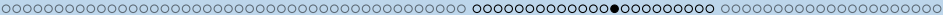
$$\Delta G = G(t = -3\text{ps}) - \langle G(t < 0) \rangle$$



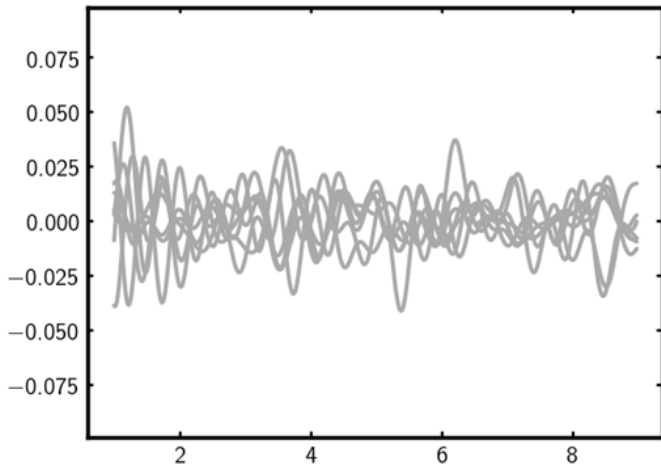


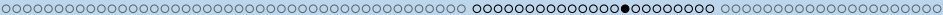
$$\Delta G = G(t = -2\text{ps}) - \langle G(t < 0) \rangle$$



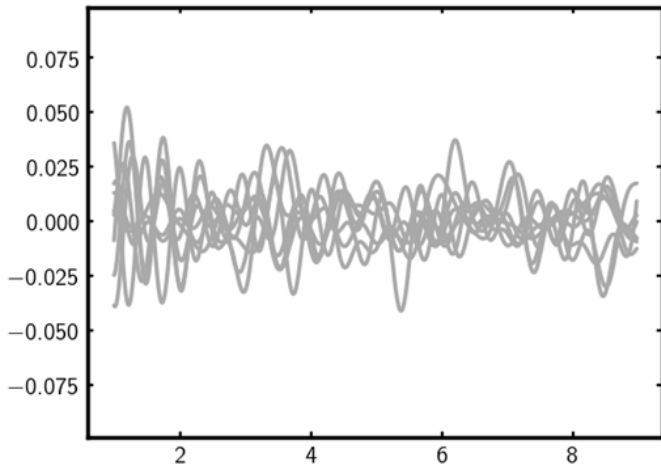


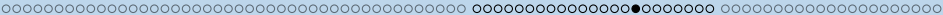
$$\Delta G = G(t = -1\text{ps}) - \langle G(t < 0) \rangle$$



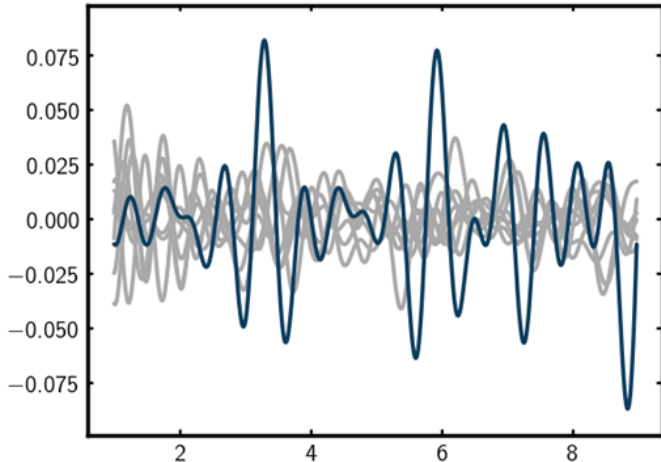


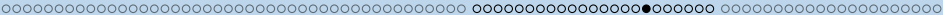
$$\Delta G = G(t = -0\text{ps}) - \langle G(t < 0) \rangle$$



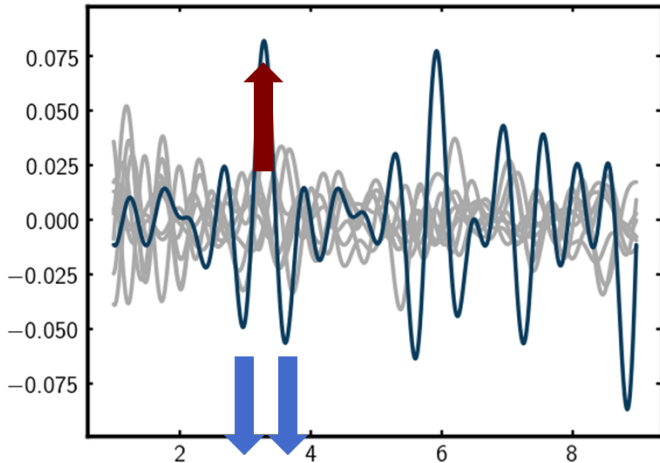


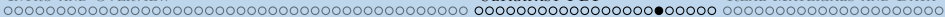
$$\Delta G = G(t = +1\text{ps}) - \langle G(t < 0) \rangle$$



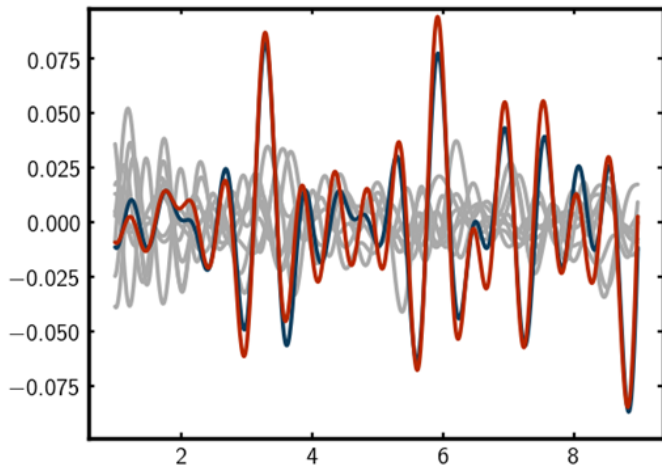


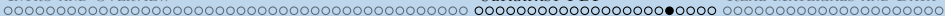
$$\Delta G = G(t = +1\text{ps}) - \langle G(t < 0) \rangle$$



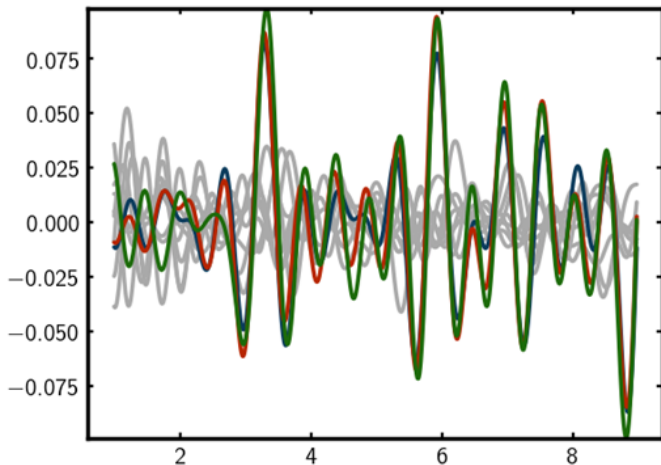


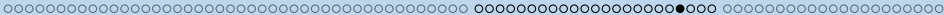
$$\Delta G = G(t = +2\text{ps}) - \langle G(t < 0) \rangle$$



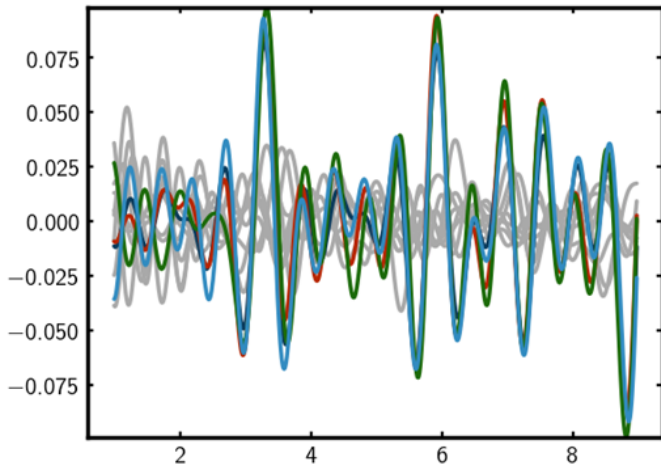


$$\Delta G = G(t = +3\text{ps}) - \langle G(t < 0) \rangle$$



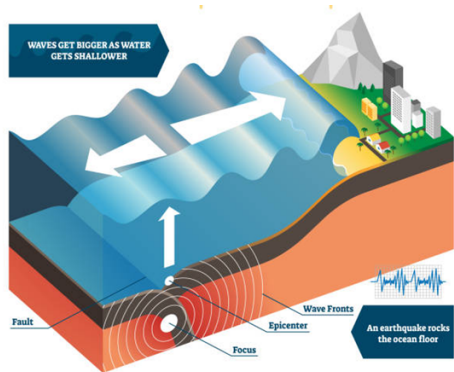


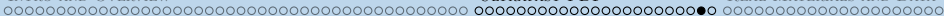
$$\Delta G = G(t = +4\text{ps}) - \langle G(t < 0) \rangle$$



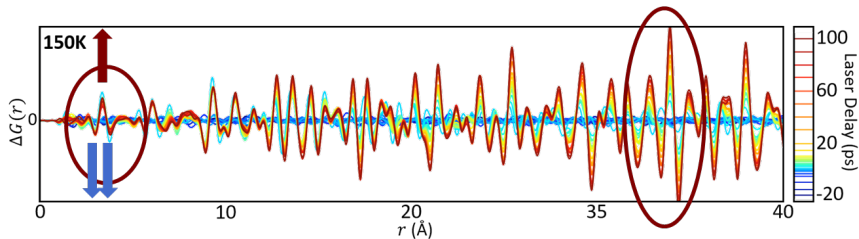
WHAT ABOUT THE TSUNAMI MODEL?

- Object destroyed/excited locally but...
- How do the effects propagate out through the lattice?

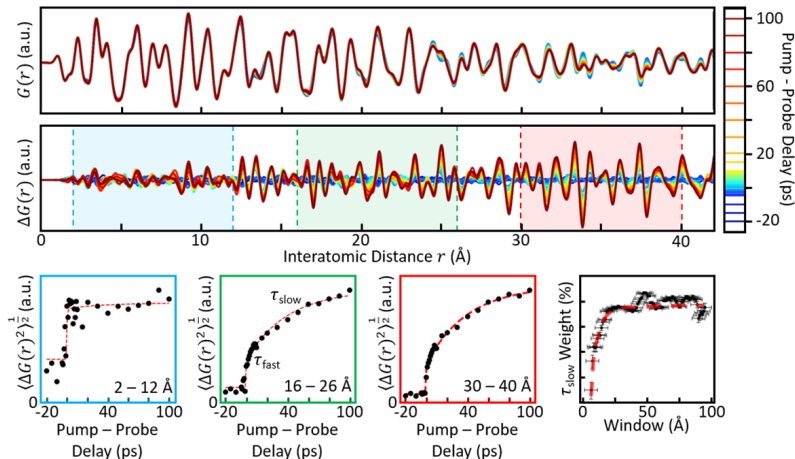




R-DEPENDENT RESPONSE



R-DEPENDENT RESPONSE



- Broad signals, in the noise
- Overlapping signals
- Weak signals

AI/ML is helping



NO WELL DEFINED FORWARD MODEL

The information is in the data but we don't know how to get it out

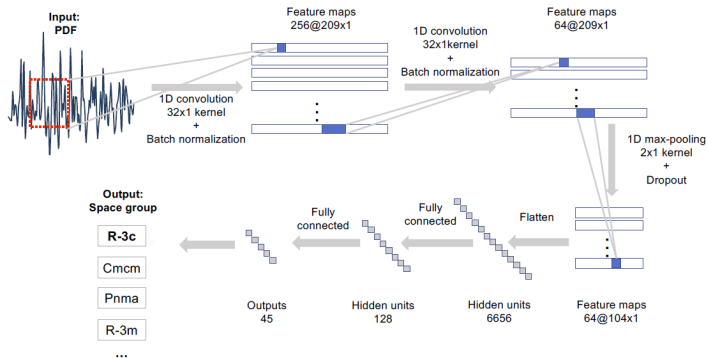


GIVEN A PDF, WHAT WAS THE SPACE-GROUP OF THE MATERIAL?

PDF, tell me your space-group

- SG info from Bragg peak systematic absences, but in the PDF?
- But the information is there (PDF is just a FT of diffraction)
- We don't know the solution to the forward problem!
- Let's try ML. It is a classification problem
- We will try a convolutional neural net

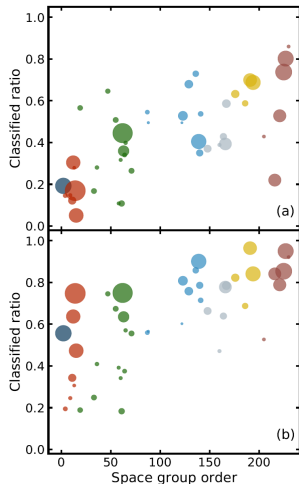
GIVEN A PDF, WHAT WAS THE SPACE-GROUP OF THE MATERIAL?



collaboration with group of Qiang Du (Columbia U)

Chia-Hao Liu, SJLB *et al.*, Acta Cryst. (2019), 10.1107/S2053273319005606

HOW WELL DID IT PERFORM?

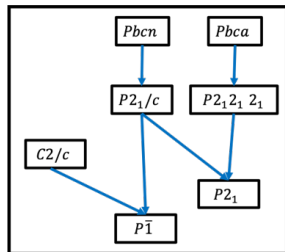
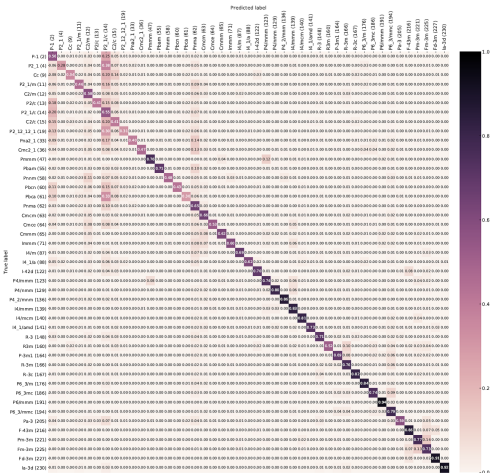


- Top six classification ratio
- (top) Logistic Regression baseline
- (bottom) CNN 91% overall
- Color indicates crystal system
- Size of spot indicates sample-size

WE LEARNED CRYSTALLOGRAPHY FROM THE MODEL

Matrix of confusion

- tear drops on space-groups
- C2/1, Pnma
- connected by group-sub-group relationships

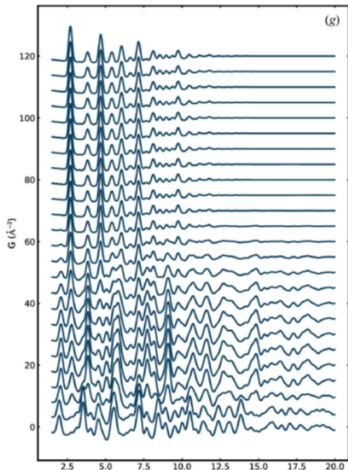


SIGNAL EXTRACTION

Separating the wheat from the chaff



IN SITU/OPERANDO DATA DURING BATTERY ELECTRODE DISCHARGE



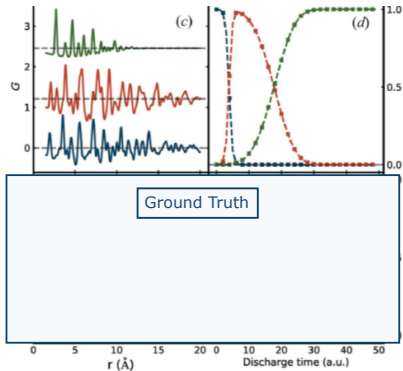
PDFs varying with time

- how do we follow chemical changes?
- can do multi-phase Rietveld refinement, but what if we don't know the phases?
- multivariate analysis: PCA, NMF
- considered as “unsupervised ML”

C.H. Liu, SJLB et.al, J. Appl. Crystallogr. (2021), 10.1107/S160057672100265X

Z. Thatcher, Acta Cryst. A (2022), 10.1107/S2053273322002522

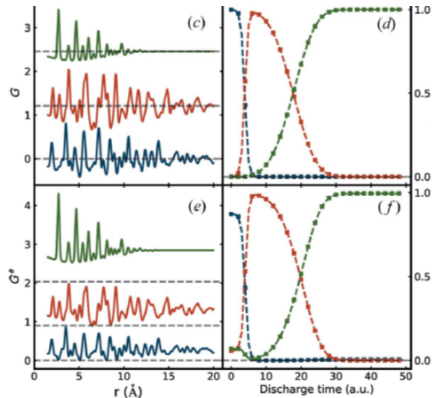
IN SITU/OPERANDO DATA DURING BATTERY ELECTRODE DISCHARGE



PDFs varying with time

- simulation
- PDFs of the components are shown
- Fractions of each phase are shown
- This is the ground-truth (the known solution)

IN SITU/OPERANDO DATA DURING BATTERY ELECTRODE DISCHARGE

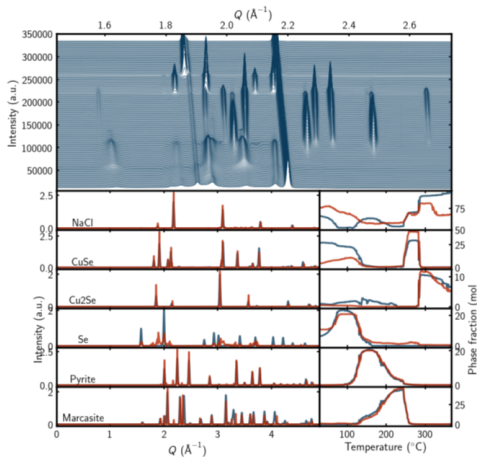


PDFs varying with time

- simulation
- PDFs of the components are shown
- Fractions of each phase are shown
- Top is the ground-truth (the known solution)
- Bottom is what was returned by NMF

The algorithm had no chemical information!

DOES IT WORK ON REAL EXPERIMENTAL DATA?



Collaboration with the groups of Jamie Neilson, Karena Chapman and Qiang Du

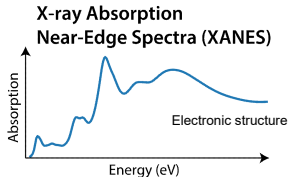
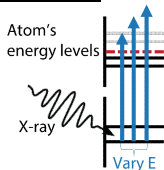
MULTI-MODAL ANALYSIS

The information is in there, but we don't know how to get it out

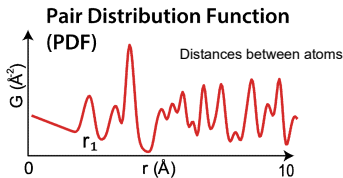
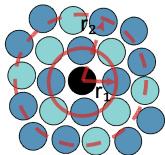


HOW CAN WE COMBINE HETEROGENEOUS DATA INPUTS?

Input 1: XANES



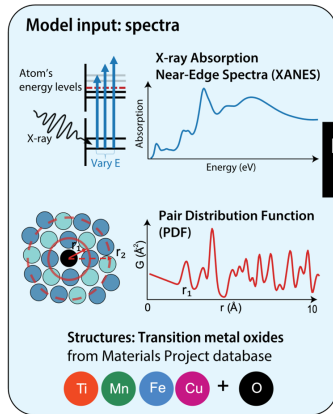
Input 2: PDF



A strategy for doing multi-modal analysis?

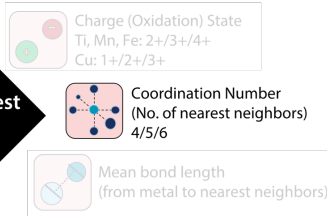
- Collaboration with Steven Torrisi (Toyota Research Institute)
- Lead researchers Tanaporn (Tina) Na Narong, Zoe Zachko
- T. Na Narong *et al.* : npj-Computational Materials (2025) 10.1038/s41524-025-01589-3

STRATEGY:

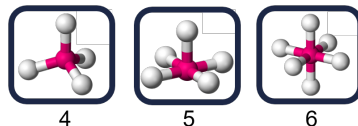


Random Forest Models

**Prediction targets:
metal's local atomic environments**



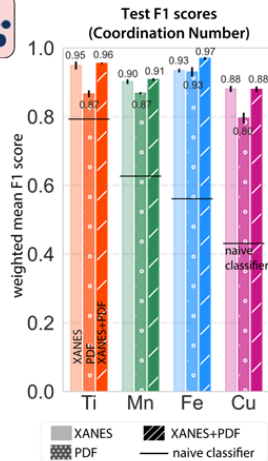
How many nearest neighbors does the metal atom have?



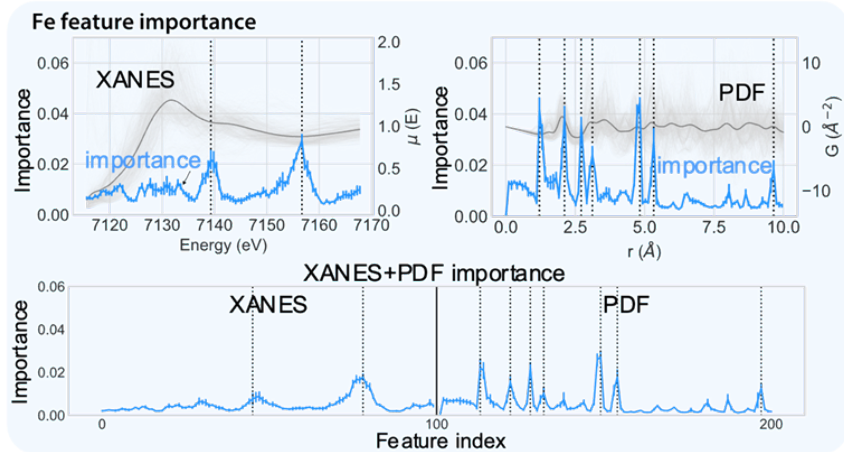


HOW DO DIFFERENT INFORMATION SOURCES PERFORM?

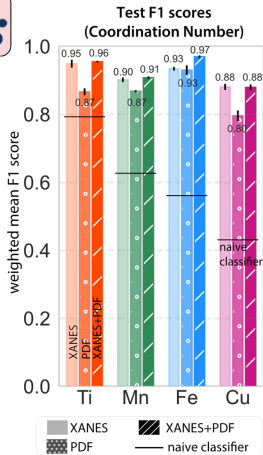
- XANES alone
- PDF alone
- XANES + PDF
- baselines



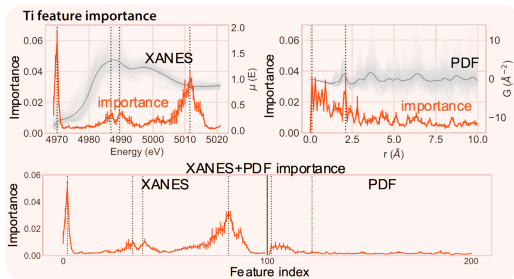
WHERE IS THE INFORMATION IN THE SPECTRA?



WHERE IS THE INFORMATION IN THE SPECTRA - Ti?



Results can look very different for a different metal





CAN WE DO STRUCTURE SOLUTION WITH GENERATIVE AI?

The goal

- Input: the PDF or a powder diffraction pattern
- Output: the structure of the material

That would be cool!

CASE 1: CLOSE PACKED STRUCTURES

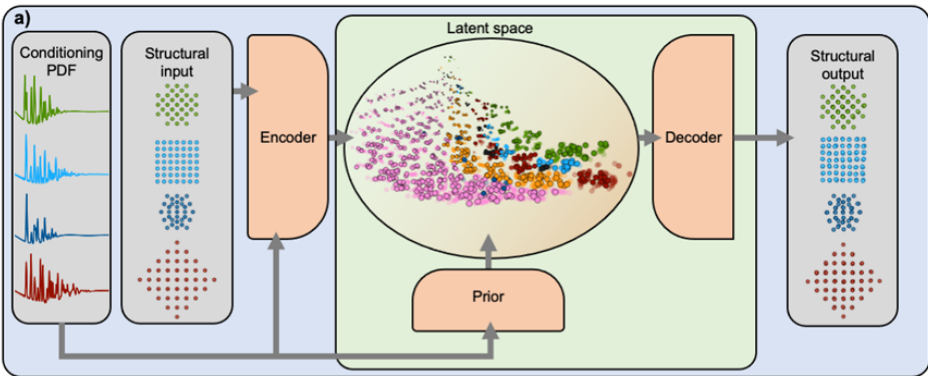
Let's do close-packed structures that interpolate

- Graph Convolutional Variational Autoencoder (CVAE)
- Collaboration with Kirsten Jensen, Raghav Selvan, U. Copenhagen
- Work of students Emil Kjaer and Andy Anker
- Kjaer, Anker, *et al.* , Digital Discovery (2023), 10.1039/D2DD00086E

Structure solution using Neural Nets

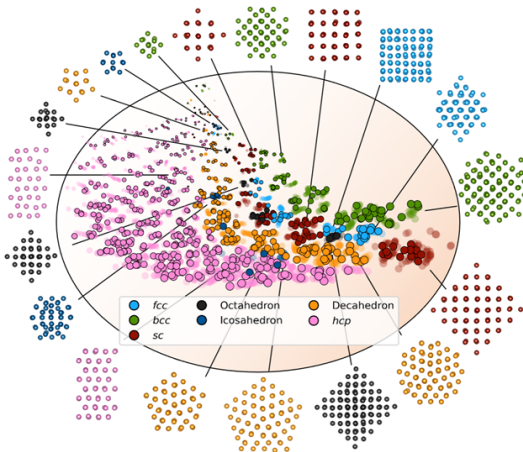


DEEPSTRUC: CLOSE PACKED NANOPARTICLES WITH A CVAE



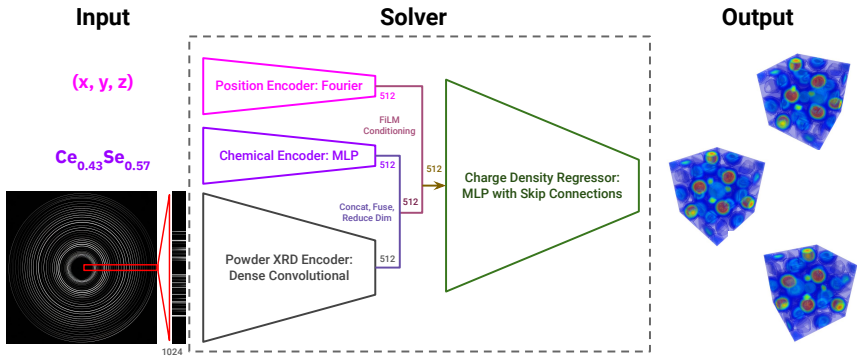


DEEPSTRUC: CLOSE PACKED NANOPARTICLES WITH A CVAE

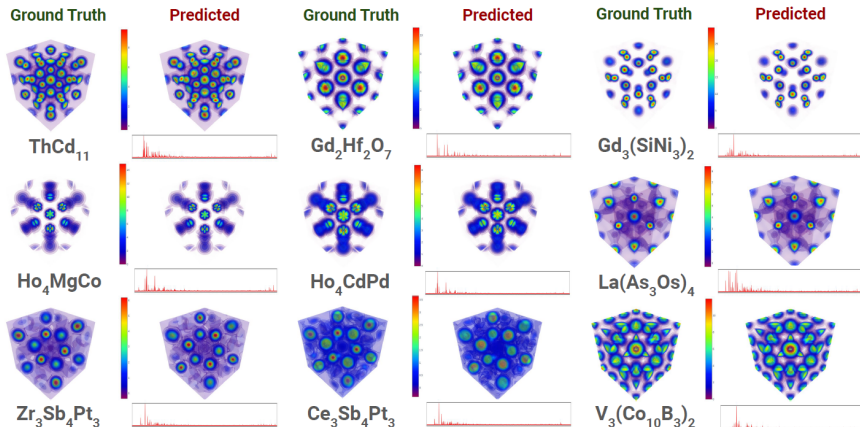


- Collaboration with Hod Lipson (Columbia University)
- Work of students Gabriel Guo and Ling Lan
- G. Guo *et al.* , npj-Computational Materials (2024)
10.1038/s41524-024-01401-8
- Graph Convolutional Variational Autoencoder (CVAE)
- input: powder diffraction pattern and composition
- output: electron density distribution
- Trained on cubic and trigonal (non-orthogonal but high symmetry) crystal systems

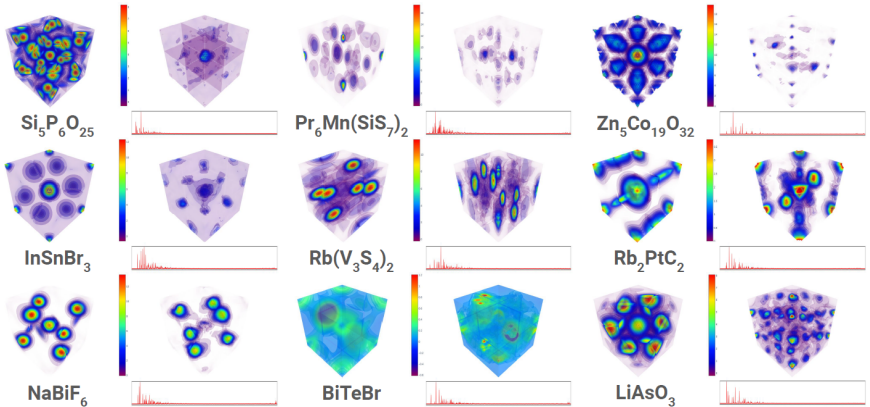
CRYSTALNET: ELECTRON DENSITY RECONSTRUCTION WITH CVAE



CRYSTALNET: ELECTRON DENSITY RECONSTRUCTION WITH CVAE



CRYSTALNET: ELECTRON DENSITY RECONSTRUCTION WITH CVAE



CASE 3: STRUCTURE SOLUTION OF NANOPARTICLES

The magic of diffusion models

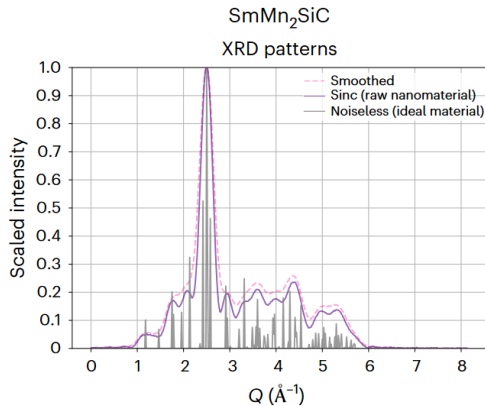
- Collaboration with Hod Lipson (Columbia University)
- Work of students Gabriel Guo and Tristan Saidi
- G. Guo *et al.* , Nature Materials, (2025)
<https://doi.org/10.1038/s41563-025-02220-y>
- input: nanostructure powder diffraction pattern and composition
- output: "structure": unit cell, coordinates and coloring



PXRDNet: STRUCTURE SOLUTION WITH DIFFUSION MODEL

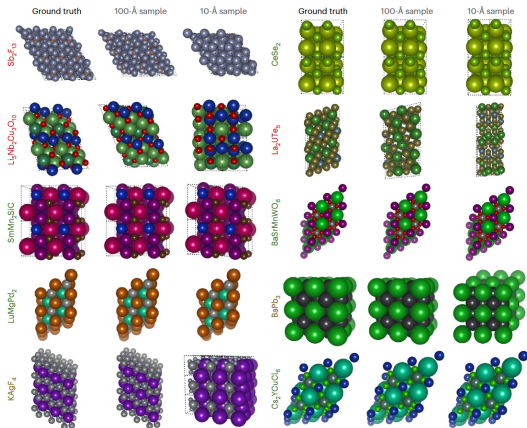
PXRDNet

- 100 Å or 10 Å NP powder data.
Low information content!
- Based on CDVAE (T. Xie, Proc. International Conference on Learning Representations (2022)).
- SE(3) equivariant graph NN autoencoder for composition, lattice parameters and number of atoms.
- Denoising diffusion via noise-conditioned score networks for the atomic coordinates

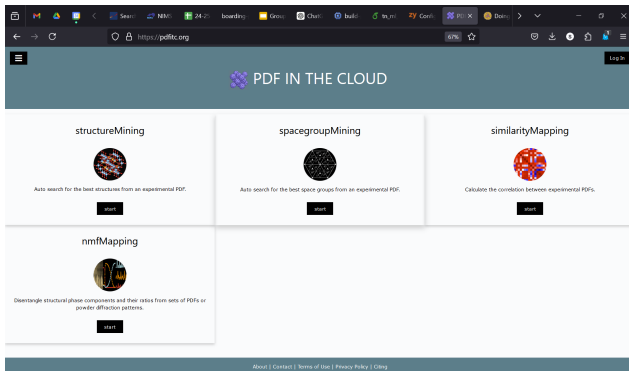




PXRDNET: STRUCTURE SOLUTION WITH DIFFUSION MODEL



DEPLOYING ML MODELS IN PRODUCTION



<https://pdfic.org>



DiffPy Community Publications Products

Please update your PDFgui to PDFgui v2.0, the file

[Get PDFgui v2.0](#)

[Credits](#)

DiffPy - Atomic Structure

A free and open source software project to provide a platform for the study of the atomic structure of materials.

DiffPy-CMI

xPDFsuite

PDFgetX3, PDFgetN3 and PDFgetS3

PDFgui

SrMise

mPDF

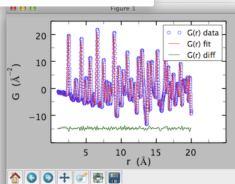
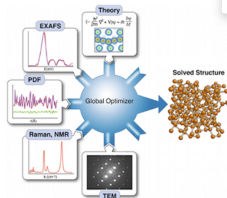
xiINTERPDF

Python Packages

10, 2023)

Python

the study of the atomic struc-



DiffPy is the home of the [DiffPy-CMI](#) Complex Modeling framework, a modular software framework for robust and extensible modeling of diffraction data. We welcome contributions to this project from the community.

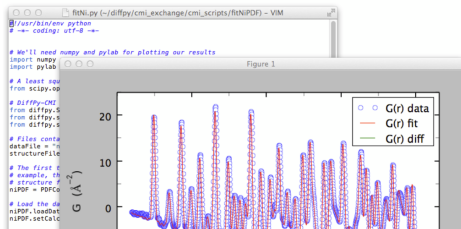
<https://diffpy.org> - home of Billinge-group software

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, nanomaterials, and amorphous materials.

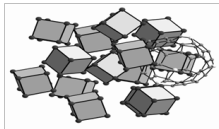
The software provides functionality for storage and manipulation of structure data and calculation of structure-based quantities, such as PDF, SAS, bond valence sums, atom overlaps, bond lengths, and coordinations. Most importantly the DiffPy-CMI package contains a fitting framework for combining multiple experimental inputs in a single optimization problem.

This is an early release of code that is under intense development, with support for installation on Unix, Linux, and Macintosh machines. The scope and documentation of the project 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 and suggestions for improvement, but please be patient and check back frequently for updates.

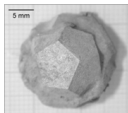


<https://diffpy.org - cmi>

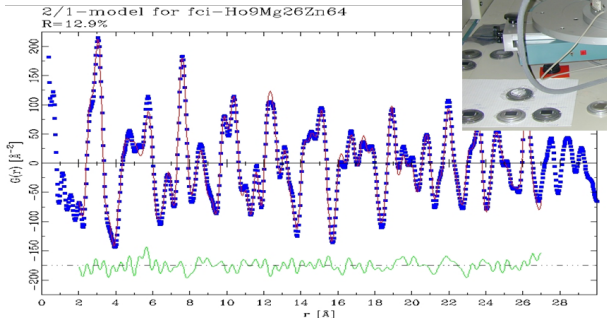
IT WAS ALWAYS POSSIBLE TO DO PDF IN THE LAB



fci-Ho-Mg-Zn



Huber Gunier diffractometer
 $Q_{\max} = 13.5 \text{ \AA}^{-1}$



Brühne et al., *Z. Kristallogr.* **219**
 (2004) 245-258

LAB PDF PROTOCOLS

 for updates

Chemistry—Methods

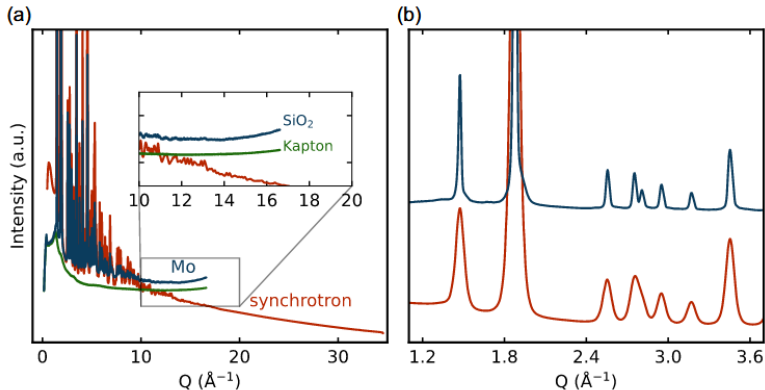
Research Article
doi.org/10.1002/cmtd.202500001 **Chemistry
Europe**
European Chemical
Societies Publishingwww.chemistrymethods.org

Testing Protocols for Obtaining Reliable Pair Distribution Functions from Laboratory X-Ray Sources Using PDFGETX3

Till Schertenleib,* Daniel Schmuckler, Yucong Chen, Geng Bang Jin,* Wendy L. Queen,
and Simon J. L. Billinge*

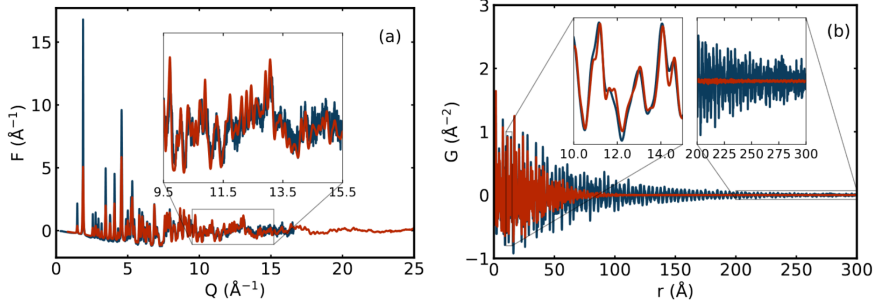
This is just our opinion, see also others' work, e.g., group of Miriam Zobel

COMPARING LAB PDF WITH SYNCHROTRON



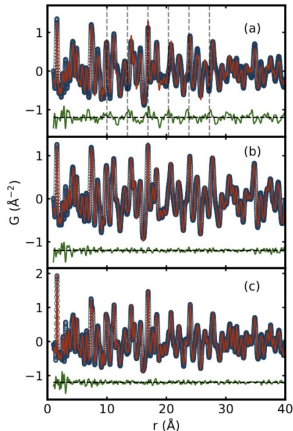
Synchrotron-lab-PDF comparison

THE MEASUREMENT METHOD AFFECTS THE RESULTING PDF



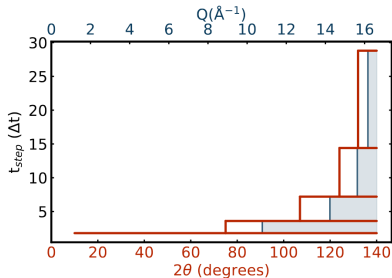
Comparison of $F(Q)$ and $G(r)$ functions of crystalline quartz collected via synchrotron measurement (red) and lab measurement (blue).

COMPARISON OF EXPERIMENTAL AND CALCULATED PDFs



- Lab data, no soler slit
- Lab data, 2.5 deg soler
- Synchrotron data

VARIABLE COUNTING: STAIRCASE SCHEME



exponential weighting of high-Q region

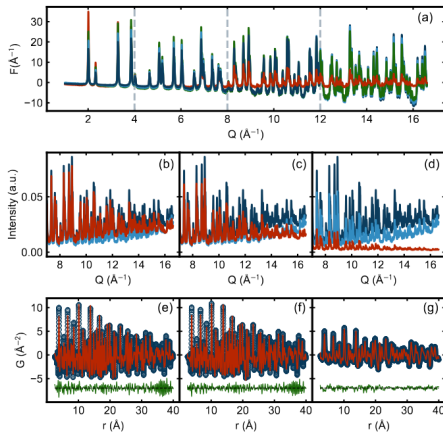
- step 0: -4 - 15
- step 1: 10 - 140
- step 2: 75 - 140
- step 3: 107 - 140
- step 4: 124 - 140
- step 5: 132 - 140

DATA REDUCTION FOR LAB PDF

A number of different options from the manufacturers. From Us

- DIFFPY.LABPDFPROC for multiplicative corrections
- followed by PDFGETX3
- Yucong Chen *et al.* , Absorption corrections. To be published in Crystal Growth & design (arXiv:2504.12499)

Also, look out for DiffractionObjects in DIFFPY.UTILS



LAB PDF IN THE 21ST CENTURY

Lab PDF can become a standard technique in every chemistry laboratory

- basic acquisition down to hours and minutes from days
- High quality PDFs very straightforwardly
- Data reduction and analysis protocols are close to being automated and straightforward

ACKNOWLEDGEMENTS



- My current and former students and post-docs
- Beamline and software teams
- Collaborators
- Funding (DOE-BES, NSF-DMR, TRI, Columbia-DSI)
- The Facilities!