Artificial Intelligence for Scattering Experiments

Thomas Proffen

Neutron Scattering Division

About me

- PhD in Crystallography from Ludwig Maximilians Universität in Munich, Germany
- Postdoc at the Australian National University in Canberra, Australia
- Postdoc at Michigan State University
- Instrument scientist at Los Alamos National Laboratory
- Diffraction Group Leader at Oak Ridge National Laboratory (SNS and HFIR)
- Director Neutron Data Analysis and Visualization at ORNL.
- Distinguished Staff Member and Director Science Initiative High Performance Computing, Modeling and Data Analytics.
- Founder of Oak Ridge Computer Science Girls.

My car

Forschungsreaktor München

NPDF at Lujan Center (LANL)

First Neutron Scattering Paper ..

Acta Cryst. (1993). **B49**, 599–604

Defect Structure and Diffuse Scattering of Zirconia Single Crystals Doped with 7 mol% CaO

BY TH. PROFFEN, R. B. NEDER AND F. FREY
Institut für Kristallographie und Mineralogie, Theresienstrasse 41, 8000 *München* 2, *Germany*

AND W. ASSMUS
Physikalisches Institut der Universität Frankfurt, Germany

(*Received* 21 *September* 1992; *accepted* 4 *January* 1993)

This layer of diffuse scattering took **several month** to collect – 180 x 120 points, ~10 min per point

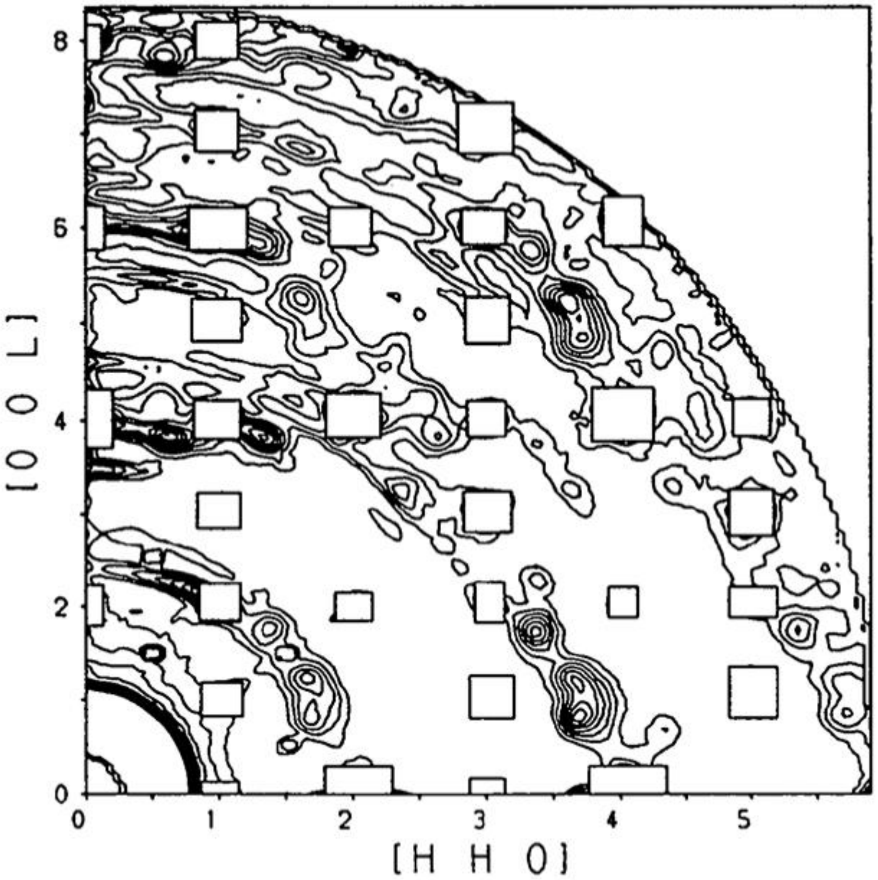
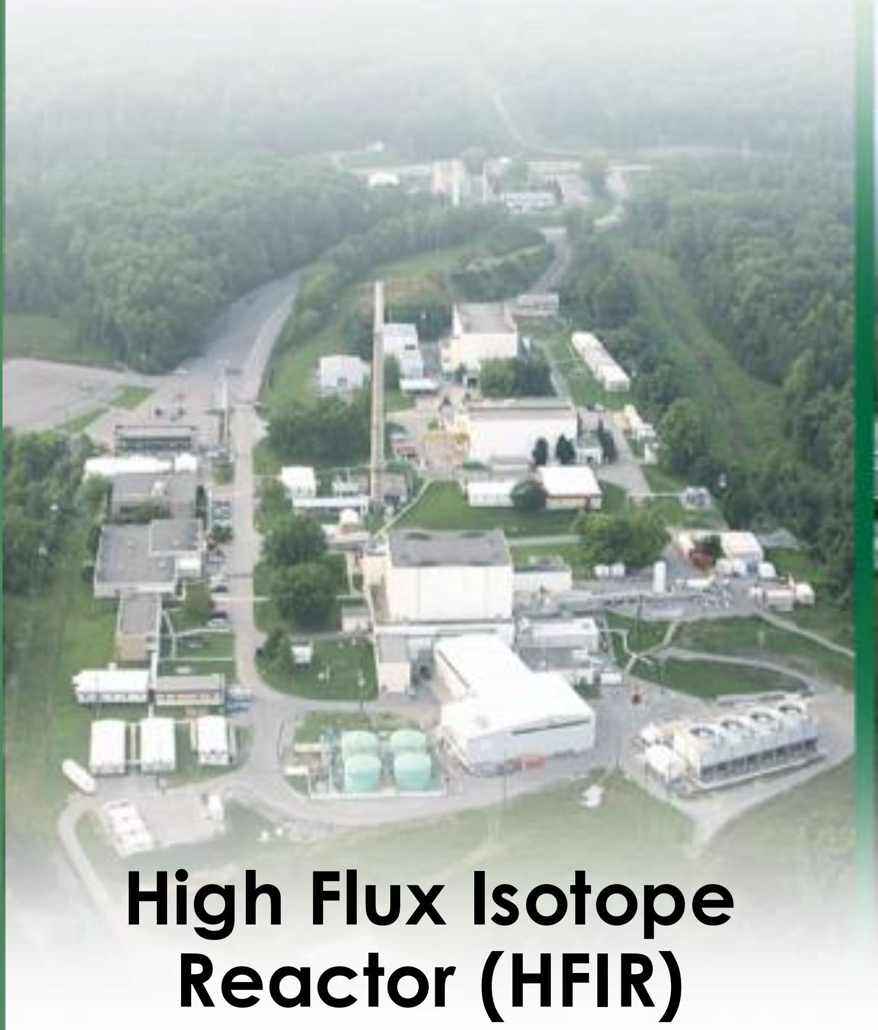


Fig. 1. Zero layer of the $[1\bar{1}0]$ zone. The intensities are stepped with linear intervals of 25 counts, the lowest intensity represented is 125 counts.

ORNL is home to two world class neutron sources



High Flux Isotope Reactor (HFIR)

690	685	95,469	1,190
Scientific Publications in CY21*	Total Unique Users in FY21	Hours of Beamtime in FY21	Total Experiments in FY21

Spallation Neutron Source (SNS)

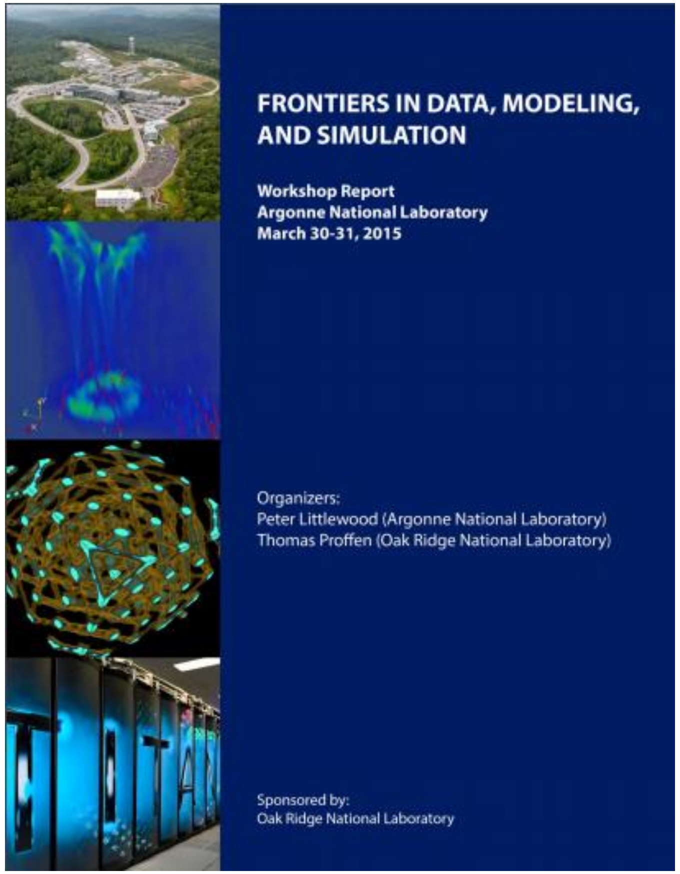
Materials research crosses facilities

Argonne NATIONAL LABORATORY

OAK RIDGE National Laboratory

Opportunities

- Multimodal analysis
- Applied Math. concepts
- Advanced Materials Modeling



FRONTIERS IN DATA, MODELING, AND SIMULATION

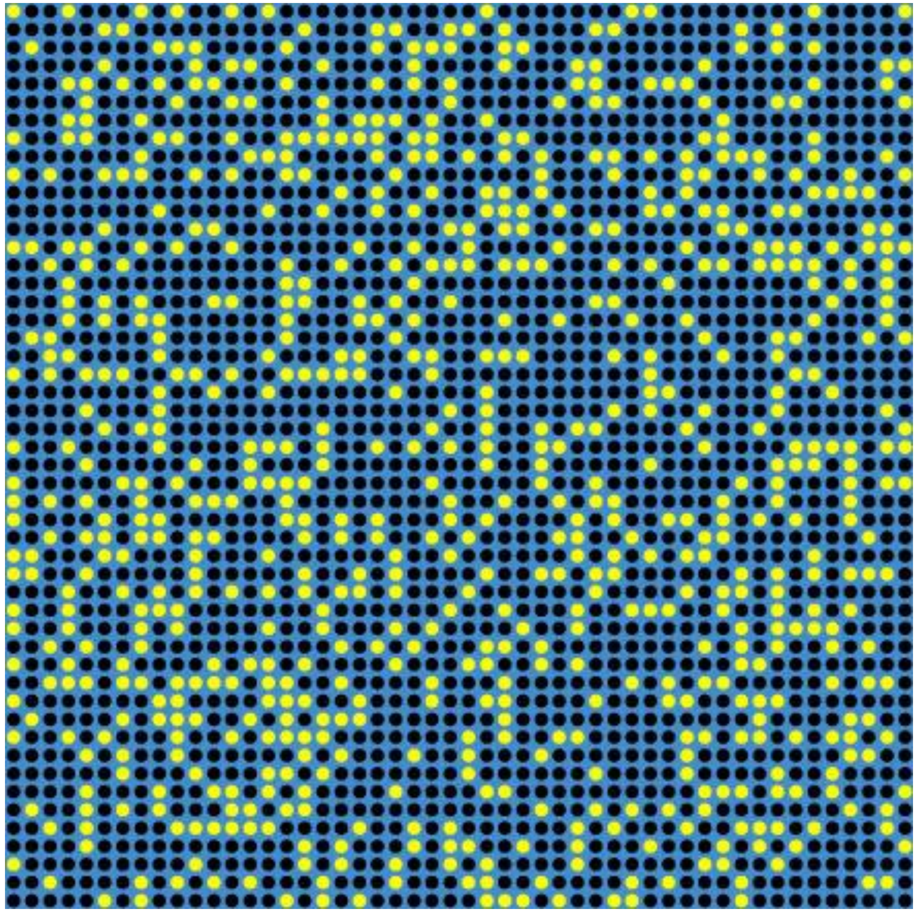
Workshop Report
Argonne National Laboratory
March 30-31, 2015

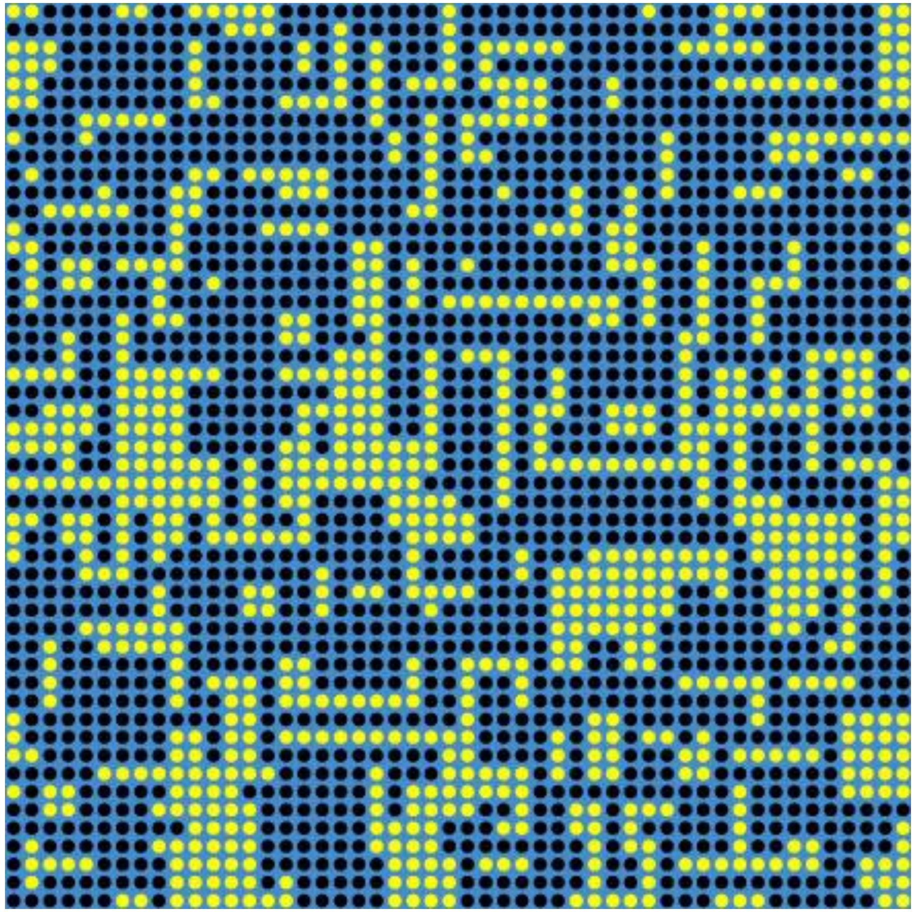
Organizers:
Peter Littlewood (Argonne National Laboratory)
Thomas Proffen (Oak Ridge National Laboratory)

Sponsored by:
Oak Ridge National Laboratory

http://neutrons.ornl.gov/grand-challenge-workshops

Diffuse scattering ?



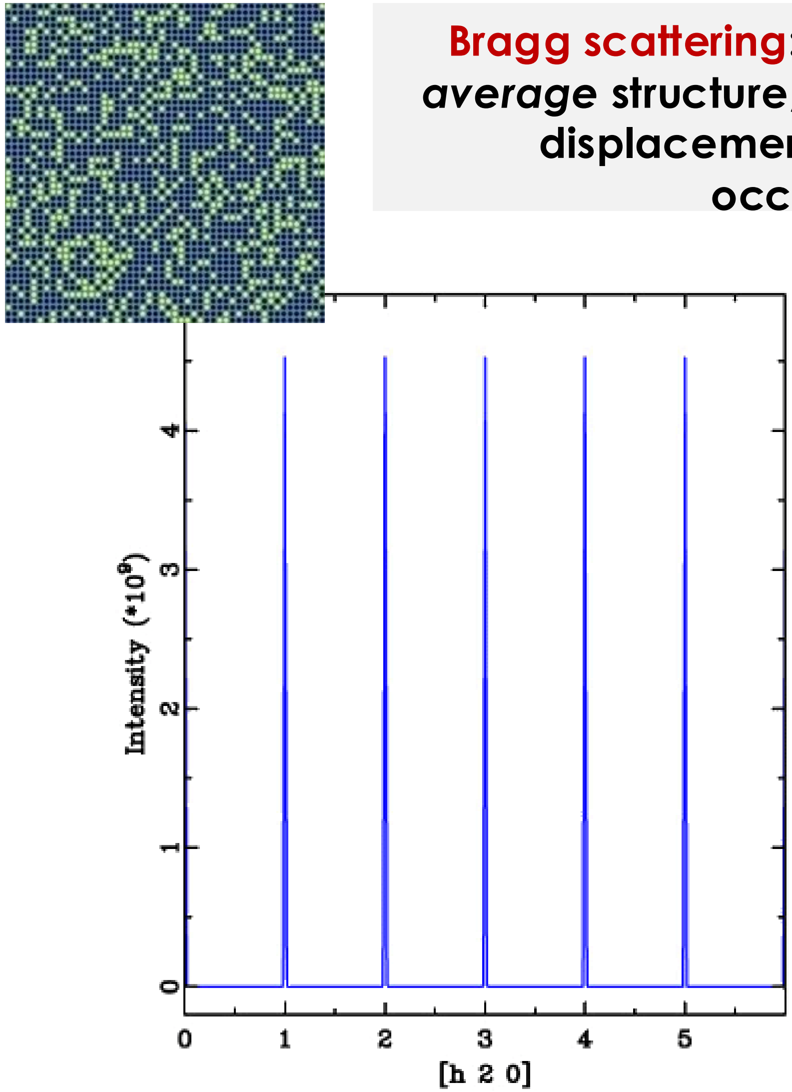


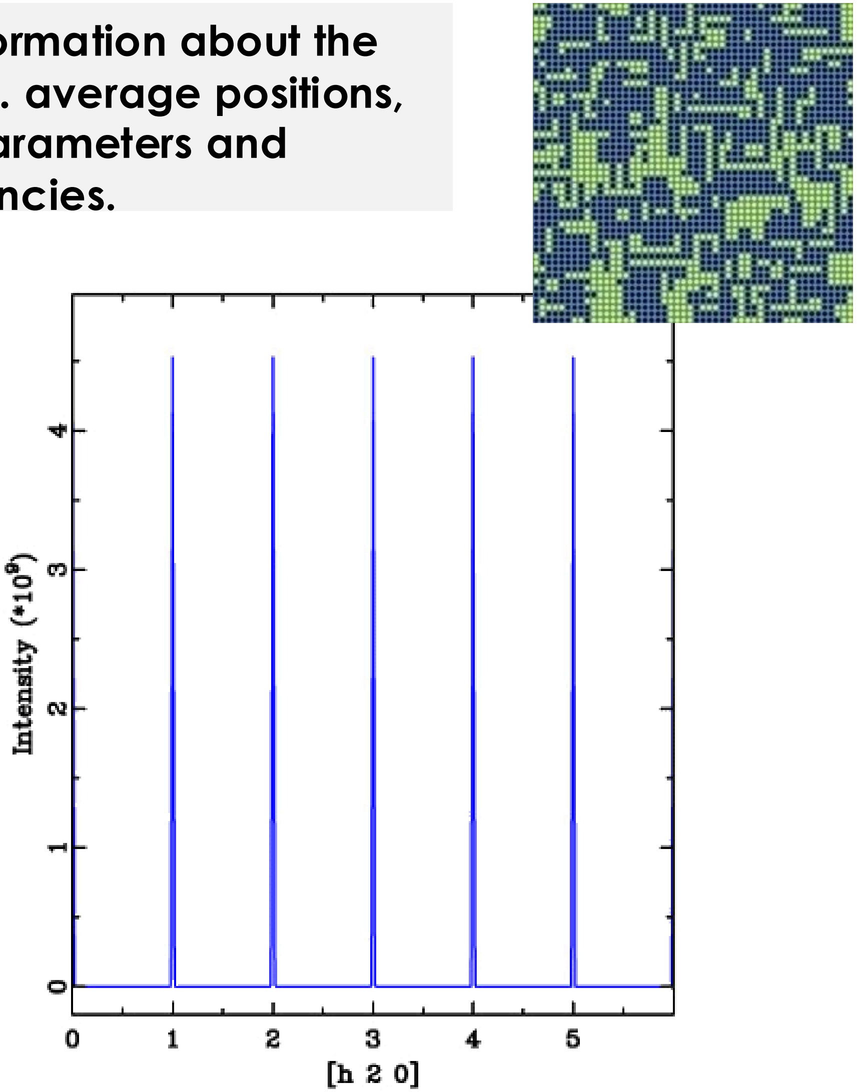
Cross section of 50x50x50 u.c. model crystal consisting of 70% black atoms and 30% *vacancies* !

Properties might depend on vacancy ordering !!

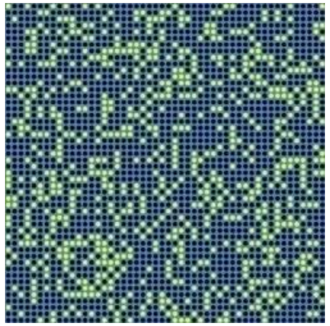
Bragg peaks are blind ..

Bragg scattering: Information about the *average* structure, *e.g.* average positions, displacement parameters and occupancies.

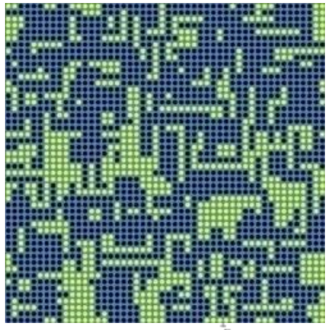


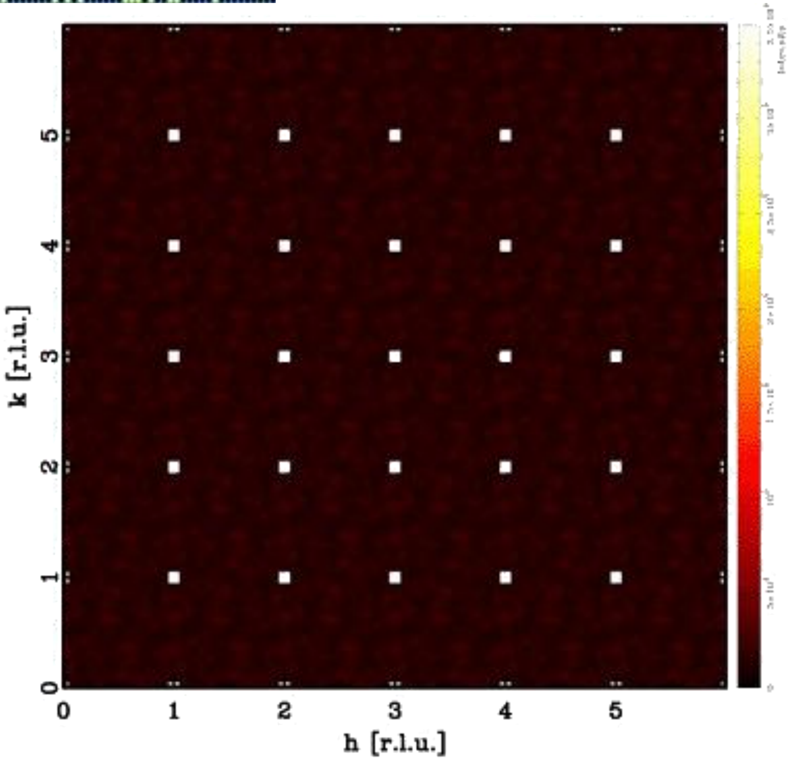


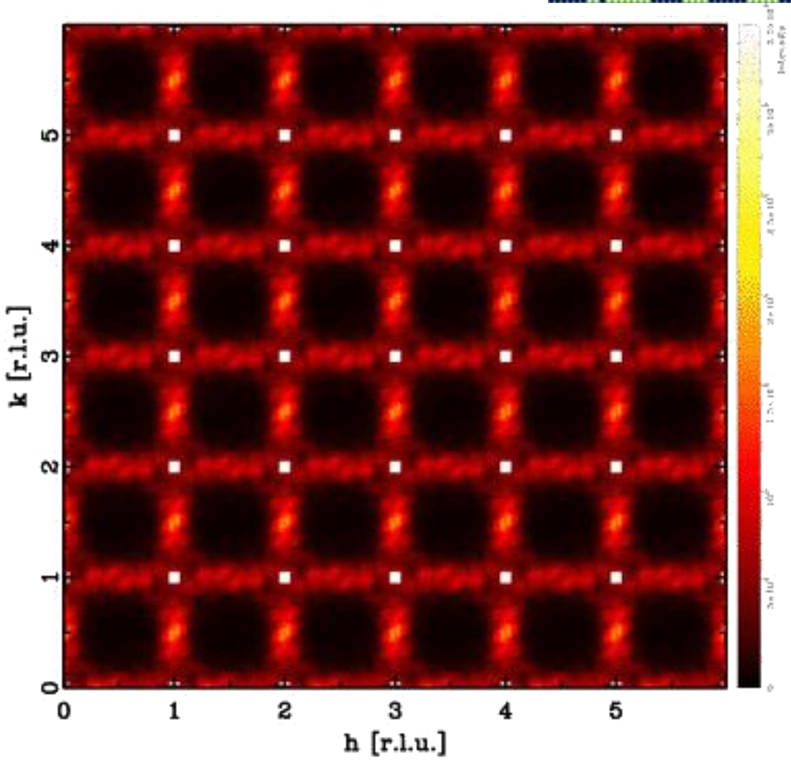
Diffuse scattering to the rescue ..



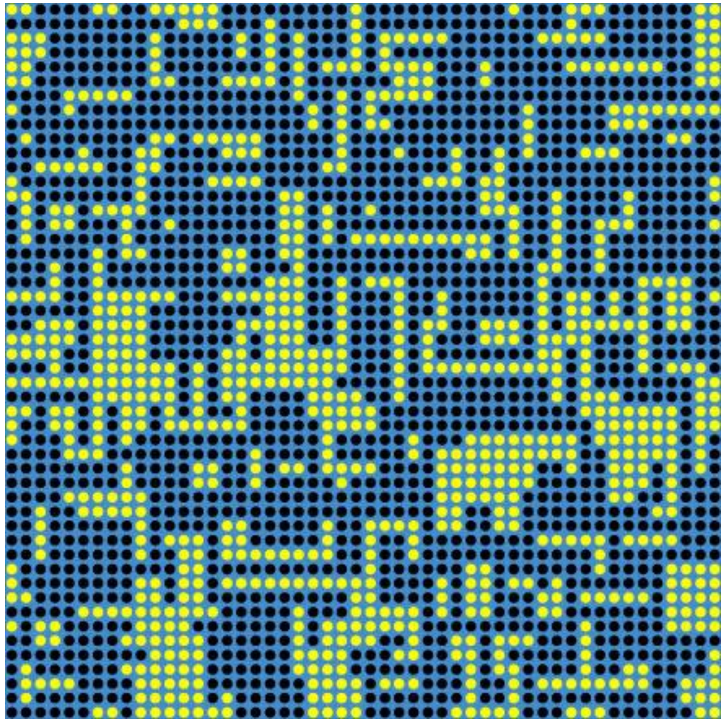
Diffuse scattering: Information about *two-body correlations*, *i.e.* chemical short-range order or local distortions.



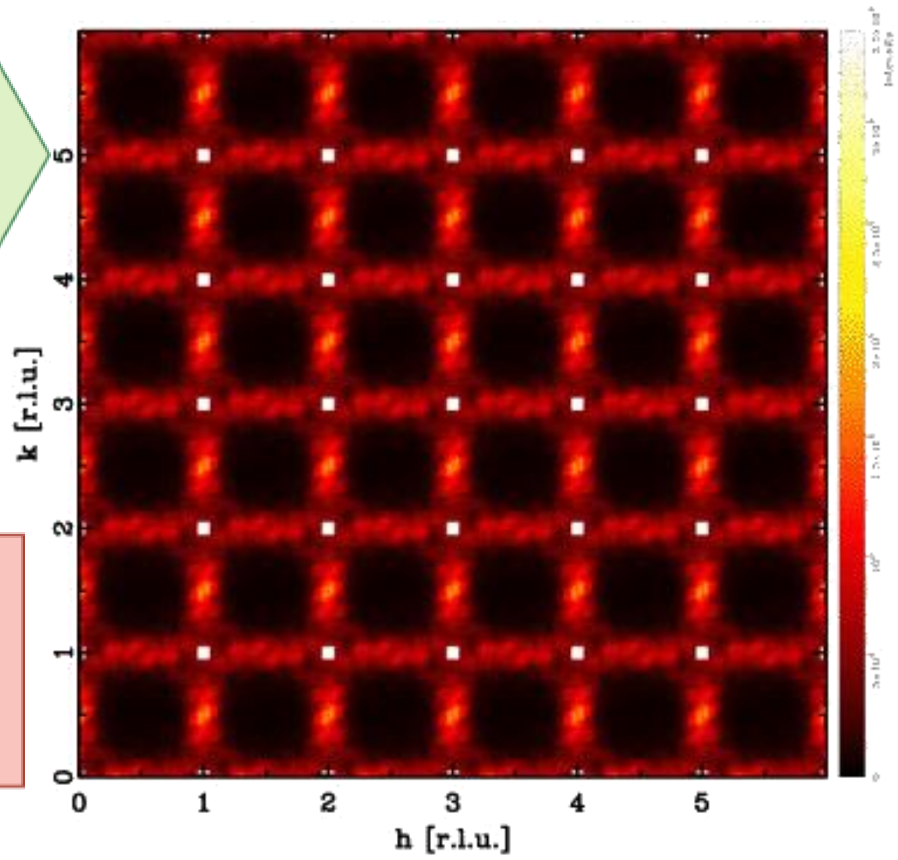




Inverse Problem *aka* Crystallographic Phase Problem



$$F(\mathbf{h}) = \sum_{i=1}^{N} f_i(\mathbf{h}) e^{2\pi i \mathbf{h}\mathbf{r}_i}$$



Intensities measured only give |**F**| and not the phase

Analyzing diffuse scattering

- **Correlation approach**: Expansion of kinematic scattering equation in terms of displacement. Yields set of two-body correlations.

- **Monte Carlo based computer simulations**: Scientist might "win" solution to the problem ..

  – Minimize total energy E: AMC
  – Minimize (observed – calculated)^2: RMC

- More: "Diffuse Neutron Scattering from Crystalline Materials" by Nield and Keen, Oxford University Press

Table 1. Summary of the properties of the different components of the diffuse intensity.

Term	$\mathbf{I_0}$	$\mathbf{I_1}$	$\mathbf{I_2}$	$\mathbf{I_3}$
Description	Short-range order (SRO) term	Warren Size-effect	Huang Scattering 1st order TDS	3rd order size term
Lattice averages involved	SRO parameters α^{ij}	$\langle X^{ij}\rangle, \langle Y^{ij}\rangle$ *etc.*	$\langle (X^{ij})^2\rangle$, $\langle X^{ij}Y^{ij}\rangle$ *etc.*	$\langle (X^{ij})^3\rangle$, $\langle (X^{ij})^2 Y^{ij}\rangle$ *etc.*
Type of Summation	cosine	sine	cosine	sine
Symmetry	symmetric	anti-symmetric	symmetric	anti-symmetric
Variation in $\boldsymbol{k}$–space	nil	linear, *i.e.* with h_1, h_2 *etc.*	quadratic, *i.e.* with $h_1^2, h_1 h_2$ *etc.*	cubic, *i.e.* with $h_1^3, h_1^2 h_2$ *etc.*
Dependence on f_A, f_B for binary	$(f_A - f_B)^2$	$f_A(f_A - f_B)$, $f_B(f_A - f_B)$	$f_A^2, f_A f_B, f_B^2$	$f_A^2, f_A f_B, f_B^2$
Number of components for binary	1	6	18	30

The Automatic Monte Carlo Method

Input:

- Observed diffuse scattering
- Starting structure (e.g. average)
- Model for disorder in terms of interaction energies for MC simulation.

Result:

- Set of interaction energies for given model that best match the data.

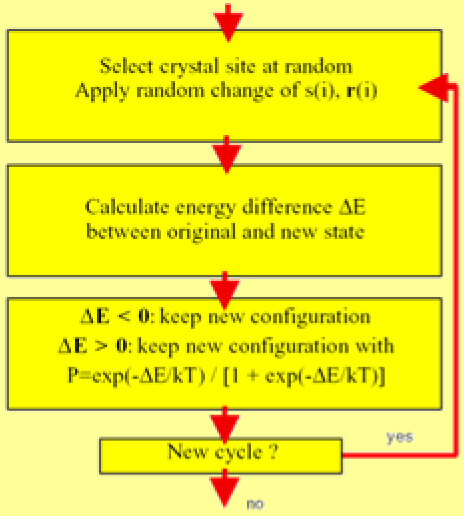
Questions:

- Finding the right model ..
- It is very slow ..

Least Squares

For each parameter in MC model

Run MC



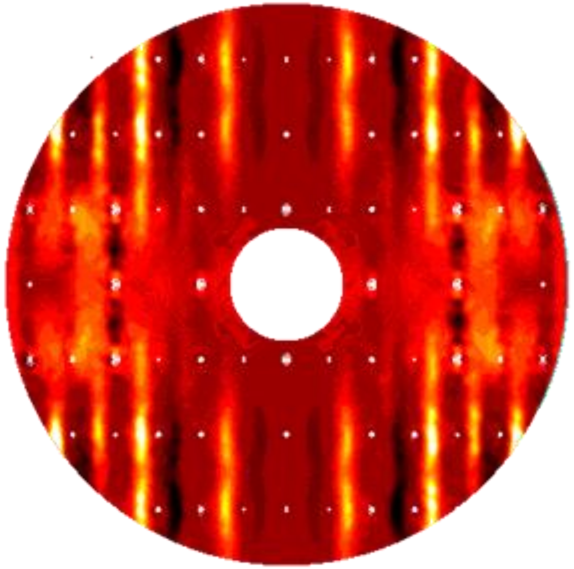
Calculate diffuse scattering

Obtain derivatives

Disorder in $Fe_3(CO)_{12}$ – AMC refinement

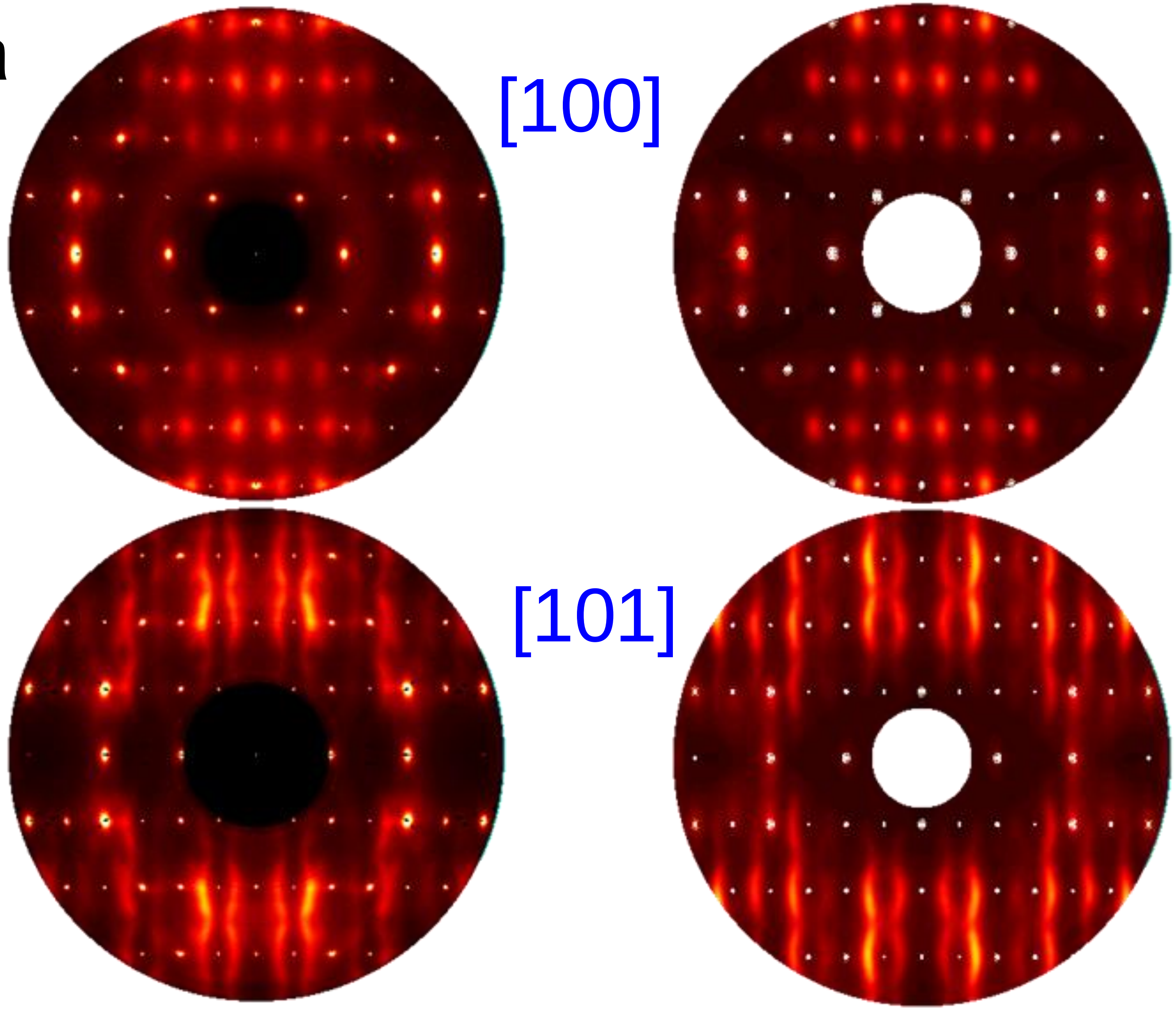
Numerical estimates
of Differentials

$$\frac{\partial \Delta I}{\partial p_i} = \sum_{hklm} \frac{\left(\Delta I_{p+} - \Delta I_{p-}\right)}{2\delta_i}$$



Difference between two
calculated diffraction patterns

Data **calculated**

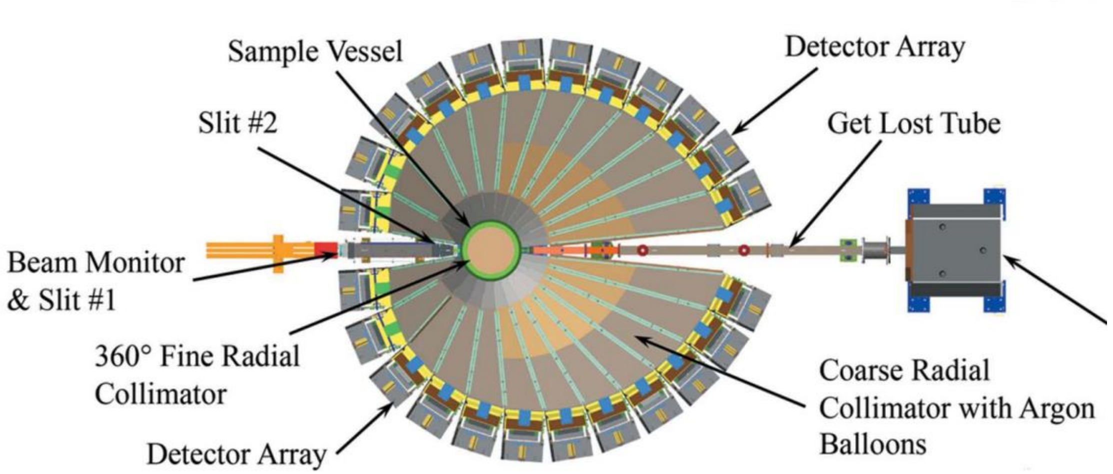


Opportunities using Machine Learning

AI is about how we use and process data. It will be, and is, transformative in knowledge-based disciplines. AI will not replace scientists, but scientists who use AI will replace those who don't.*

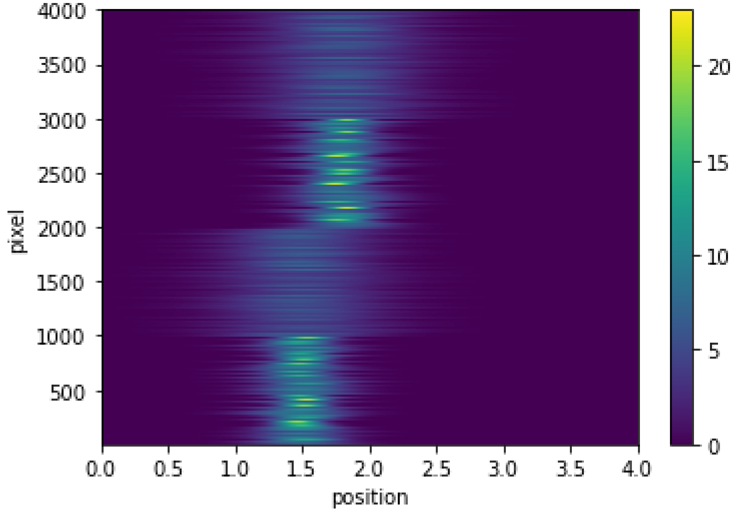
*Modified from a quote in the Microsoft report, "The Future Computed: Artificial Intelligence And Its Role In Society"

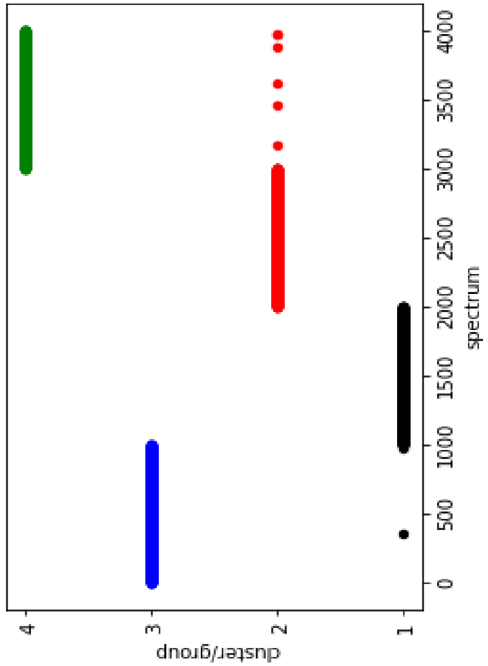
Unsupervised Machine Learning – Instrument calibration



Unsupervised clustering algorithm for Time focusing and selection of groups of detectors with ‘similar’ features, e.g. resolution

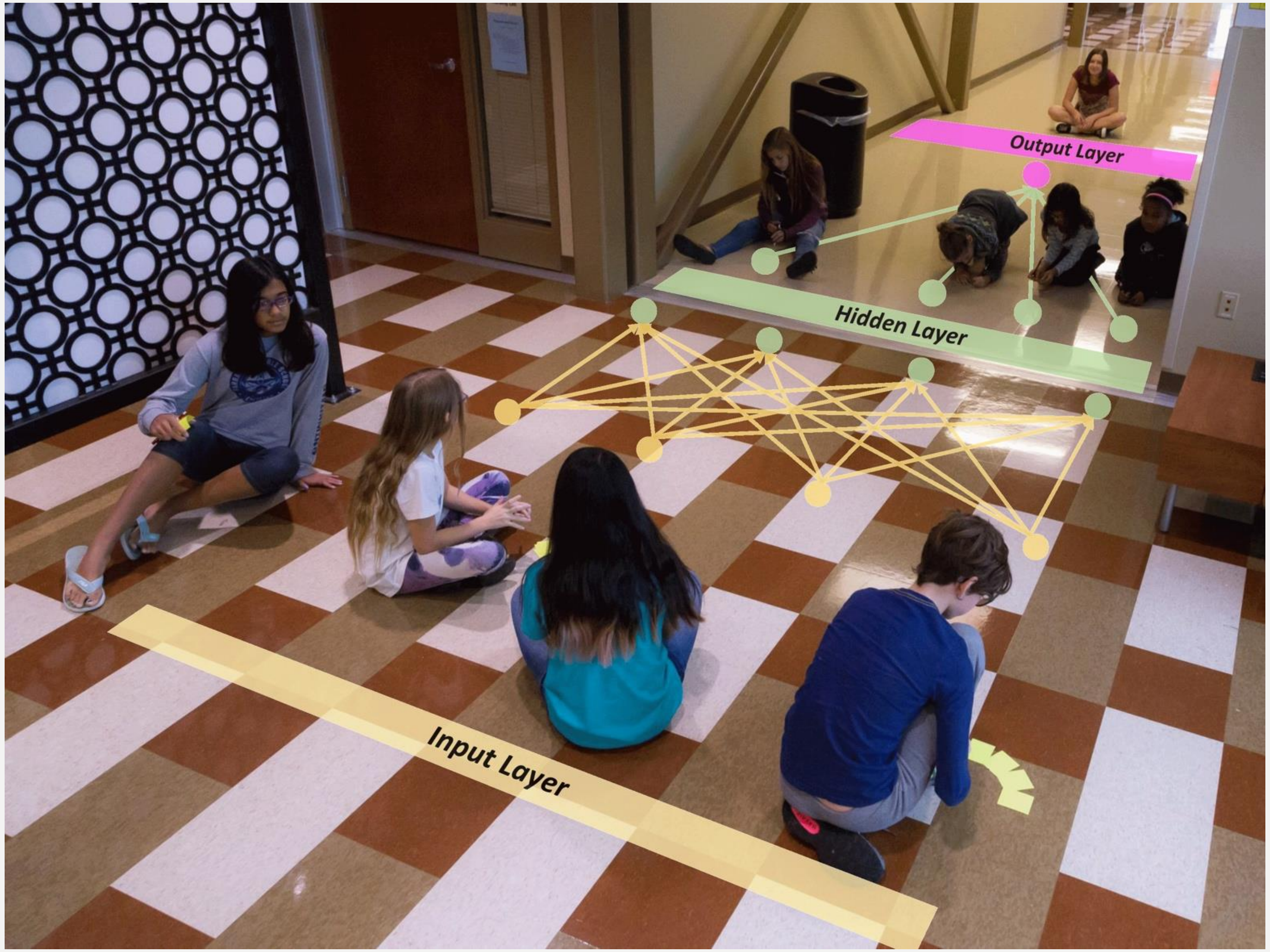
POWGEN





Work by Yuanpeng Zhang

(SUPERVISED) MACHINE LEARNING



ORCS Girls
www.orcsgirls.org

MACHINE LEARNING

A machine learning method takes a bunch of data and "learns" from it!

Data

Label: Not a pig

Label: Pig

Label: Pig

Label: Not a pig

Machine Learning Method

Machine's rules for determining whether or not an image is a pig

DID IT “LEARN” SOMETHING?

Label: Not a pig

Label: Pig

Label: Pig

Label: Not a pig

Training Data

The data we give to the machine learning method to learn from

Label: Not a pig

Label: Pig

Testing Data

The data we hold out and use to check to see if the method actually learned something!

DEEP LEARNING

Simulated scattering 'images'

- Small Angle Scattering
- Diffraction
- Diffuse Scattering
- Quasi Elastic Scattering

Labels

- Relate to model / parameters
- Related to topology
- Good/Bad

Training Data

The data we give to the machine learning method to learn from

Testing Data

The data we hold out and use to check to see if the method actually learned something!

Machine Learning for classification

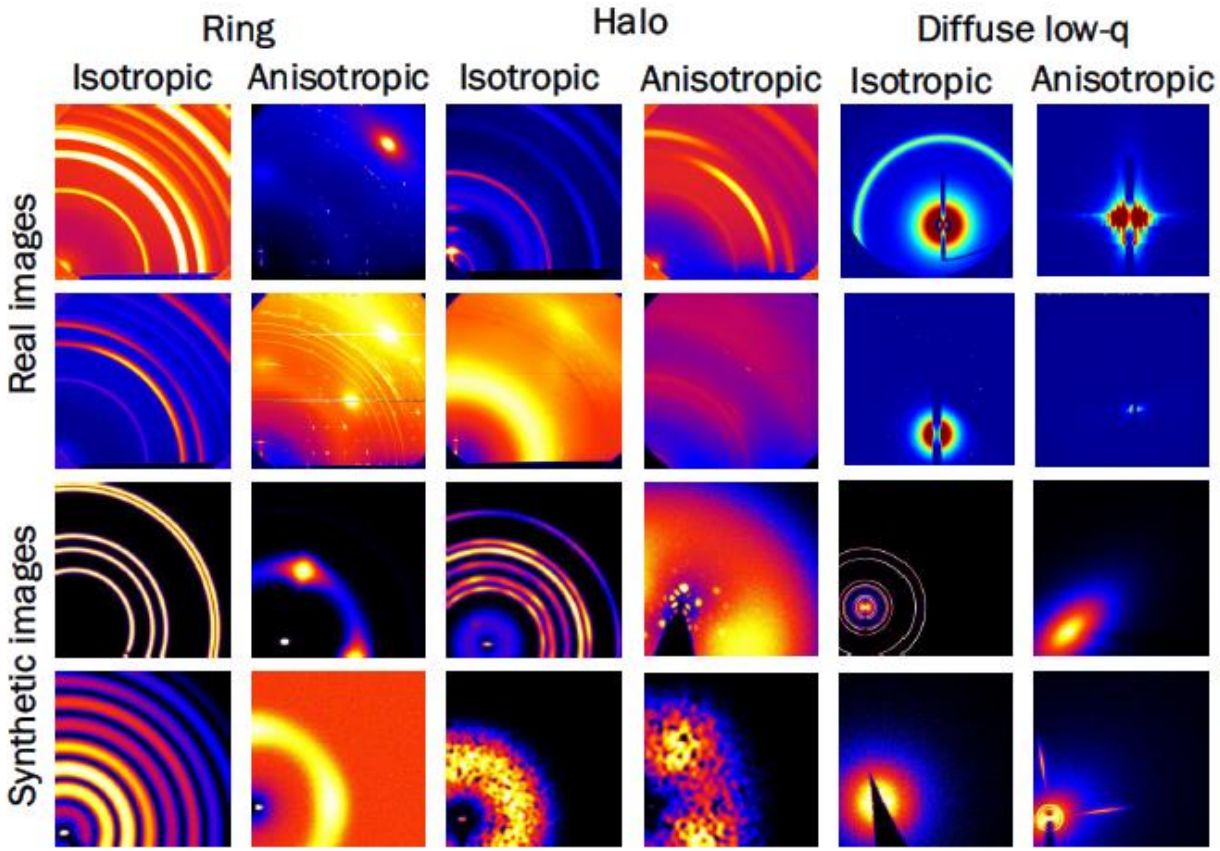


Figure 2: Comparison between synthetic images and real experimental images. The first and second rows are real experimental images, while the third and forth rows are synthetic images. Images in the same column have the same attribute. From left to right, the attributes are: Ring: Isotropic, Ring: Anisotropic, Halo: Isotropic, Halo: Anisotropic, Diffuse low q: Isotropic, and Diffuse low q: Anisotropic. Visually, synthetic and real images are indiscernible.

2017 IEEE Winter Conference on Applications of Computer Vision

X-ray Scattering Image Classification Using Deep Learning

Boyu Wang[1], Kevin Yager[2], Dantong Yu[2], and Minh Hoai[1]
[1]Stony Brook University, Stony Brook, NY, USA
{boywang, minhhoai}@cs.stonybrook.edu
[2]Brookhaven National Laboratory, Upton, NY, USA
{kyager,dtyu}@bnl.gov

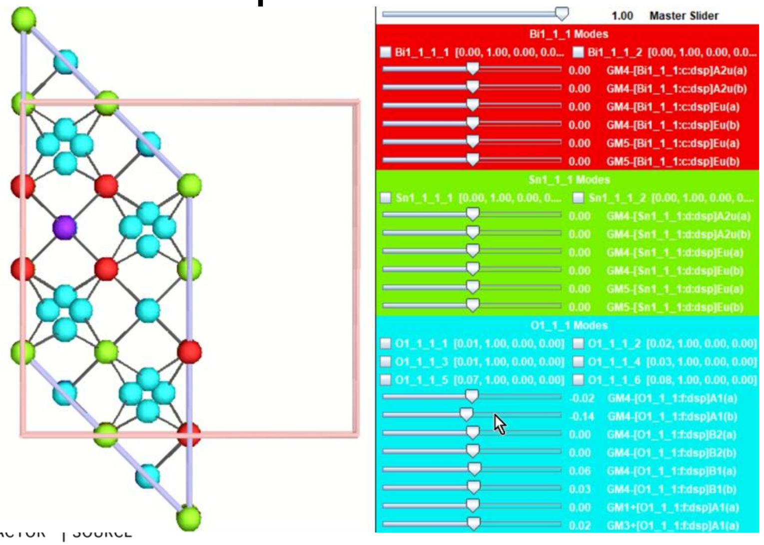
XsymNet:
ML + Exhaustive Symmetry for Phase Transitions

Objectives with XsymNet

- Lower barrier for subtle or complex phase transition studies
- Identify SG, lattice parameters, and distortions modes from powder diffraction data

Exhaustive Symmetry - ISODISTORT

- Provides symmetry adapted distortion modes to model the phase transition



XsymNet Workflow

1) Generate Subgroup tree (SGT) with ISODISTORT Method 3
2) Create 250-1000 perturbations of each subgroup member by randomly choosing:
   - Strain Mode Amplitudes
     - 1 to 6 modes depending on symmetry
     - Random(-0.01, 0.01)
   - Displacement Mode Amplitudes
     - Gaussian(0, $\sigma = 0.33$)
   - BEQ Intensity – Thermal Parameters
3) Simulate powder patterns of all perturbed structures
4) Train XsymNet to classify powder patterns by subgroup symmetry
5) Classify Experimental diffraction data

Slides thanks to Dayton G. Kizzire - kizziredg@ornl.gov

XsymNet: ML + Exhaustive Symmetry for Phase Transitions

Slides thanks to Dayton G. Kizzire - kizziredg@ornl.gov

XsymNet – Convolutional Neural Network

- Accurately classifies subgroup symmetry to powder patterns
- Automated Rietveld refinement on top 5 subgroups → scientist reviews results

Simulated Validation Data

Classification	Accuracy Metric	α phase	β phase
Subgroup (547 classes)	Top 1	89.2%	87.5%
	Top 5	99.5%	98.2%

Experimental Data – $Bi_2Sn_2O_7$

Confidence Rank	α phase	β phase
1	0176	**0152**
2	**0088**	0077
3	0236	0383
4	0544	0169
5	0183	0170

Parent Symmetry CIF → ISODISTORT Subgroup Tree → XsymNet Classification → Rietveld Refinement of Top 5 → Scientist Review

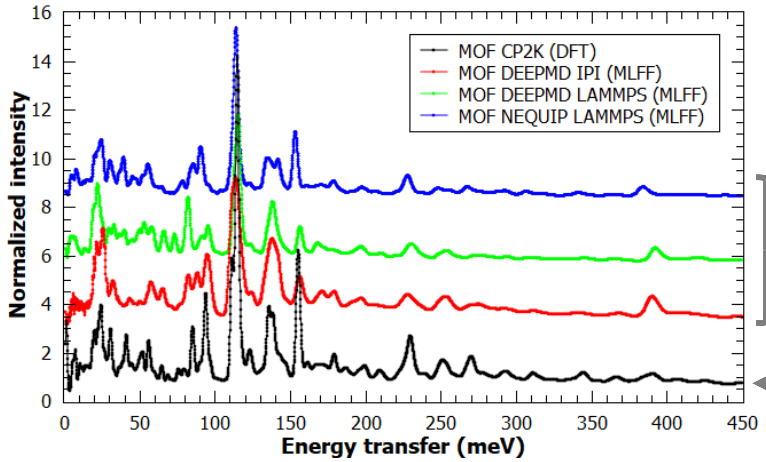
Slides thanks to Yongqiang Cheng - chengy@ornl.gov

Machine learning force fields (MLFFs) for neutron scattering

Task:	DFT calculations to generate training datasets	Training of MLFFs	Atomistic modeling with MLFFs	Neutron scattering simulation	Analysis, visualization, and interpretation
Software:	VASP/CP2K/etc.	DeePMD/NequIP	LAMMPS/i-PI/etc.	OCLIMAX	Mantid/Dave/etc.
Hardware:	CADES/HPC	DGX	Analysis/PC	Analysis/PC	Analysis/PC

DeepMD: Zhang et al. Phys. Rev. Lett. 120, 143001 (2018)
NequIP: Batzner et al. https://arxiv.org/abs/2101.03164 (2021)

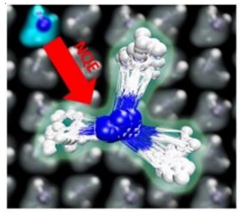
✓ Simulation of vibration and INS spectra of complex materials

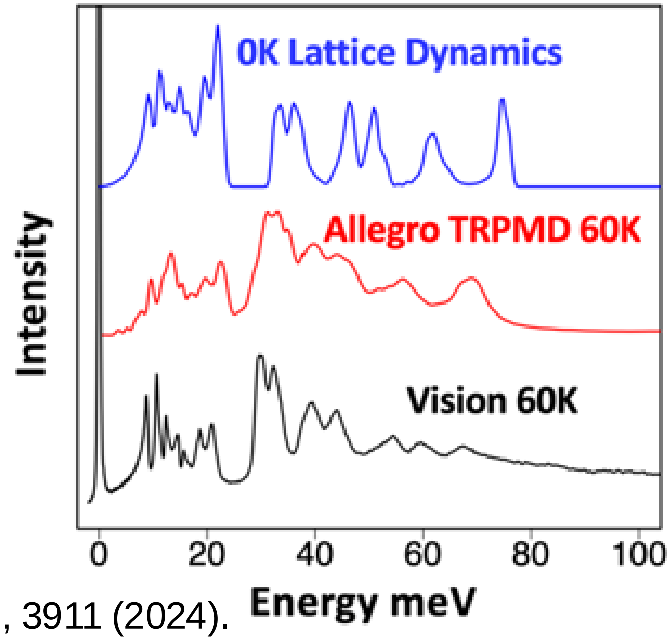


MLFF: Minutes on PC

DFT: Days on CADES

✓ Nuclear quantum effects in spectroscopy



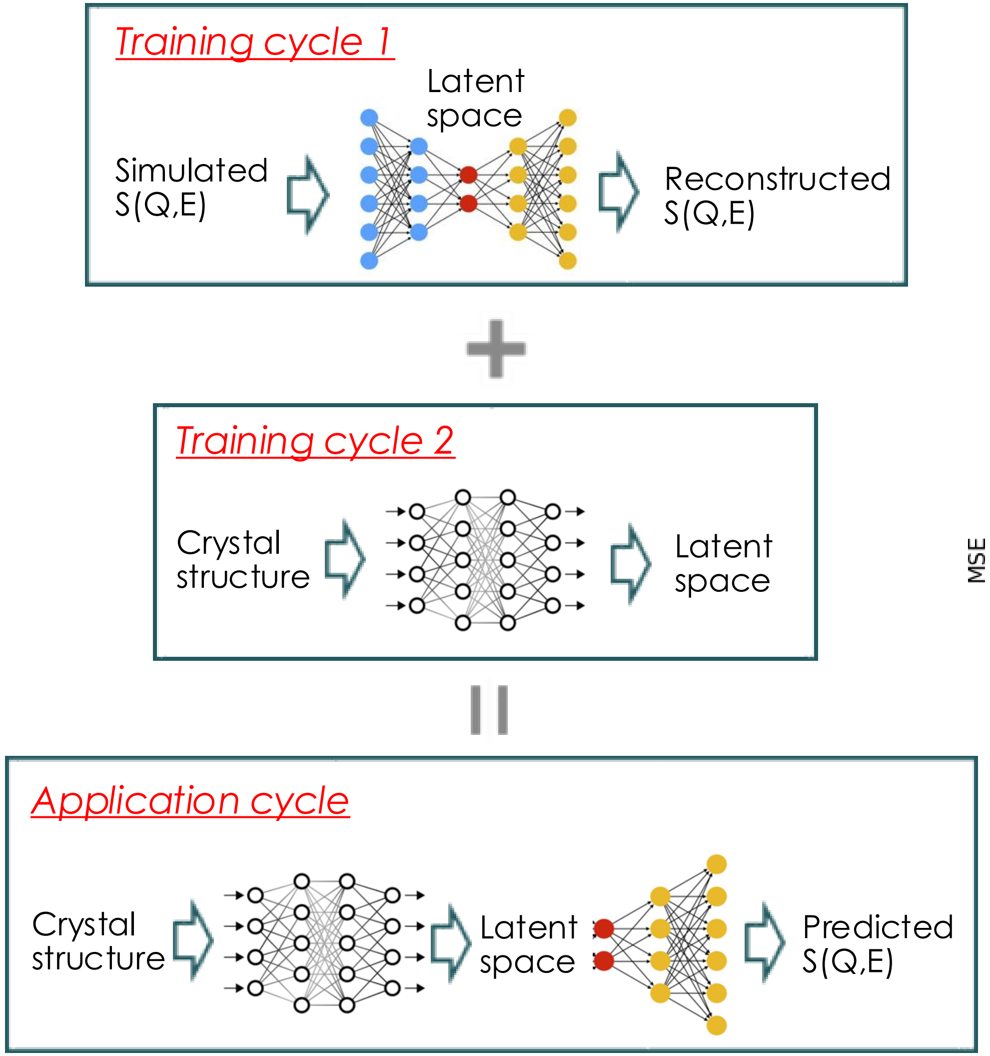


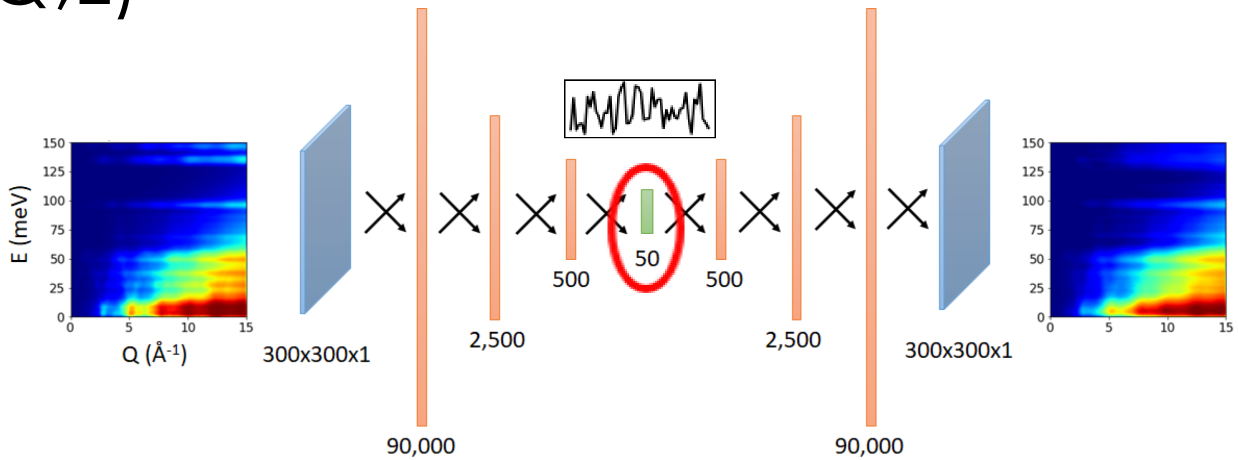
Linker, T.M. et al. Nat Commun 15, 3911 (2024).

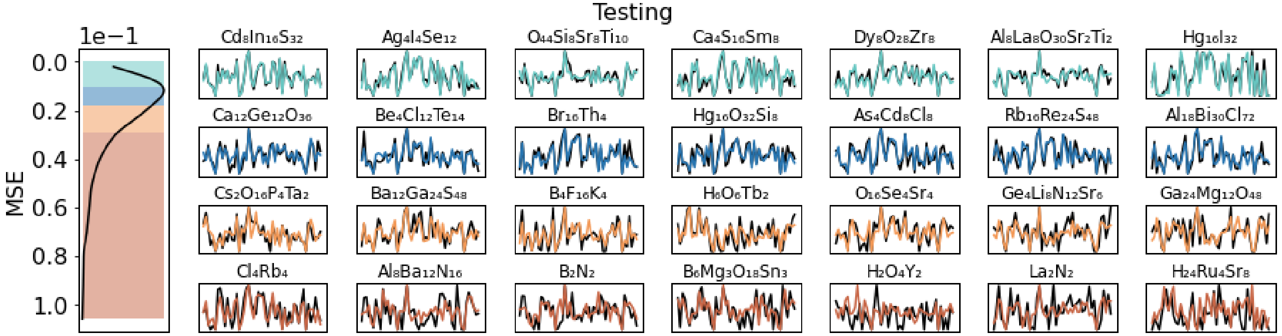
10,000 speedup and linear scaling with size, while inheriting spectroscopic accuracy from DFT:

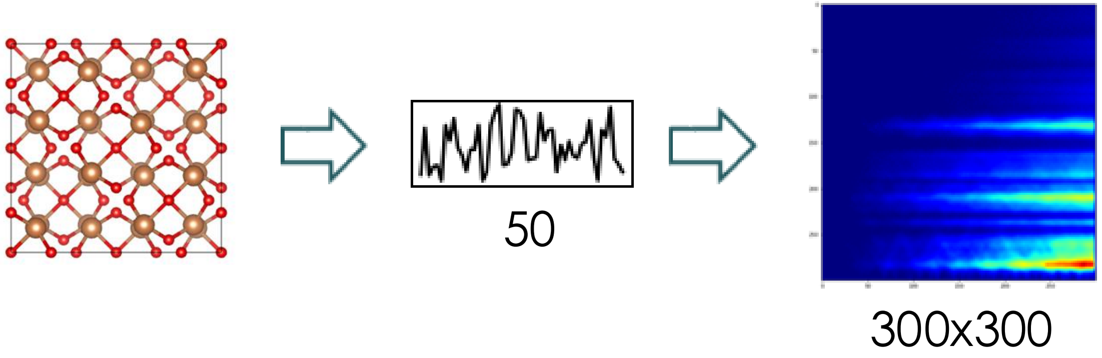
- Disordered, defective, or distorted crystals
- Heterogeneous structure (interface, boundary, guest-host systems)
- Long-range correlations
- Slow dynamics and rare events
- Nuclear quantum effects

Direct prediction of powder S(Q,E)









Cheng, Y., et al. Mach. Learn.: Sci. Technol. 4, 015010 (2023).

Slides thanks to Yongqiang Cheng - chengy@ornl.gov

Analysis and feature detection in large volumes of diffuse x-ray and neutron scattering from complex materials

Thomas Proffen, Ray Osborn, Rick Archibald, Stuart Campbell, Ian Foster, Scott Klasky, Tashin Kurc, Dave Pugmire, Michael Reuter, Galen Shipman, Chad Steed, Chris Symons, Ross Whitfield, Doug Fuller, Guru Kora, Mike Wilde, Justin Wozniak

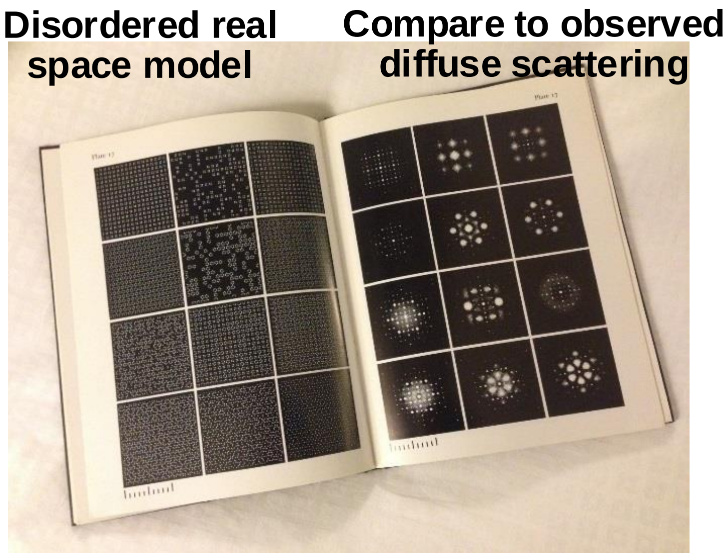
Facilities/Resources
SNS, APS, ALCF; OLCF; and CADES at ORNL

Argonne NATIONAL LABORATORY

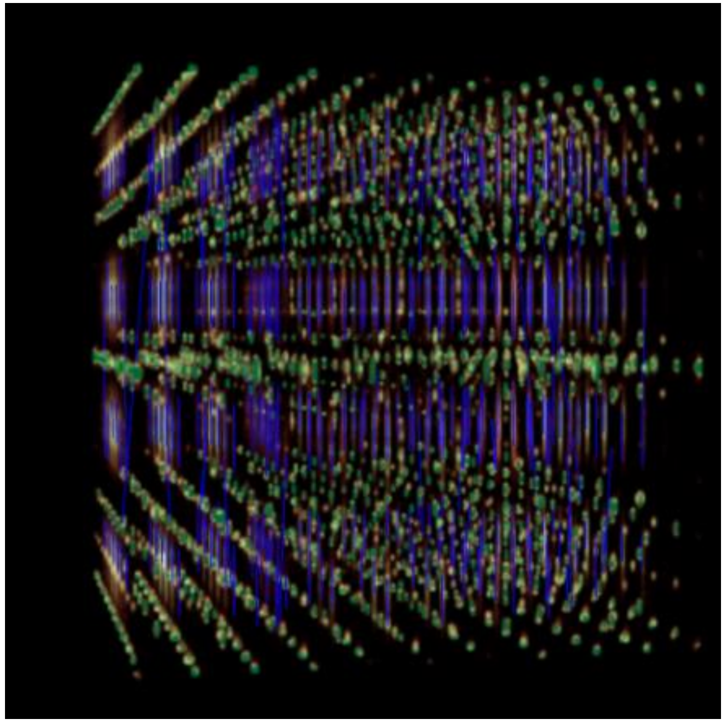
OAK RIDGE National Laboratory | HIGH FLUX ISOTOPE REACTOR | SPALLATION NEUTRON SOURCE

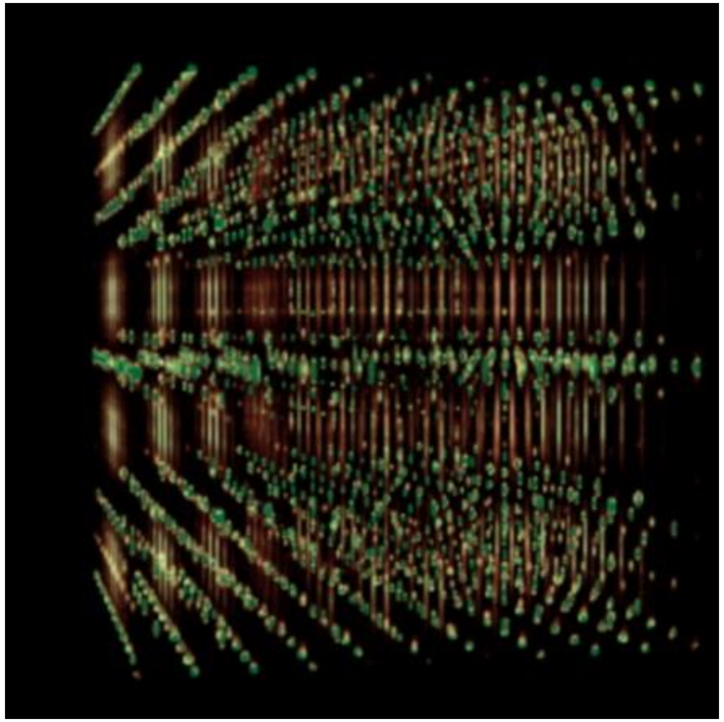
DOE Science Data Pilot Project

- **Diffuse scattering** contains information about **disorder in materials** which is critical to understand function.
- **Novel approach using pattern recognition and machine learning**.
- Aligned with science needs of CORELLI and TOPAZ.

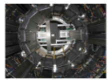


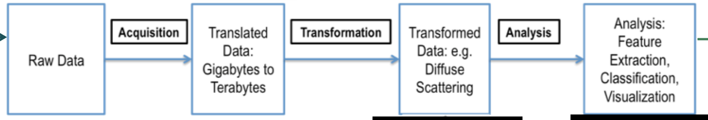
Atlas of Optical Transforms, Harburn, Taylor and Welberry (1975)



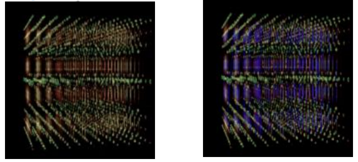


High Level Demonstration Workflow

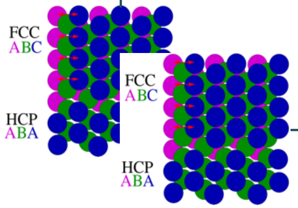




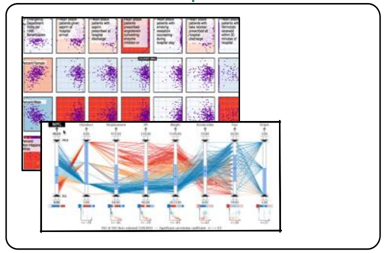
SNS – TOPAZ
Neutron Scattering



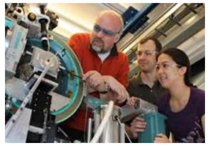
Feature detection "Streaks" in Diffuse scattering data



Classification of "streaks" into candidate stacking faults

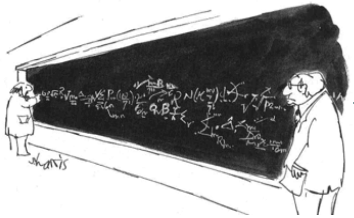


Scientific visualization and visual analytics interface presenting detected features and candidate stacking faults based on classifiers



APS - X-ray Scattering coming soon

Diffuse scattering simulations over parameter space



Knowledge base of experiments and simulations

Data sharing and catalog of results

DISCUS SIMULATION PACKAGE

EDEN Exploratory Data analysis ENvironment

swift

globus online

VisIt

ADIOS

Challenges

- What are the correct labels?
- Sparse data.
- Data management and ‘ML friendly’ metadata.
- Correct normalization for scientific data.

Thank you

NXS Lecture - Thomas Proffen:
“Machine Learning and AI for Scattering Experiments”

Thomas Proffen
tproffen@ornl.gov

http://neutrons.ornl.gov