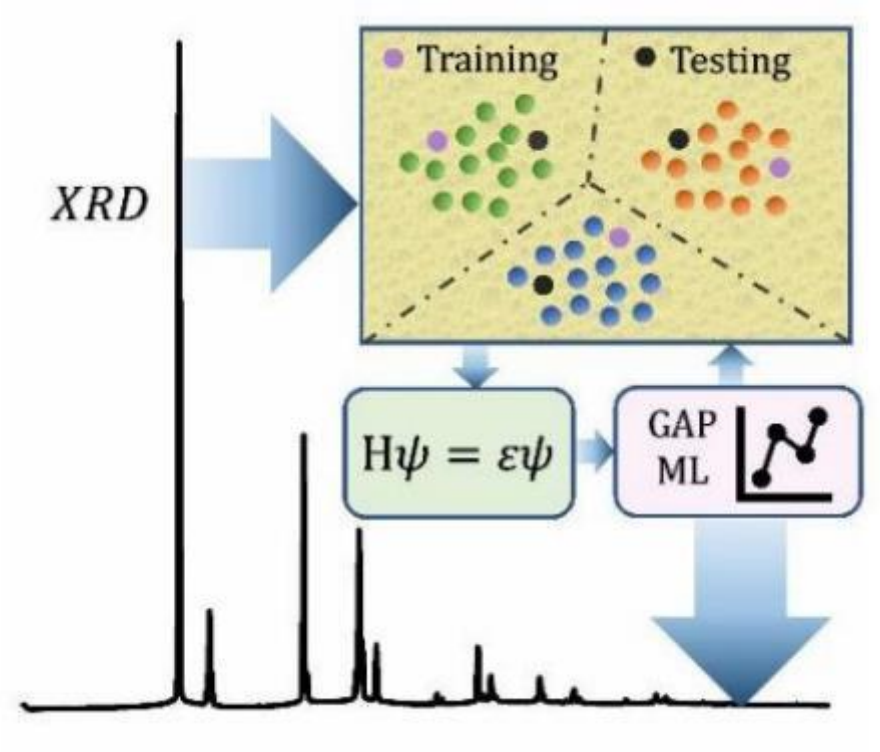
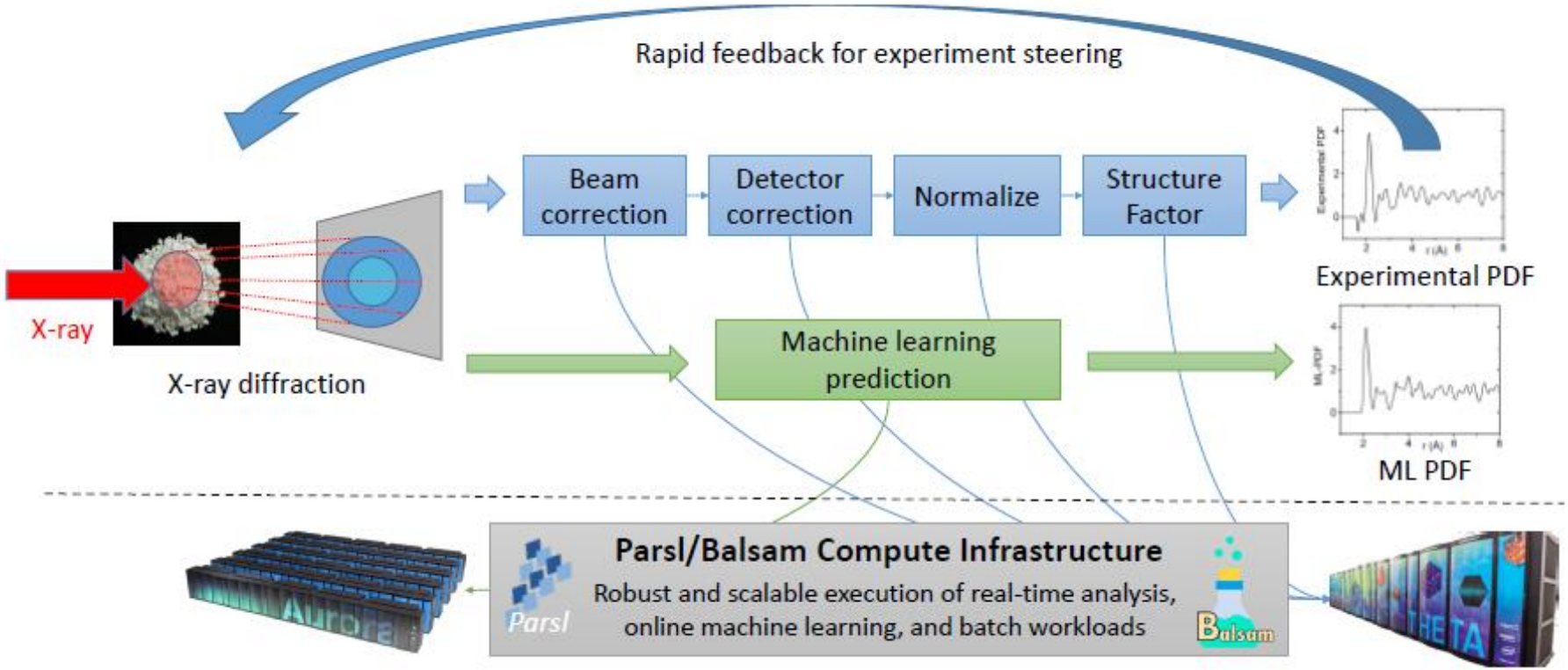
Automated x-ray PDF analysis and ML modeling



Chris Benmore (XSD)

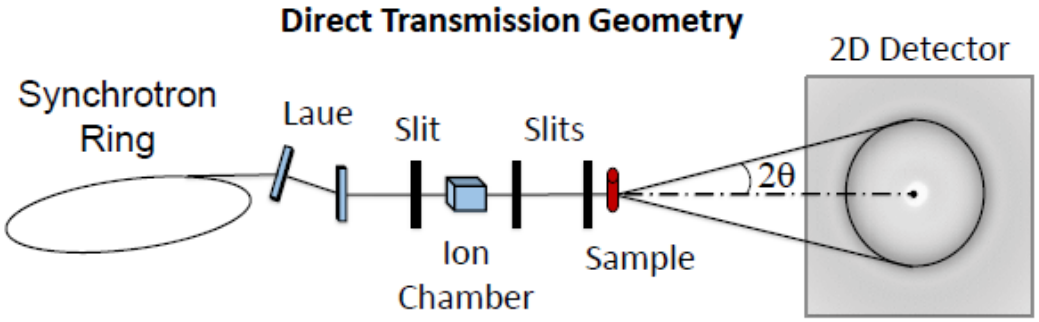
K. Chard (DSL), G. Sivaraman (DSL), R. Vescovi (DSL), R. Chard (DSL), T. Uram (ALCF), M. Salim (ALCF), Á. Vázquez-Mayagoitia (CSD), M. Stan (AMD), R. von Dreele (XSD), L. Gallington (XSD), G. Csányi (U. Cambridge)…

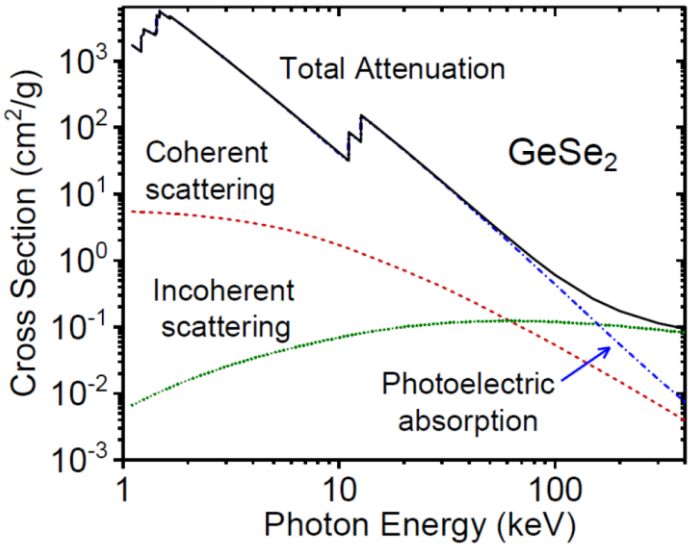
Pair Distribution Function Overview

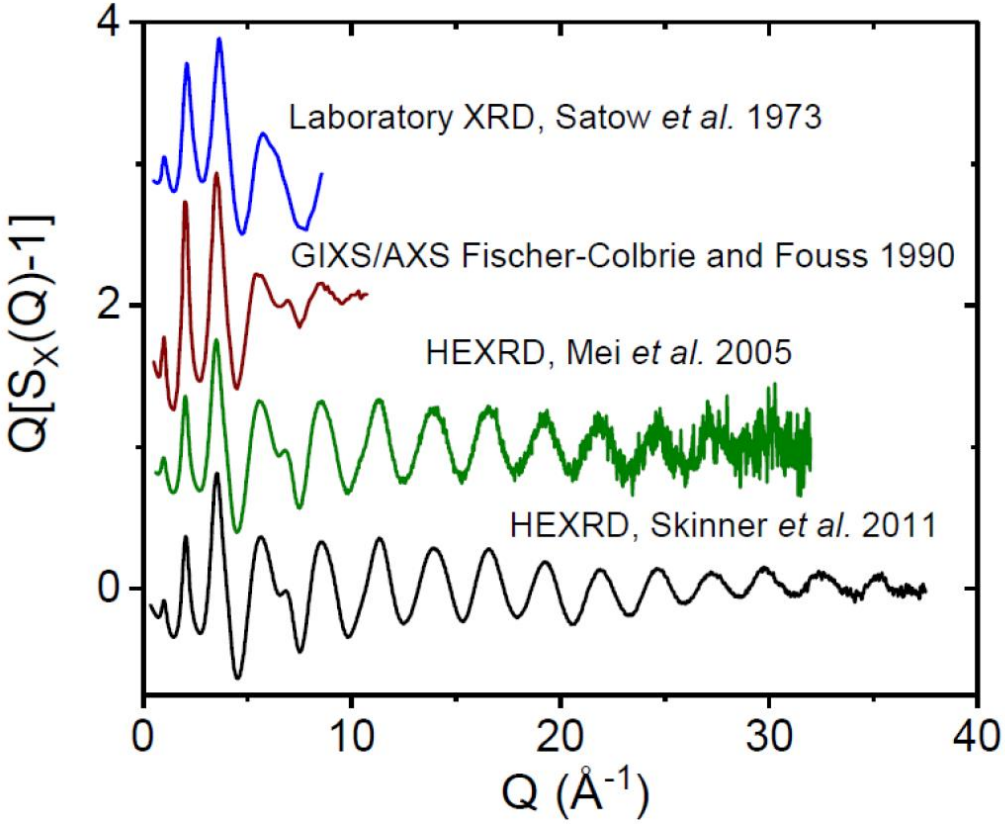


- Automated Data reduction ✓
- Consistency checks for data quality ✗
- ML modeling ✓

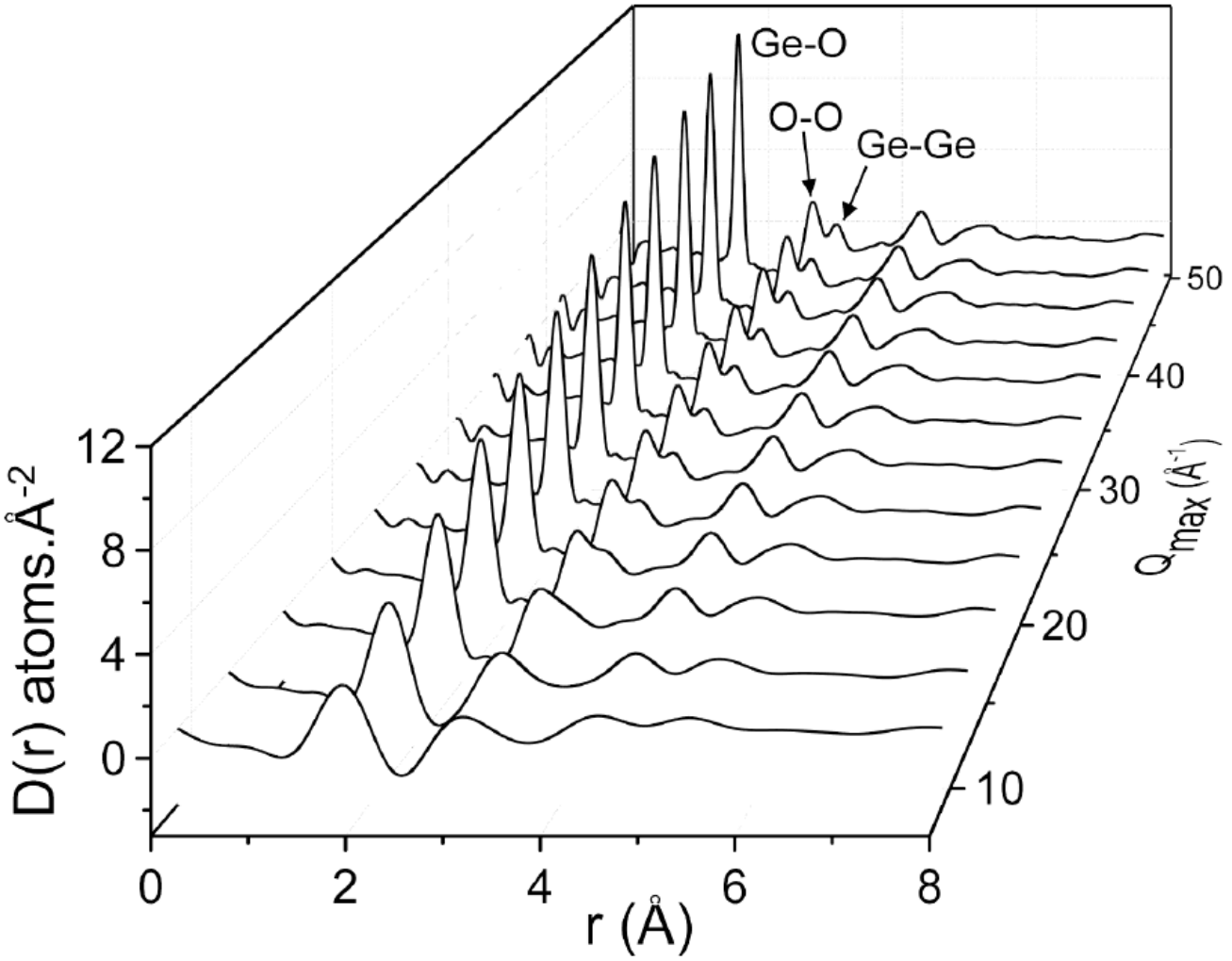
Total Scattering







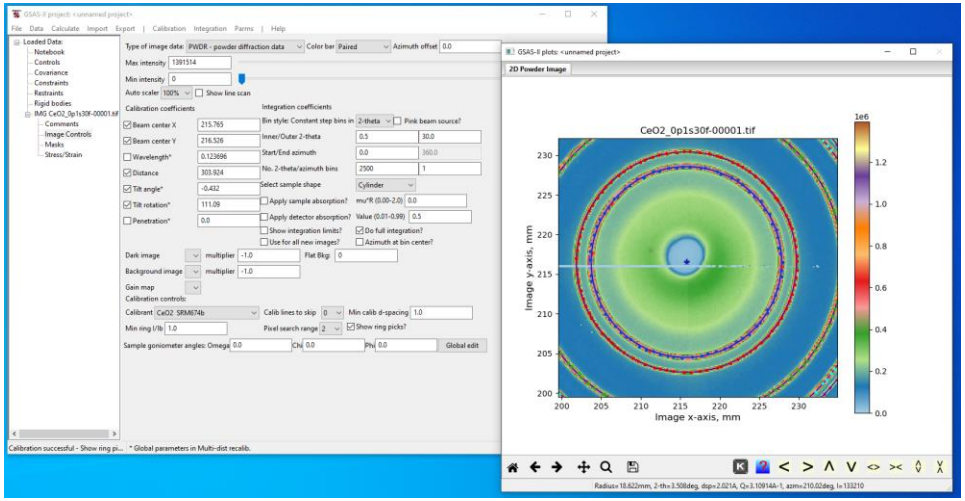
Access to High Momentum Transfers



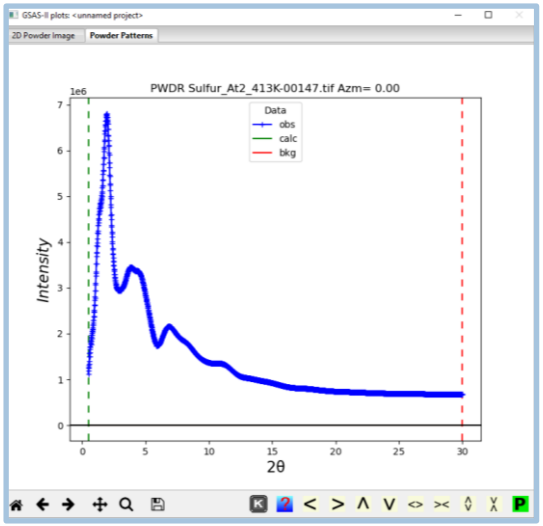
X-ray and neutron diffraction from glasses and liquids.
C.J. Benmore. *Comprehensive Inorganic Chemistry III, Book Chapter, 2021.*

GSAS-2

Automated PDF analysis (LDRD)



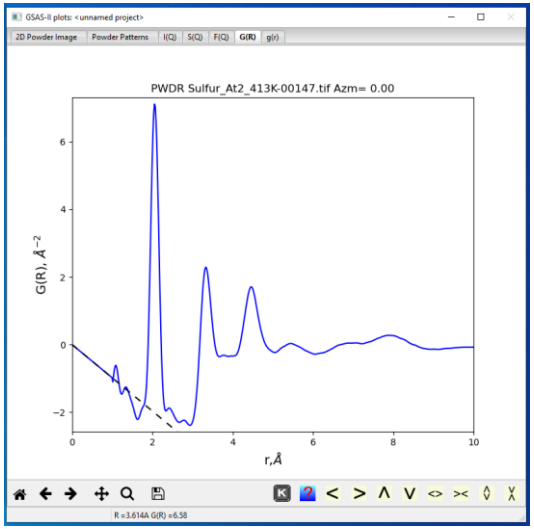
Globus - Parsl/Balsm Infrastructure - GSAS-II



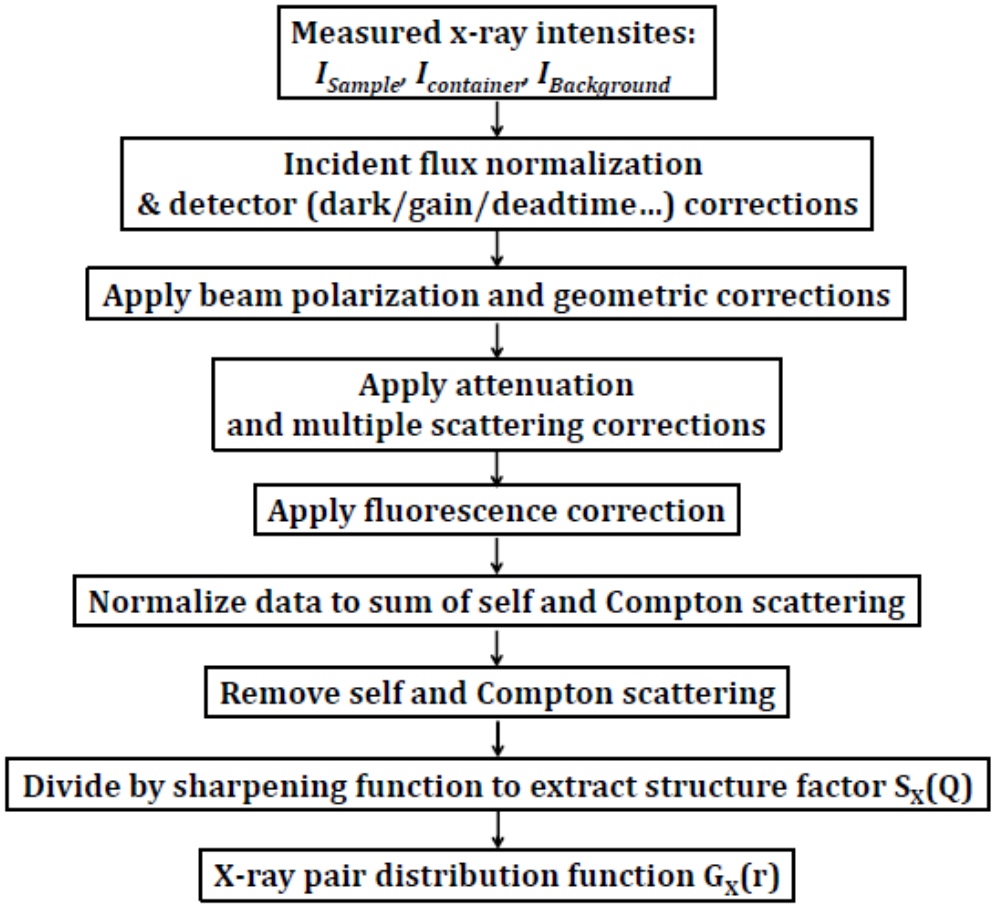
Set up .gpx file at beginning of experiment

Outputs PDF Within a few seconds of measurement

Credit: Darren Govoni
Software engineer



Not all PDF analysis software is the same…



Other commonly used PDF software

PDFgetX2: *a GUI-driven program to obtain the pair distribution function from X-ray powder diffraction data* ✓

GudrunX: *Routines for reducing total scattering data* ✓

PDFgetX3: *a rapid and highly* ***automatable program*** *for processing powder diffraction data into total scattering pair distribution functions* ✗

Most x-ray and neutron PDF courses neglect importance of corrections and experimental set up.

What do you want from your data ?

Phase Identification

Amorphous vs. Crystalline

Peak Positions

Q-calibration
Polarization…

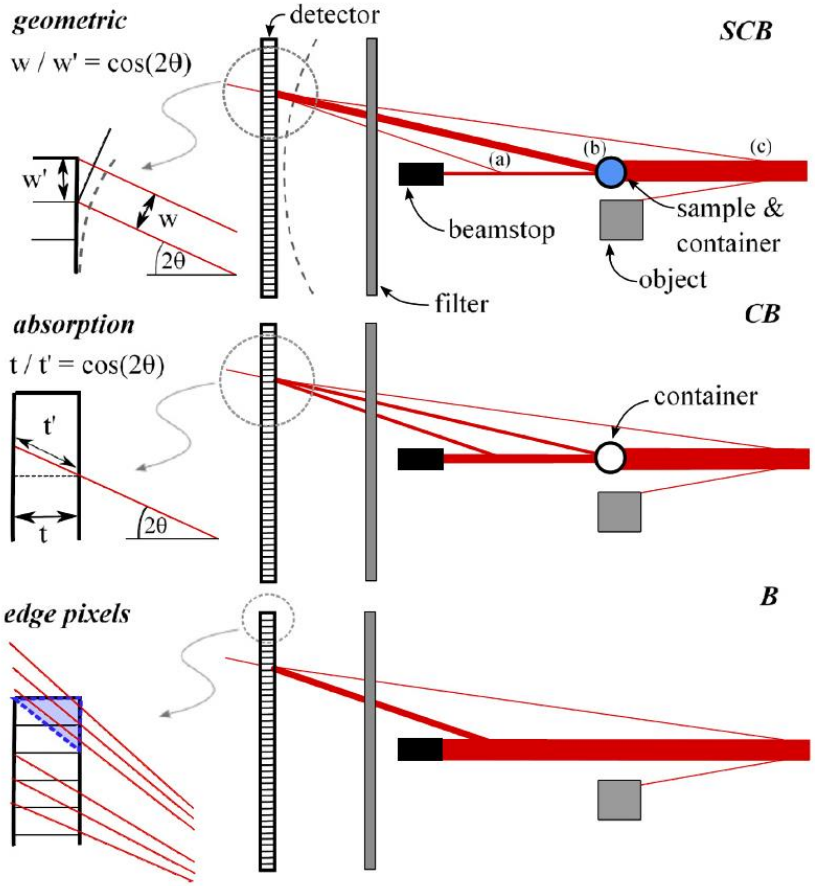
3D structural molecular model

Coordination numbers

Intra- & Inter-molecular structures

Absolute normalization
Quantitative Corrections
Meets all consistency checks...

Detector Corrections…



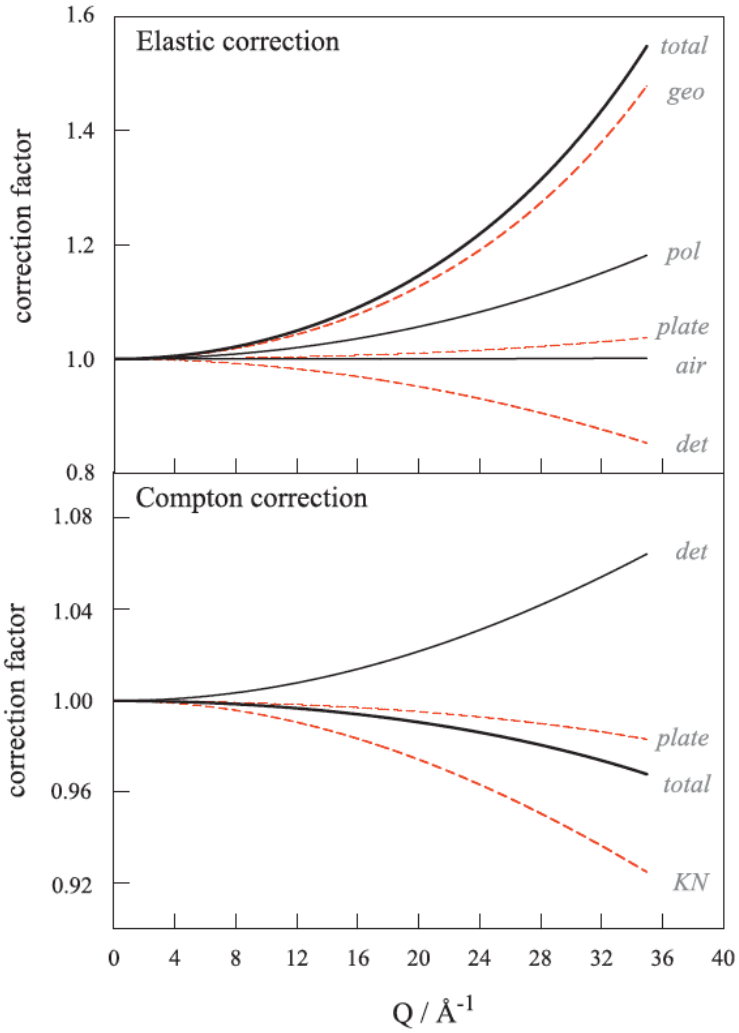


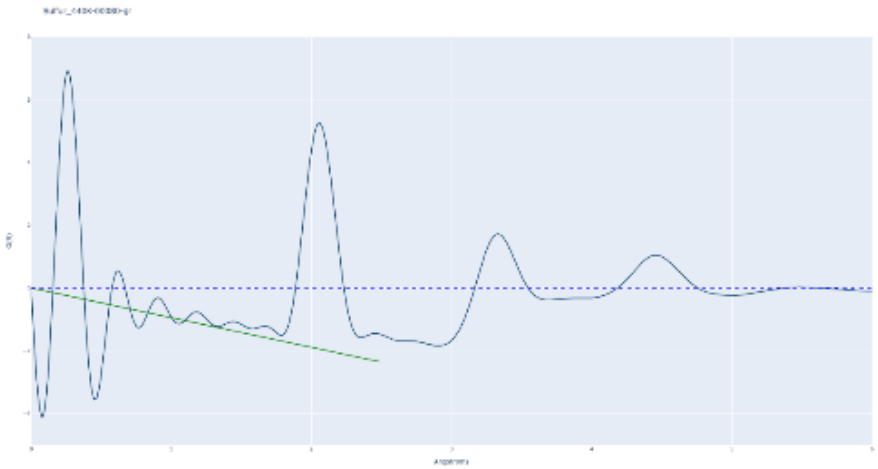
Fig. 6. Top: multiplicative correction factors for a CsI scintillator image plate detector at 115 keV (i.e. the setup at APS beamline 11-ID-C). *Geo*=geometric correction, *pol*=polarization correction, det=detector oblique absorption correction, plate=plate filter correction, air=air absorption correction. Bottom: correction factors to the Compton scattering due to its differing energy (which are additive in $S(Q)$). KN=Klein–Nishina quantum mechanical correction to the Compton cross-section.

Area detector corrections for high quality synchrotron X-ray structure factor measurements.
L.B. Skinner, C.J. Benmore, J.B. Parise.
Nuc. Instr. & Meth. A 662 (2012) 61.

Accuracy counts…

Number of input parameters	Sample dependent corrections @100 keV	$I(Q=25)$ correction relative to $I(Q=0)$		% Error in correction	Consequent error in $I(Q)$ (%)	
		SiO_2	La_2O_3		SiO_2	La_2O_3
1–4	Sample attenuation	0.99	0.93	2	0.02	0.14
1–4	Fluo. sample attenuation	–	0.79	4	–	0.28[a]
1	Fluo. detector absorption	–	1.00	1	–	< 0.001
		100 keV	60 keV		100 keV	60 keV
Sample independent corrections						
1	Polarization	1.13	1.38	0.5	0.06	0.14
0	Geometric	1.23	2.04	0.25	0.05	0.13
1	Detector absorption	0.90	0.89	1	0.1	0.1
1	Filter attenuation[b]	1.03	1.11	2	0.06	0.20
Energy dependent Compton corrections[a]						
1	Detector energy dep.	1.05	1.07	1	0.05	0.07
1	Filter en. dep.	0.99	0.99	2	0.02	0.02
0	Klein–Nishina	0.95	0.93	< 0.1	< 0.005	< 0.007
				Total	0.34	0.66

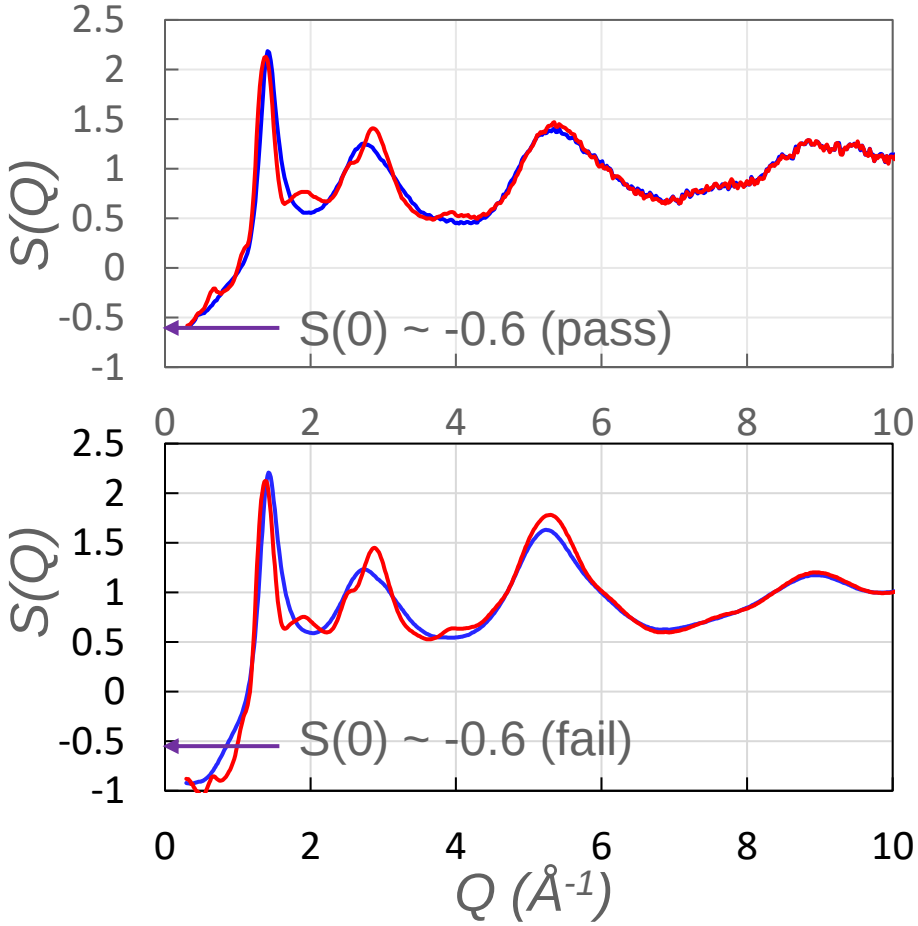
Consistency checks…

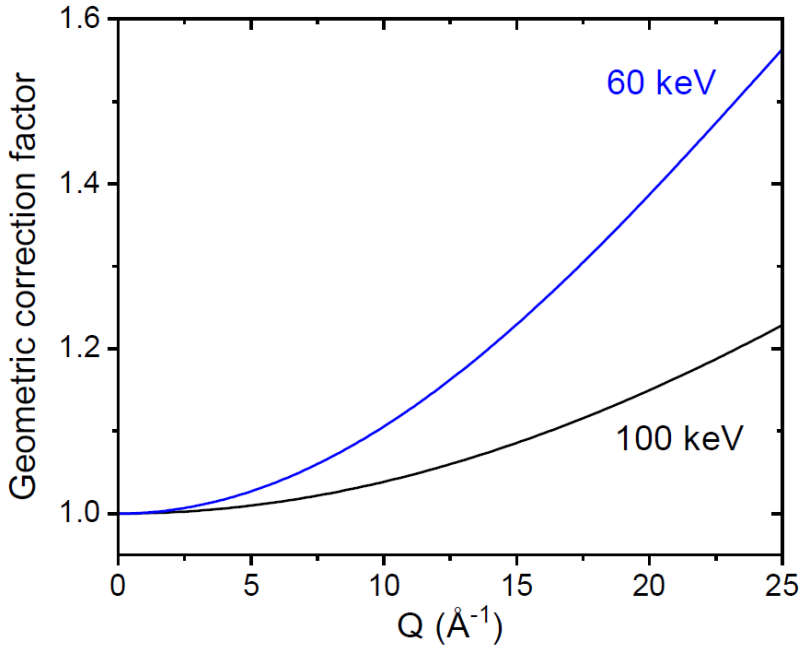


- **Automated GSAS-II analysis**
- **Check low-r limit** agrees with bulk density line.
- **Fourier back-transform** can assess level of systematic error – only implemented in GudrunX

Other consistency checks…

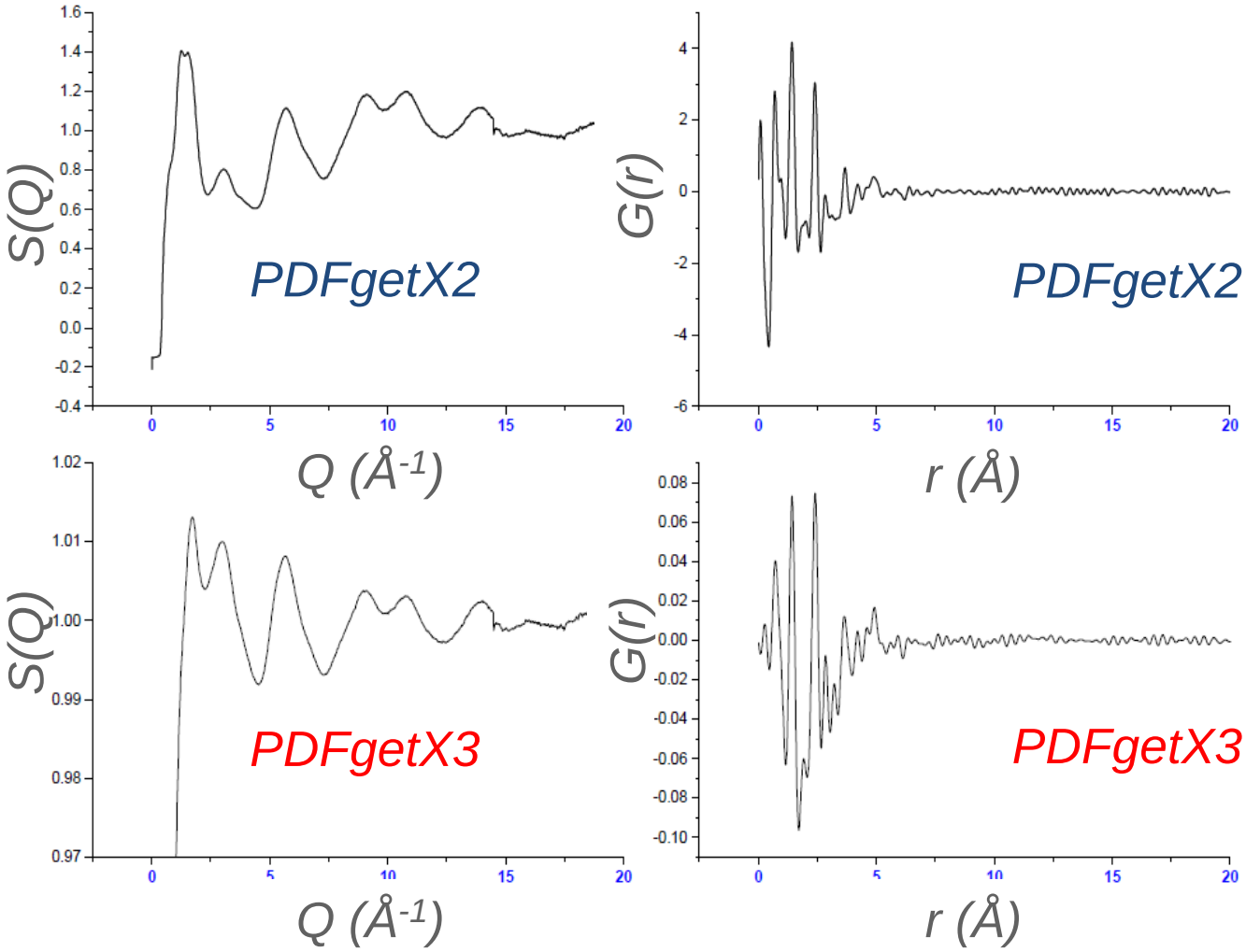
Example 1. User data on same low-Z sample at different temperatures measured on different diffractometers at different energies





Effects of approximations used in automation…

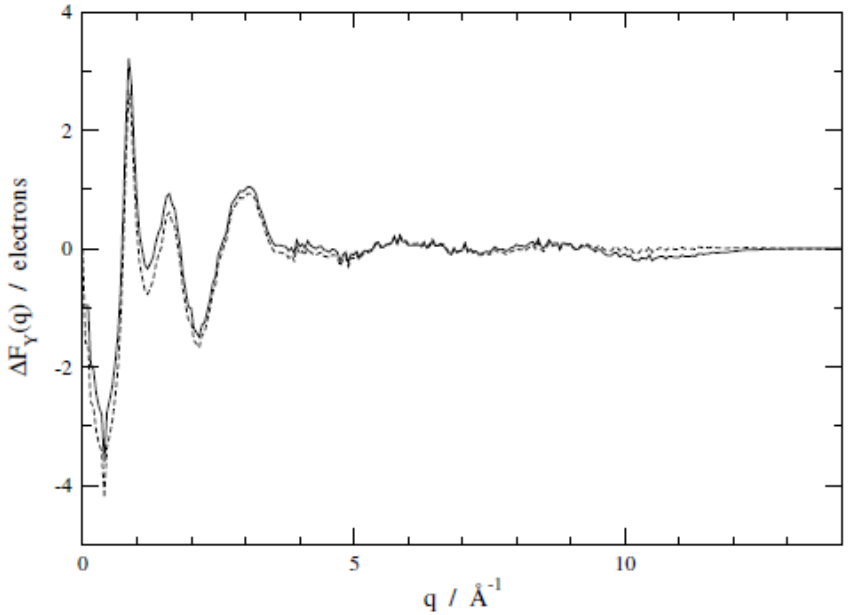
Example 2. User data on same low-Z sample analyzed using different software

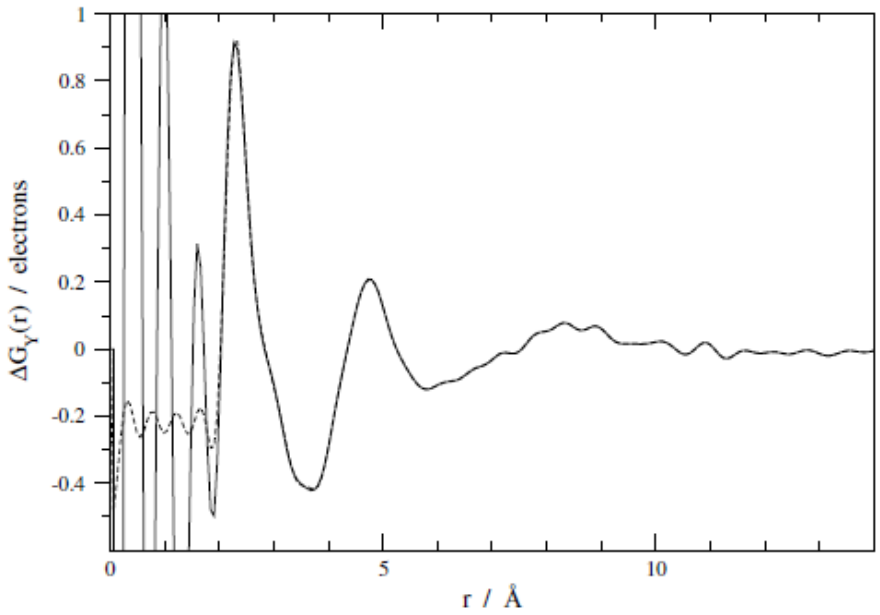


The Fourier back-transform

"A much better indication of the overall accuracy, including systematic errors, is given by the Structure in the correlation function at low-r below the first true peak, ***provided the data have not been massaged before publication to conceal their true quality***."

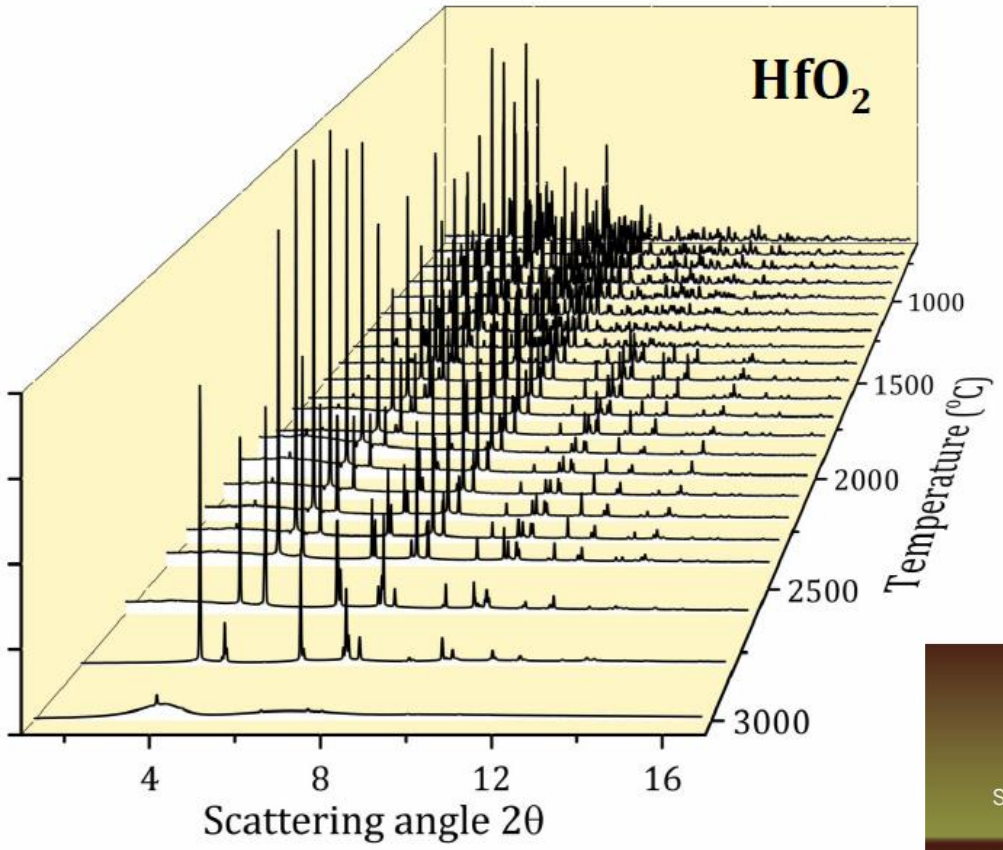
What have we learned from 60 years of diffraction studies ?
A.C. Wright, J. Non-Cryst. Solids 179 (1994) 84.

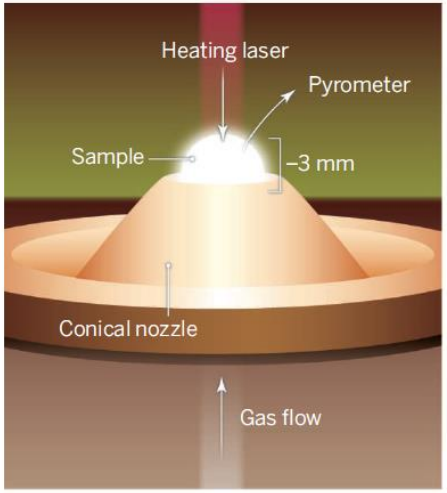




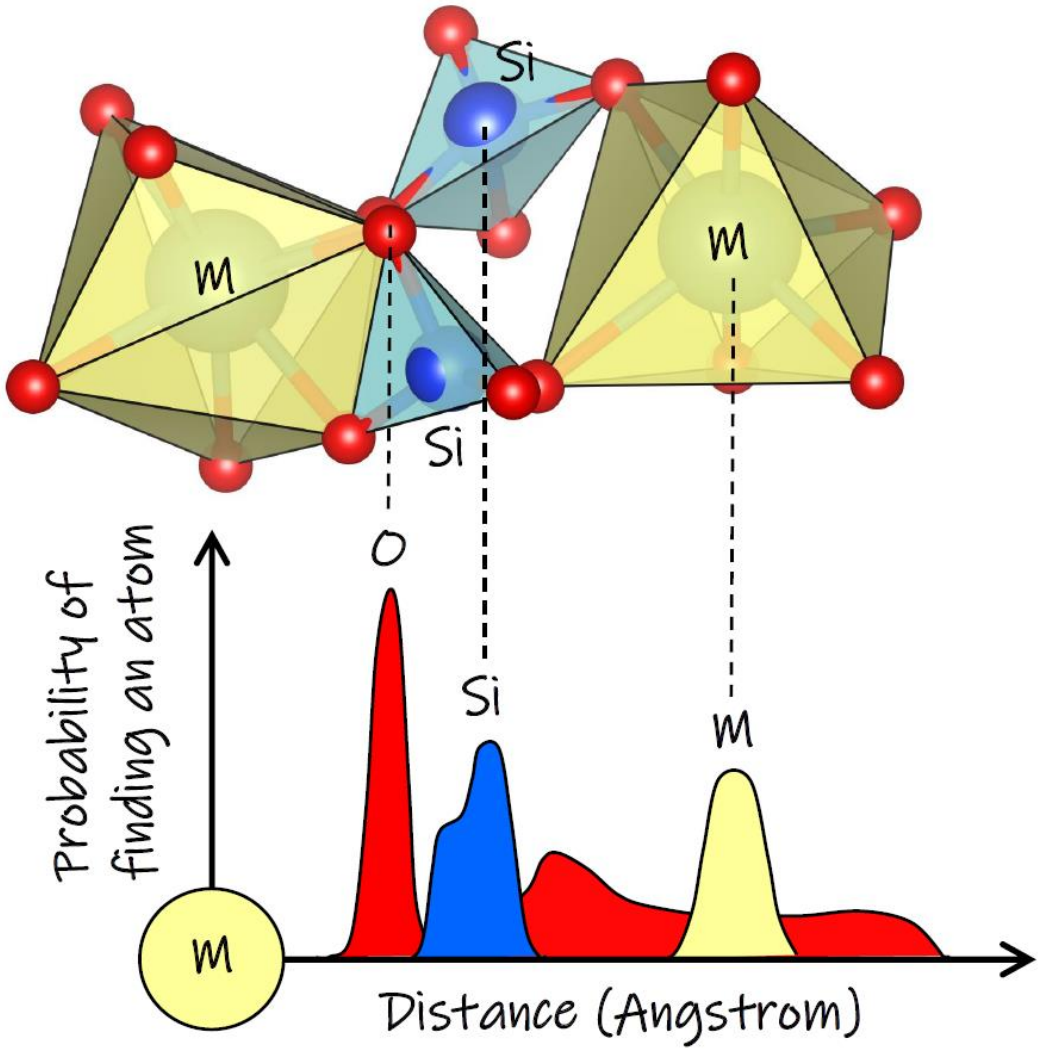
Example from *Neutron and X-ray diffraction studies of liquids and glasses*
H.E. Fischer, A.C. Barnes, P.S. Salmon. Rep. Prog. Phys. 69 (2006) 233.

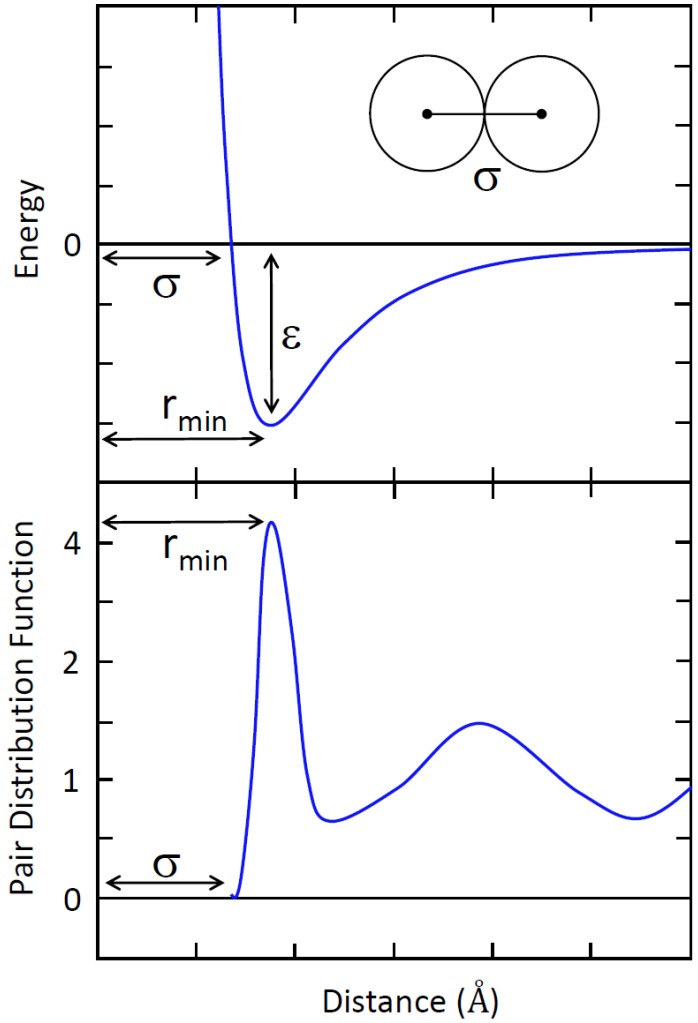
Aerodynamic levitation & Laser heating





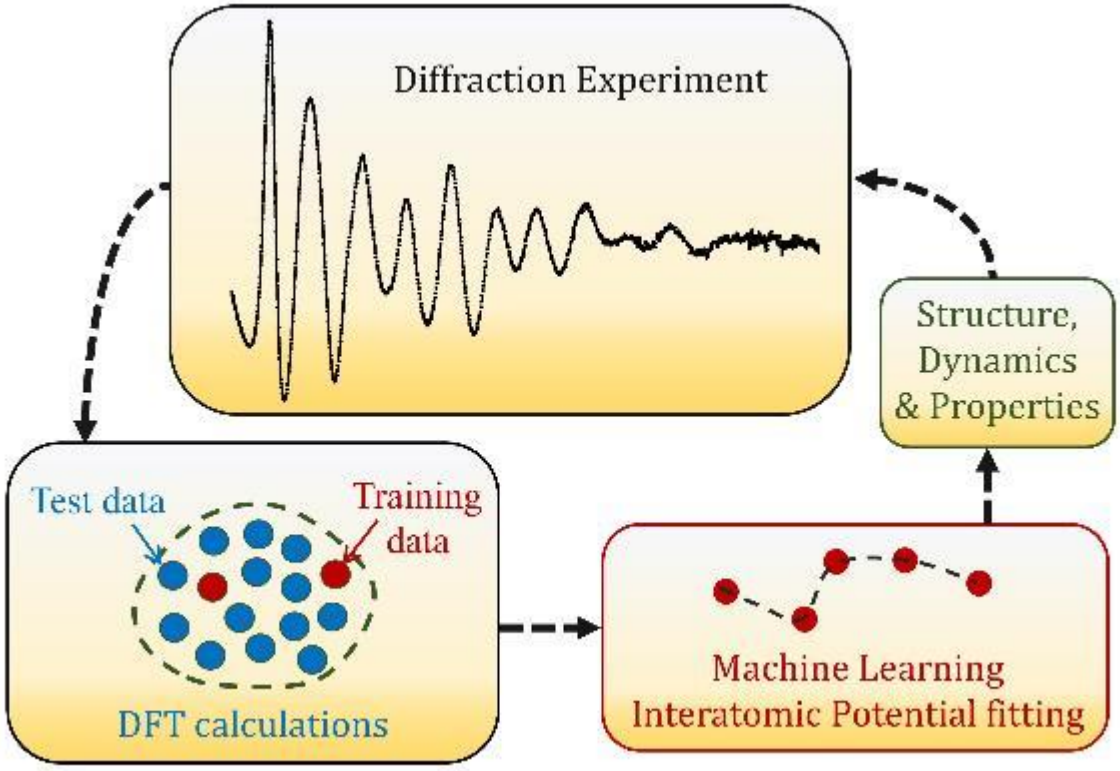
Local structure





Probing Melts, Glasses and Amorphous Materials: Neutron Scattering in Earth Sciences. C.J. Benmore & M.C. Wilding. *Elements, 2021.*

Experimentally Driven Automated Machine-Learned Interatomic Potentials



Underlying principle:

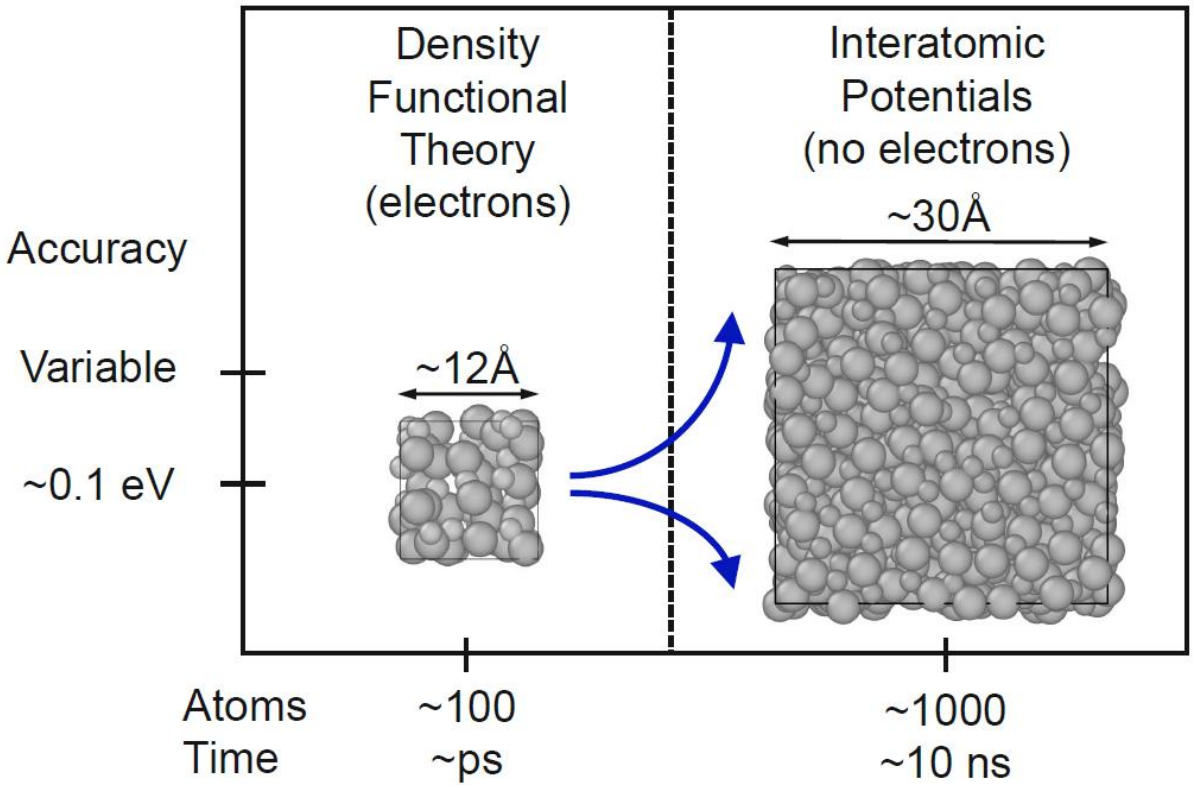
If ***thousands of AIMD simulations*** are performed, some will agree with the diffraction data – how best to get there is a computer science problem !

Gaussian Approximation Potentials I

A. Bartok & G. Csanyi, *Int'l J. Quantum Chemistry* 2015

$$E = \sum_{\alpha} \sum_{i \in \alpha} \varepsilon_i^{\alpha} + \text{long range contributions}$$

local energy functionals



Atomistic Modeling & the Potential Energy Surface

- Pair and N-body interactions
- GAP Parameters fitted from DFT
- Run larger box classical MD simulations
- Extract structure, density, diffusion constants, viscosity, conductivity…

Gaussian Approximation Potentials II

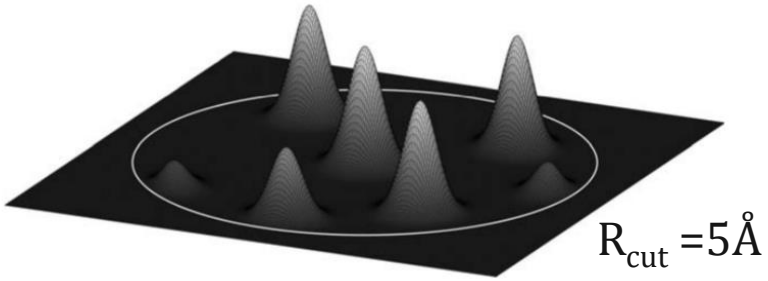
Machine Learning method

- Direct functional between the atomic configuration and energy
- Only uses reference electronic structure
- Initial multi-year effort to develop a general model for a single material !

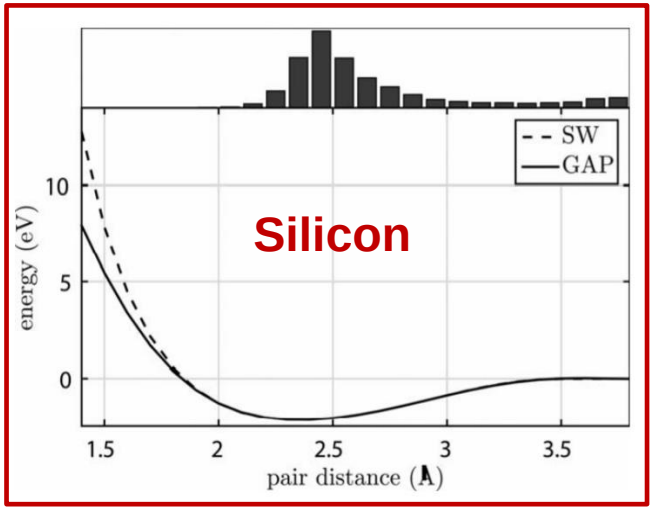
Challenges

- Sampling the correct training data to probe property of interest
- Finding the right hyperparameters for the chosen method of fitting for required model accuracy i.e. energy meV/atom.
- Local minima for metastable materials

Smooth Overlap of Atomic Positions (SOAP)



Deringer *et al. Nat. Comm.* volume **11** (2020) 5461



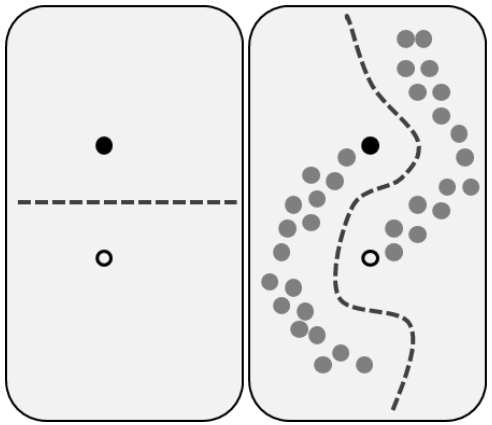
Bartók *et. al.*, *Phys. Rev. X*, **8** (2018) 041048

Active Learning: Testing and training

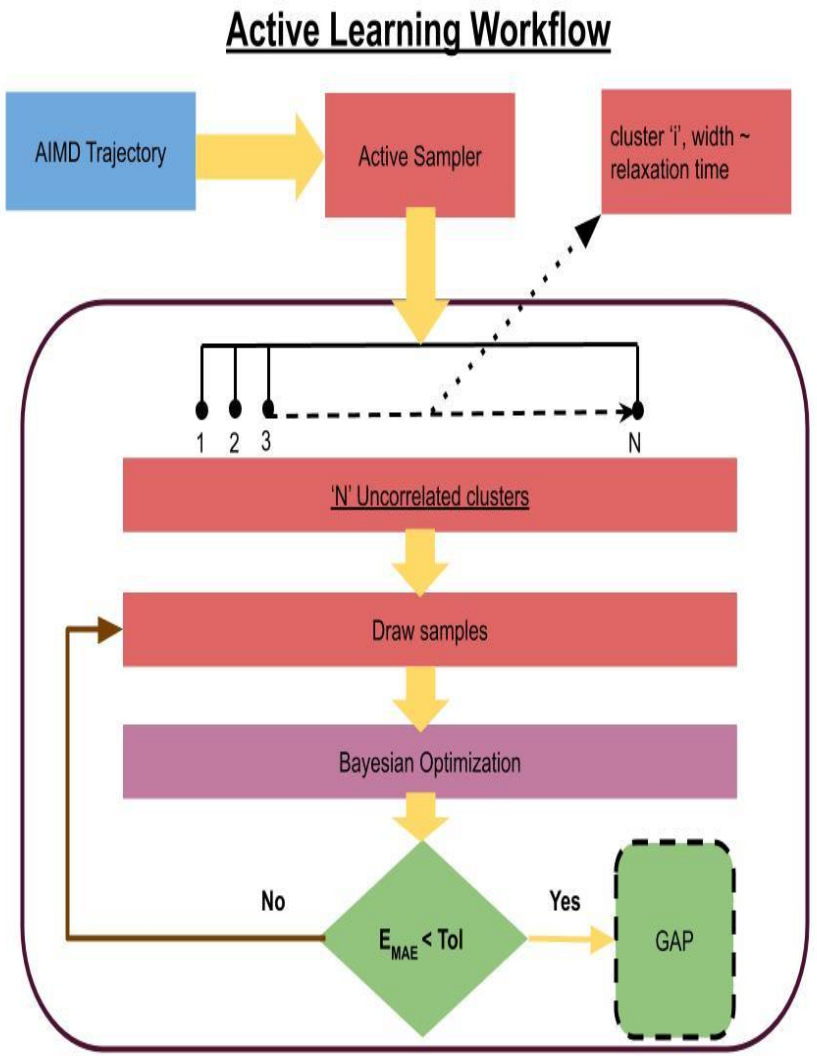
- Large pool of unlabeled data

- Active sampling exploits the underlying patterns embedded in the data to guide supervised learning

- Supervised ML model arrives at an accuracy with **minimum** training datasets



S. Dasgupta, *Theoretical Computer Science* **412**, 1767 (2011)



Hand picked vs Active learning

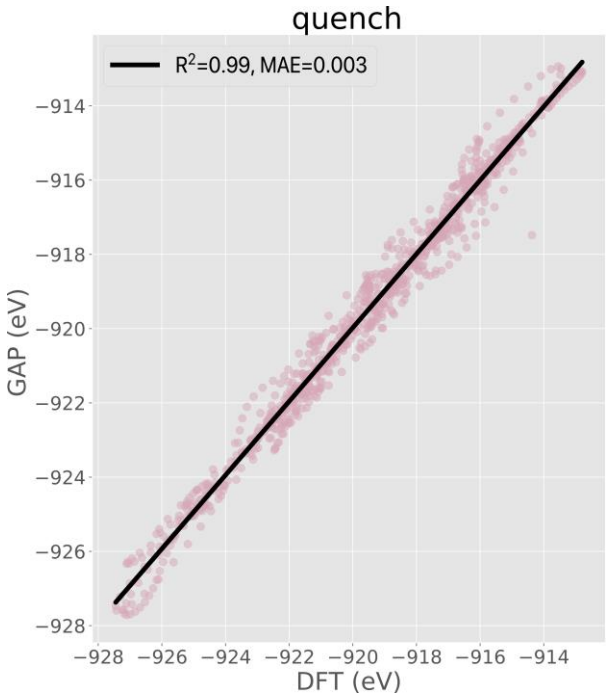
Glassy HfO_2 calculated 33,000 atomic configurations
AIMD (NVT) VASP

Melt- quench protocol

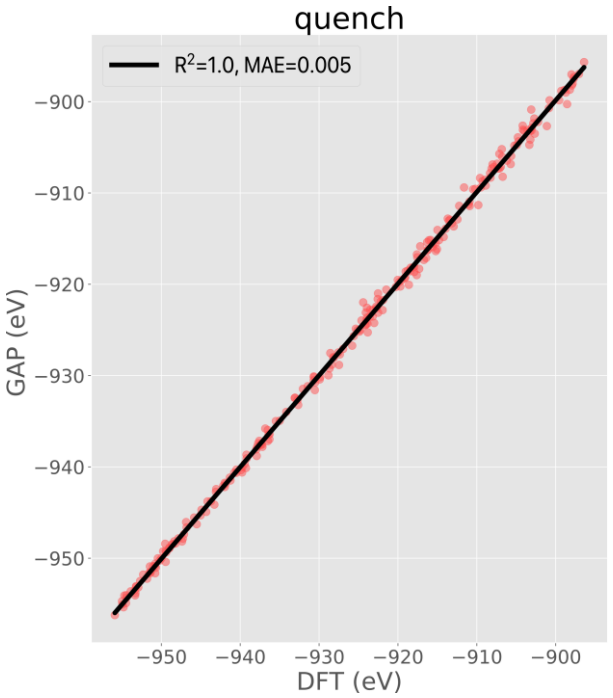
Random particles in cubic box melted at 3600K

Quenching to 300K at a rate of 100K/ps

Glassy dataset sampled at 300K



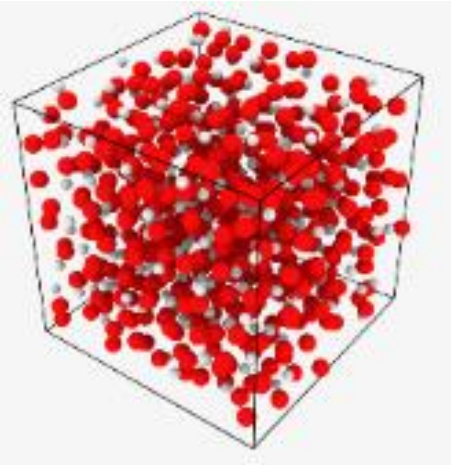
950 training configurations
>30 iterations (sleepless intern)
= 2.90 % Sample size

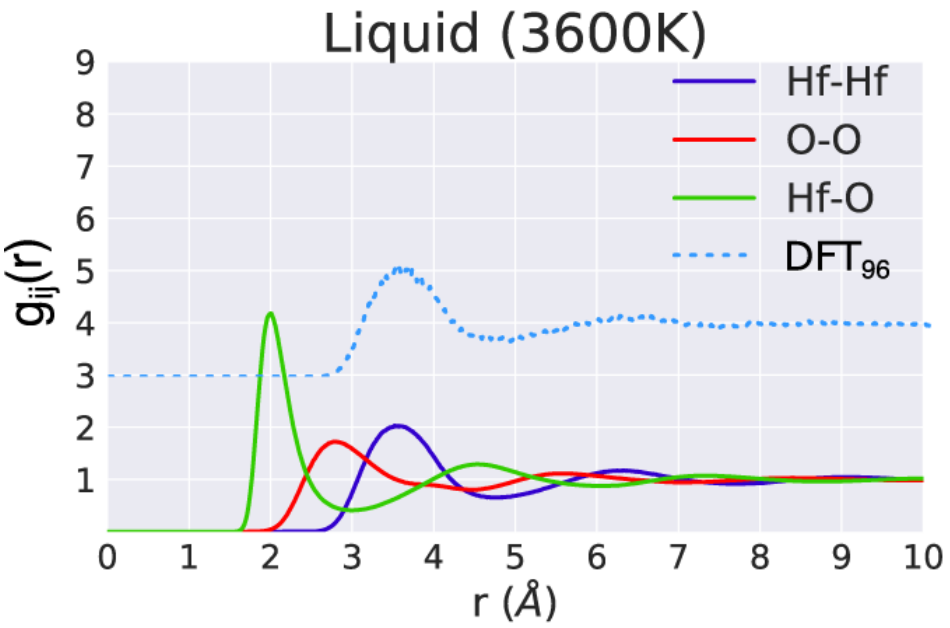


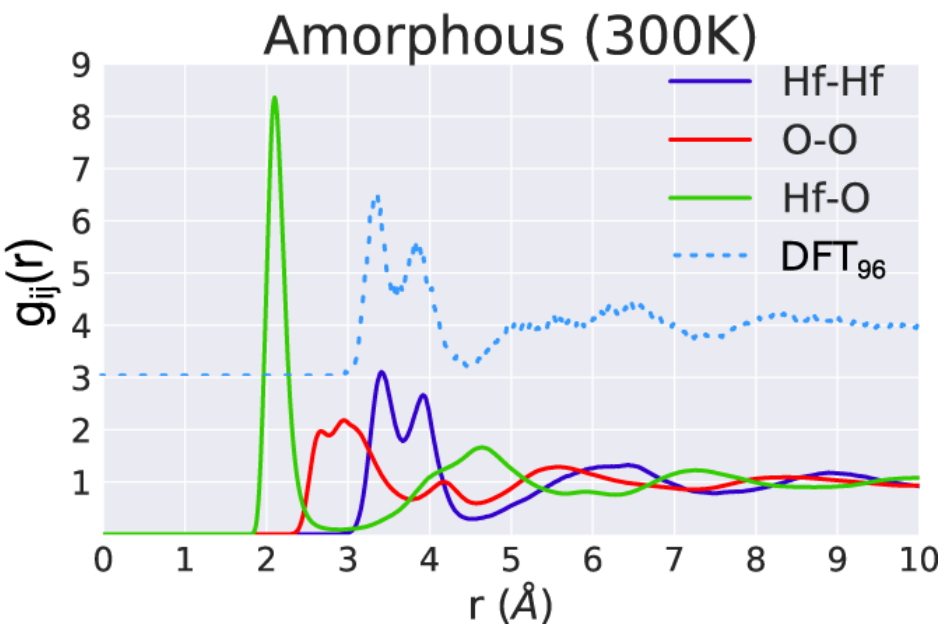
260 training configurations
11 iterations
0.78. % Sample size

From DFT to classical MD for HfO_2

- Classical MD x64 system size of DFT
  - x100 longer simulations with ML
    - Quench rate of 1.0 K/ps

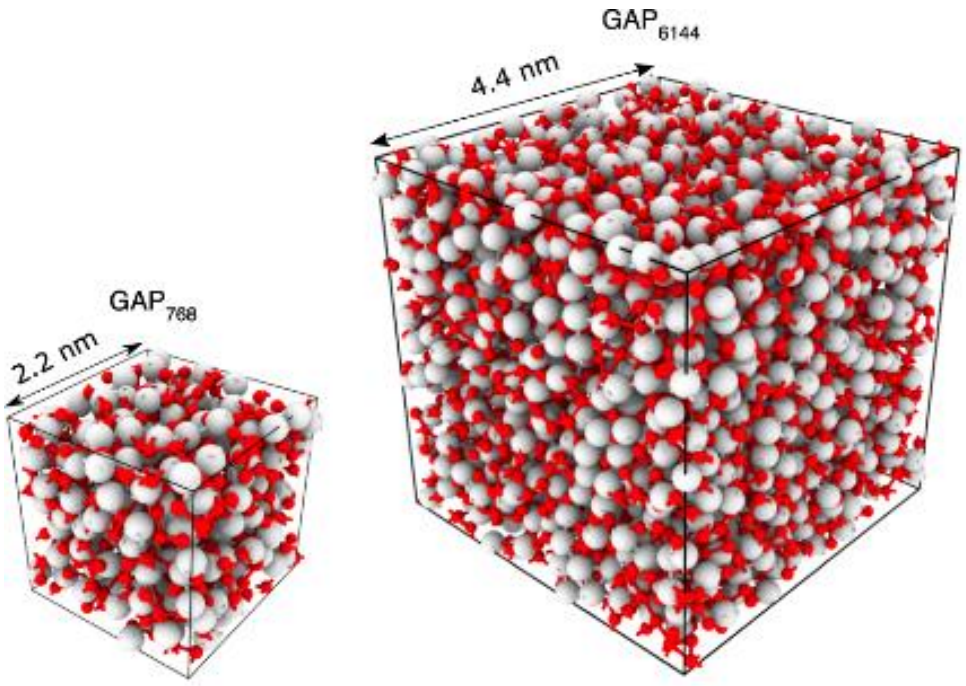


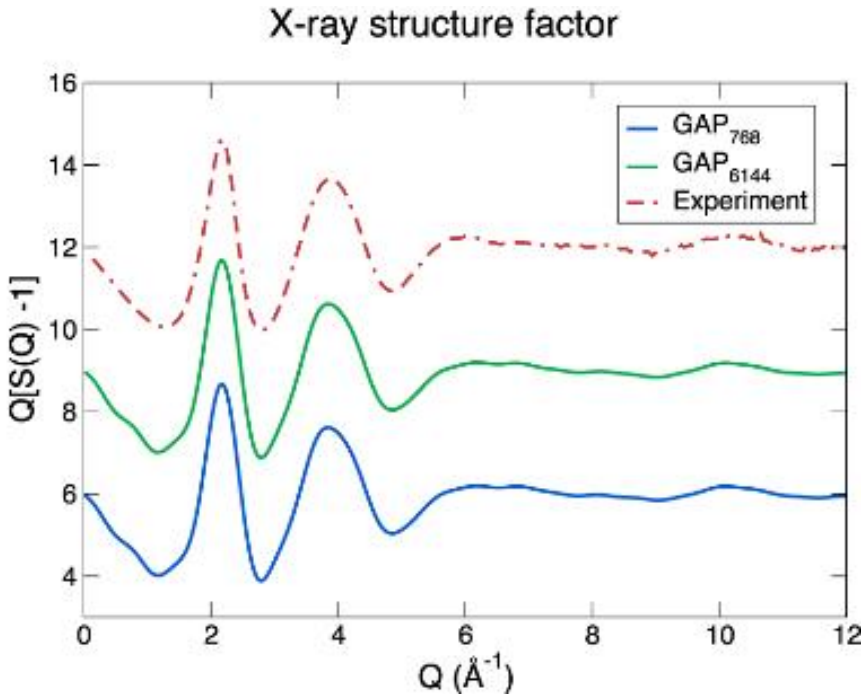




Validation with experiments

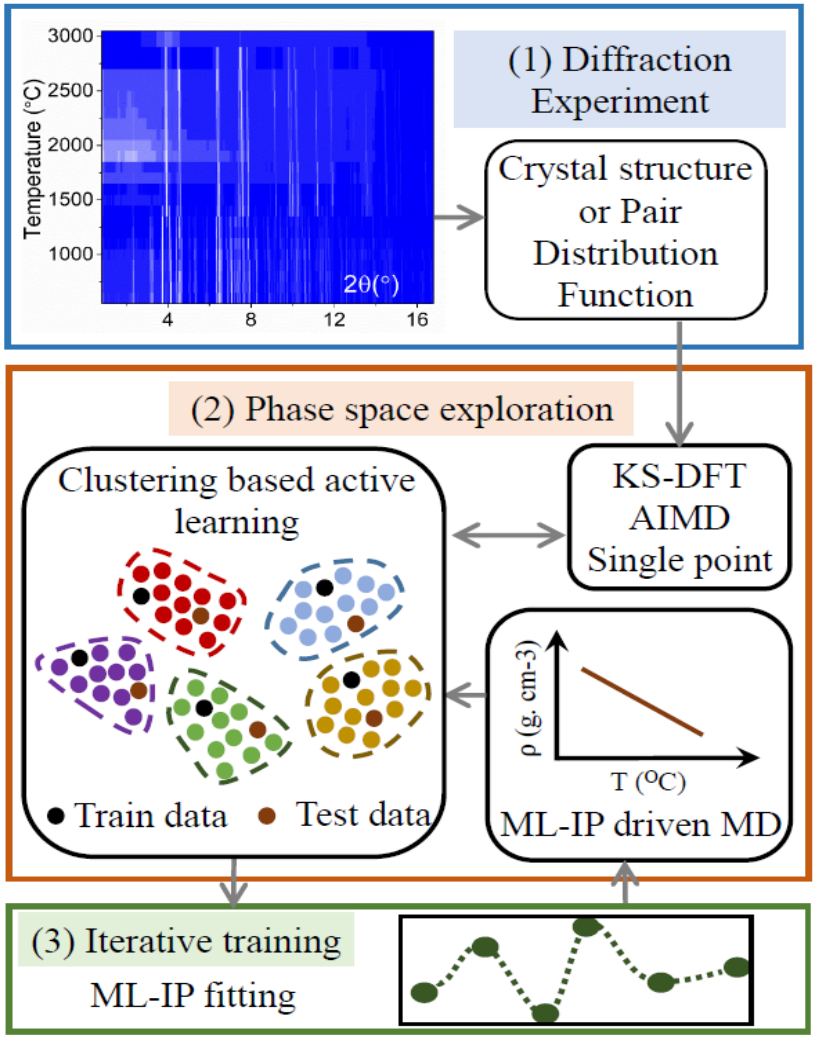
DFT (VASP) to MD (LAMPS, NVT) : Liquid HfO_2 at 2900^oC

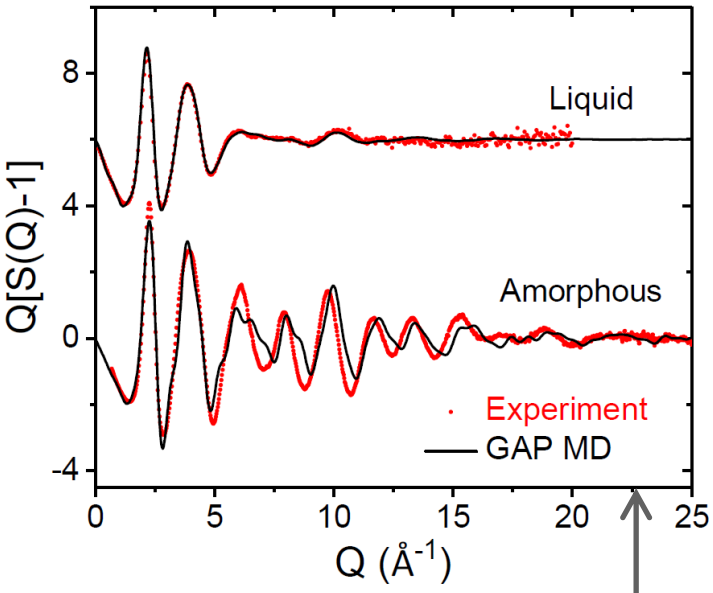




Active Learning Driven Machine Learning Inter-Atomic Potentials Generation.
G. Sivaraman et. al. *npj Computational Materials, 2020.*

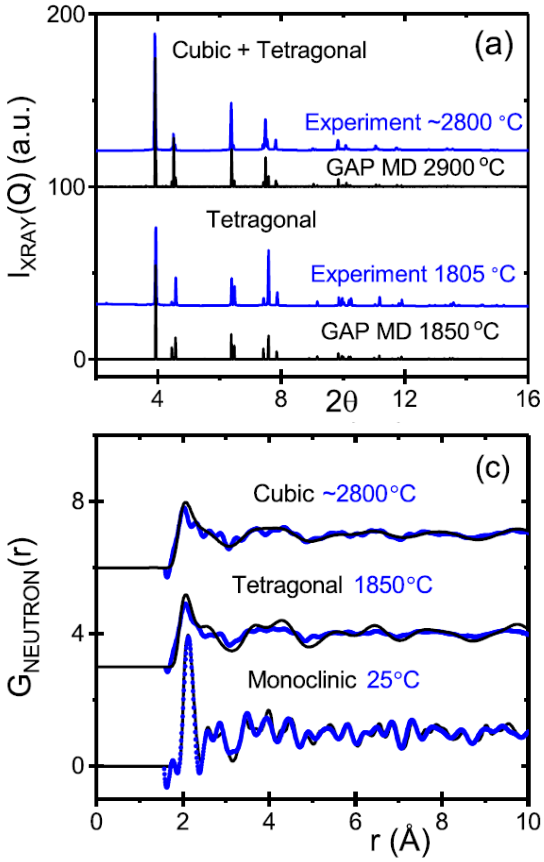
Experiment Driven Automation



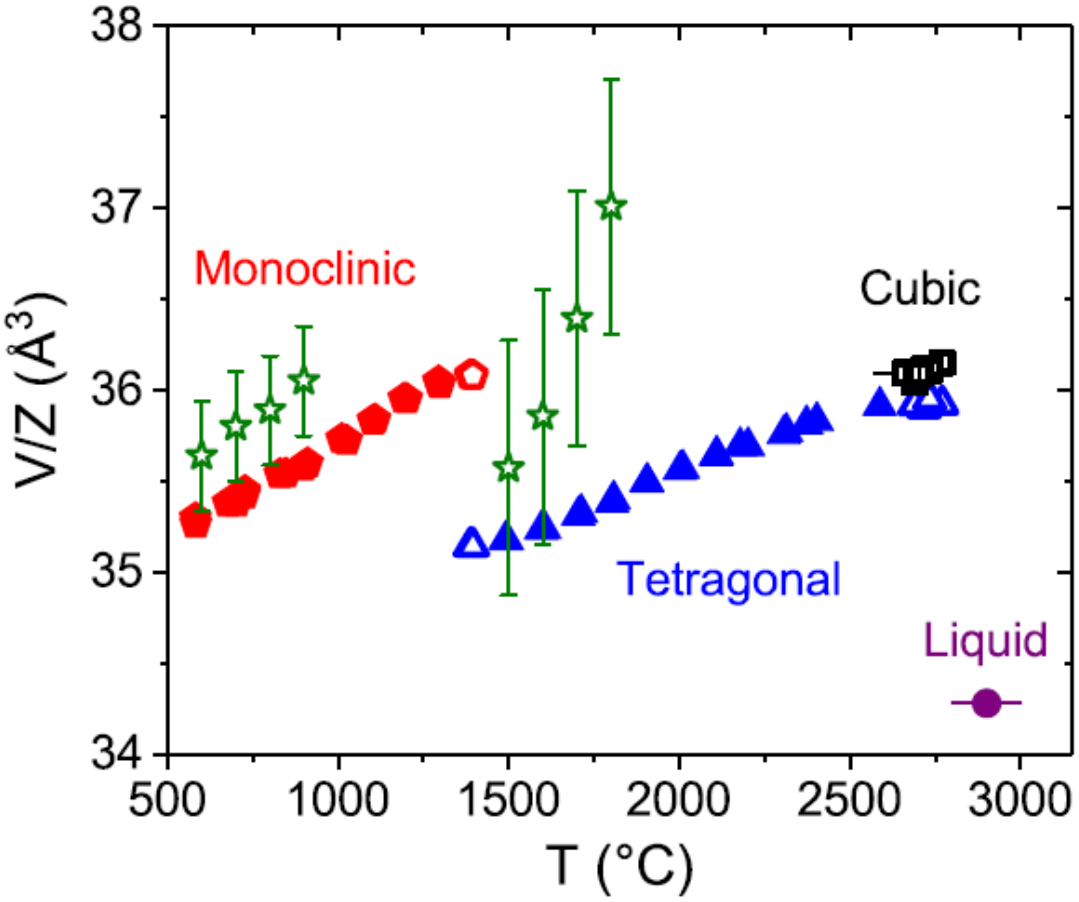


- X-ray data drives construction of *effective* classical pair potential from DFT
- The experimentally synthesized structures are **not necessarily the lowest energy structure**
- AL trains atomic and electronic structural models in near *ab initio* accuracy with PDF
- Highlights need for **accurate PDF data**

“Global” Machine Learnt Potentials



Simulated melting point is within 1% of experiment



PHYSICAL REVIEW LETTERS **126,** 156002 (2021)

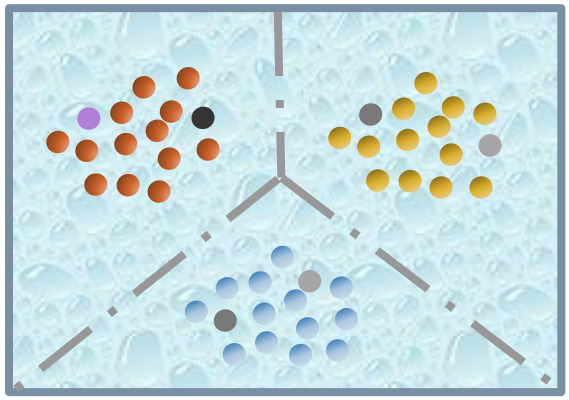
Editors' Suggestion

Experimentally Driven Automated Machine-Learned Interatomic Potential for a Refractory Oxide

G. Sivaraman *et. al.*

New machine Learning Scheme for Disordered Materials

- **Experimental PDF data** drives active learning algorithm
- Tests **AIMD simulations** using a Gaussian Approximation Potential
- Classical GAP MD simulations reproduce all the experimental phases with **near *ab initio* precision**.



- Larger box size allows **quench rates of 1.0 K/ps** not accessible via AIMD
- The method **significantly reduces** model development time and human effort
- Potential limitations in future for large amount of kernels.

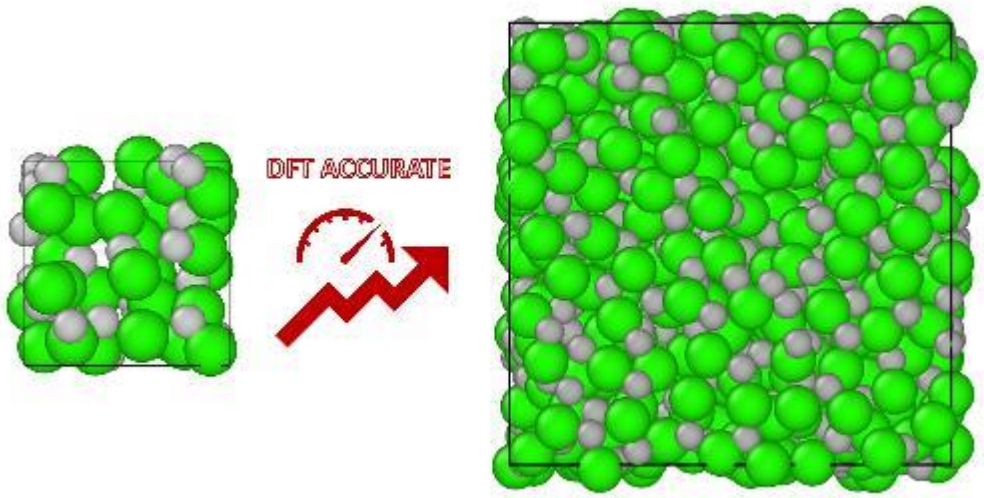
Alternative M approaches

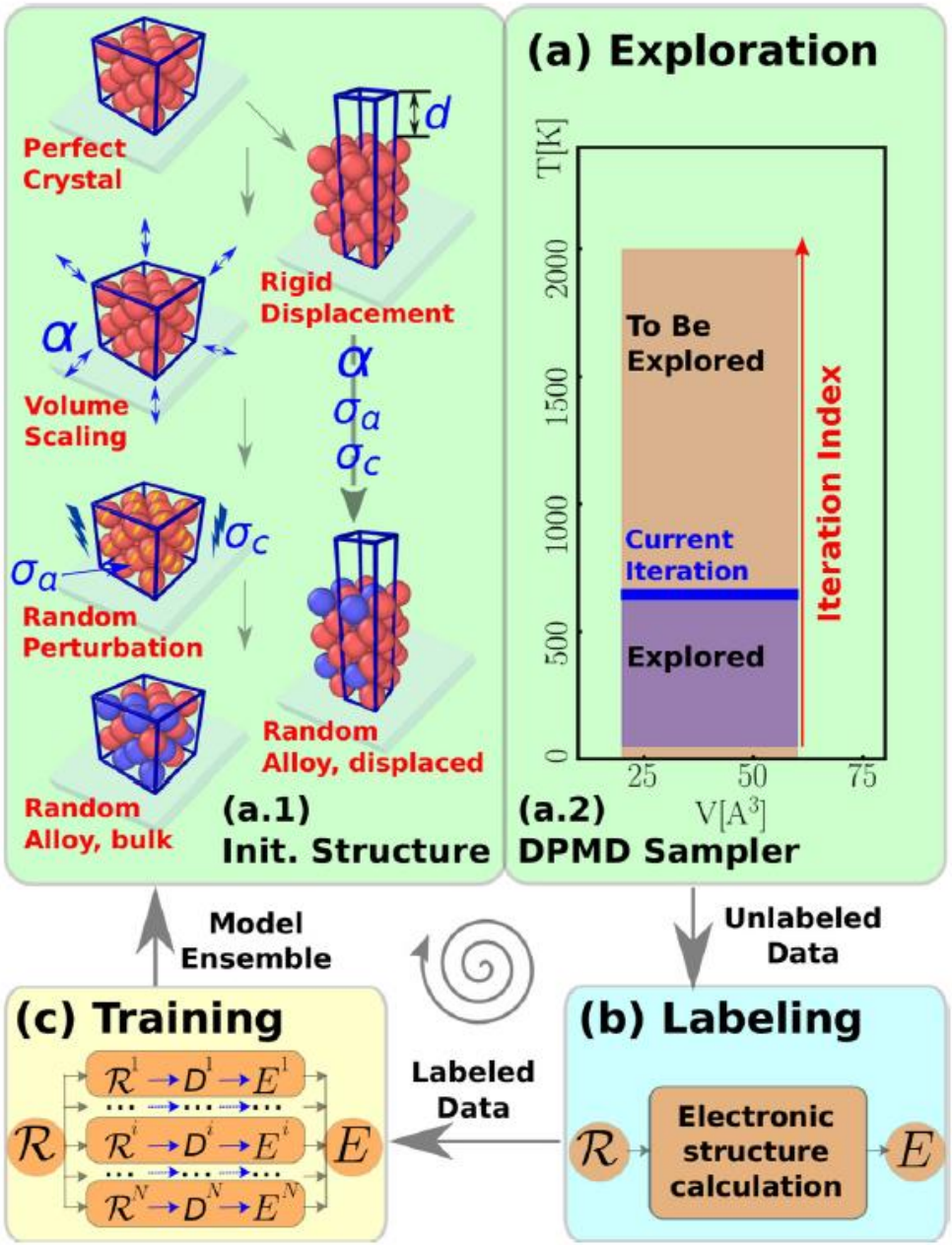
- Other ML potentials for molten salt simulations e.g. Li *et al. Cell Rep. Phys. Sci.,* (2021)
- GAP approach superior to neural network based models when dealing with small quantities of training data that are computationally expensive
  - Li *et al.* required **112,000** training samples.

GAP model used **~600** in Sivaraman *et al. J. Phys. Chem. Lett.* 2021.

Future

- Experimental consistency data analysis checks are **essential**
  - Metadynamics, complex systems, oxidation states….





Deep Potential Generator Using Neutral Networks

Construction of accurate and transferable ML models of the Potential Energy Surface

e.g. liquid Mg-Al alloy

Zhang *et al.* Phys. Rev. Mat. 3 (2019) 0233804

THE JOURNAL OF PHYSICAL CHEMISTRY C

J. Phys. Chem. C 2020, 124, 25760−25768

pubs.acs.org/JPCC Article

DFT Accurate Interatomic Potential for Molten NaCl from Machine Learning

Samuel Tovey, Anand Narayanan Krishnamoorthy, Ganesh Sivaraman, Jicheng Guo, Chris Benmore, Andreas Heuer, and Christian Holm*