

Fitting and Modeling PDF Data

Graham King

Brockhouse Beamlines

Canadian Light Source

June 19th, 2024



Canadian
Light
Source Centre canadien
de rayonnement
synchrotron

THE BRIGHTEST LIGHT IN CANADA | lightsource.ca

Different Methods of Modeling PDF Data

- Small Box: Unit cell like description
- Reverse Monte Carlo (large box): Large array of 10^5 atoms.
- Peak Positions and Intensity Integration: For coordination numbers in amorphous or liquid samples.



Small Box Fitting

- Uses a unit cell-like description for Rietveld-like refinement in real space.
- Can refine cell dimensions, atomic positions, and displacement parameters.
- Uses each inter-atomic distance as the center of a peak, which has a Gaussian shape with the width proportional to the thermal parameter.

Advantages and Disadvantages of the Small Box Method

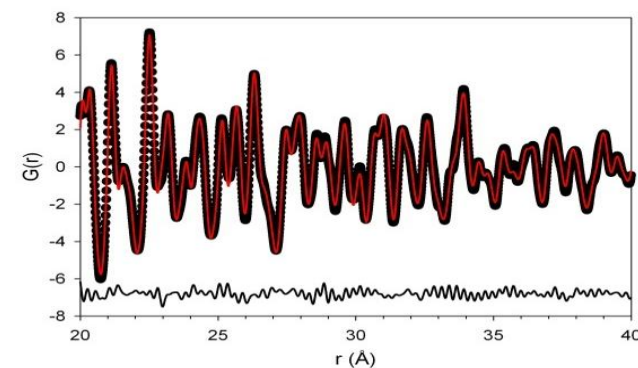
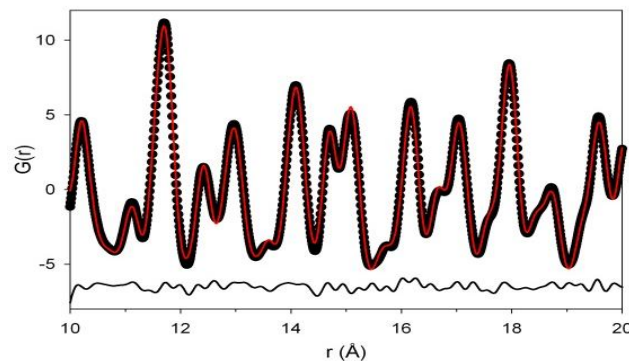
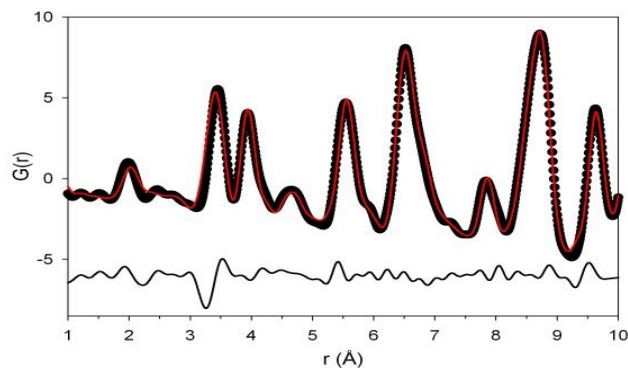
- Refinements are quick and easy to interpret.
- Q dependent dampening and broadening can be modeled.
- Can use space group symmetry to constrain the refinement.
- Can refine multiple phases.
- Cannot handle systems with large numbers of different coordination environments or complex patterns of displacements.
- Cannot be used for amorphous systems.



Box Car Fitting

Often local distortions do not correlate very far, so one model is needed to fit the low- r region of the PDF and the average structure is needed to fit the high- r region.

This reveals over what length scale the local deviations correlate.



A PDF fit with an average structure model over three r -ranges. Notice how the average structure does not fit the low- r region well but does fit the high- r regions well.



Refineable Parameters

Unit cell dimensions and atomic positions determine the positions of peaks.

Thermal parameters define the widths of the peaks.

Correlated motion accounts for the fact that nearby atoms tend to move in-synch, while far apart atoms do not. This sharpens this first few peaks compared to the high- r peaks, particularly the first peak.



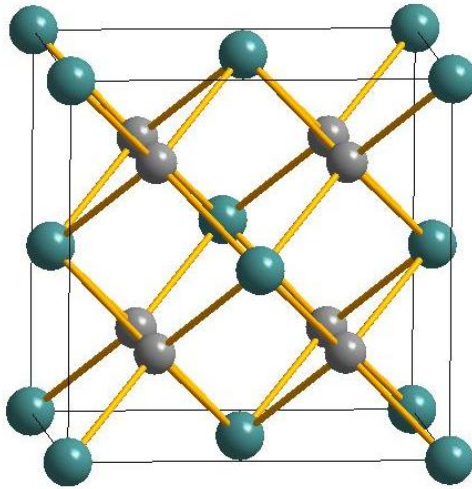
Refine $\delta 2$ (RT or LT) or $\delta 1$ (HT). Alternatively, can refine s_{ratio} using $rcut$.

Q_{damp} and Q_{broad} are instrumental, determine by fitting a highly crystalline standard and fix.

The $sp_{diameter}$ will refine the particle size, if it is roughly spherical.

PDF Simulations

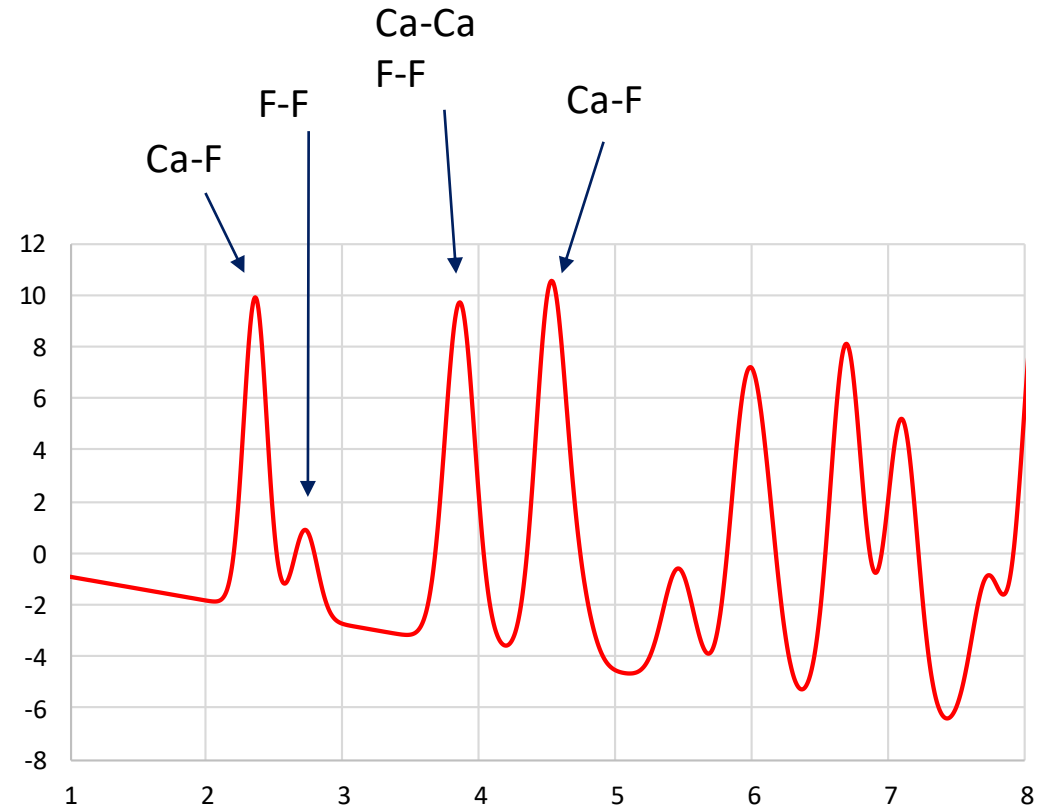
Small box programs like PDFgui can also be used to simulate PDFs. This lets you know what to expect and see how distinguishable the PDFs of different models will be before doing the experiment.



The crystal structure and
PDF of fluorite (CaF_2).

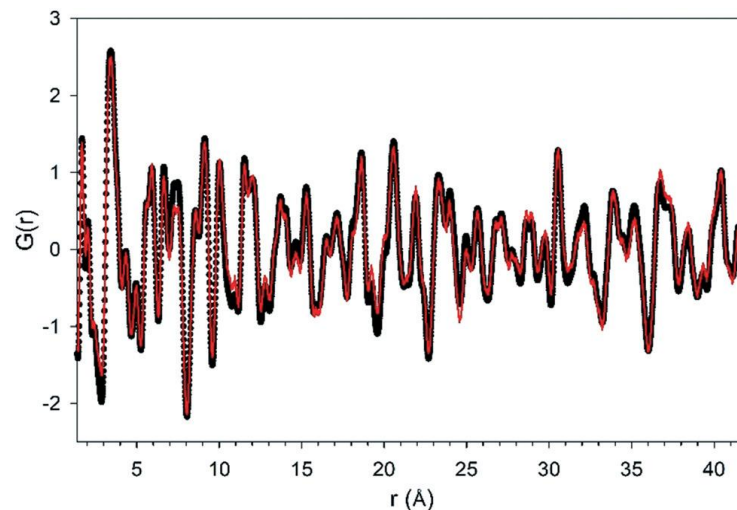
Gray = Ca^{2+}

Teal = F^-



What is RMC Fitting

RMC stands for Reverse Monte Carlo. It is sometimes referred to as large box fitting. It is a global optimization method which uses random moves to fit a configuration of atoms to data.



It is probabilistic, which means it will produce the most random distribution which is consistent with the data. It is flexible and powerful but care must be taken.



The RMC Process

- A large configuration of atoms is generated as a starting point (typically on the order of 10^4)
- An atom is randomly moved a small amount.
- If the move makes the fit better, it is accepted.
- If the move makes the fit worse, it may be rejected or accepted. There is a value σ which determines how much worse the fit can be but still have the move accepted. The quality of the fit is measured as χ^2 .

$$\chi^2 = \sum_j (y_j(\text{exp}) - y_j(\text{calc}))^2 / \sigma^2$$

The probability of a worse move being accepted is proportional to:

$$P = \exp(-\Delta \chi^2 / 2)$$

This allows the refinement to escape from local minima while still progressing towards a better fit.

- This process is repeated millions of times until an equilibrium is reached.

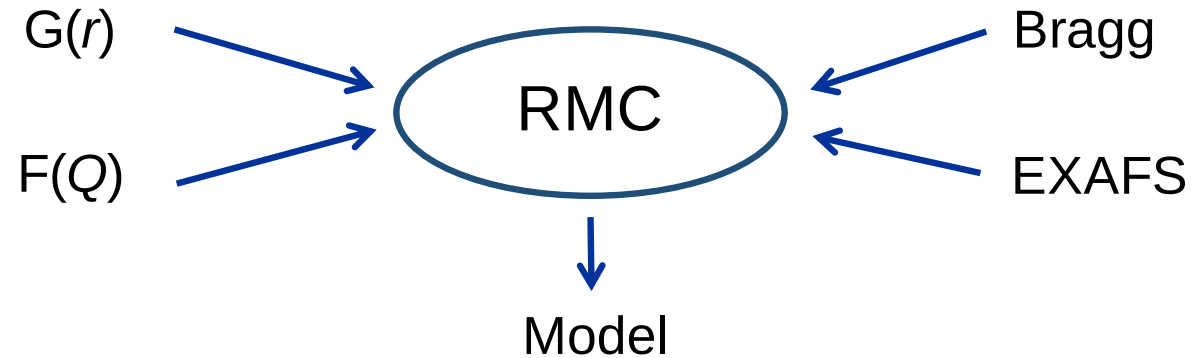


Advantages and Disadvantages of RMC

- Not constrained by symmetry.
 - Easy to produce a starting model.
 - Can produce solutions you did not imagine.
 - Can fit multiple types of data as well as use constraints.
 - Can fit complex systems with a large number of different local environments.
 - Can model the short, medium, and long range order all with the same configuration.
 - Can handle amorphous materials with no long range order.
-
- Computationally expensive, takes a day to several days to run.
 - Interpretation of results trickier due to large number of parameters.



RMC Using Multiple Data Types

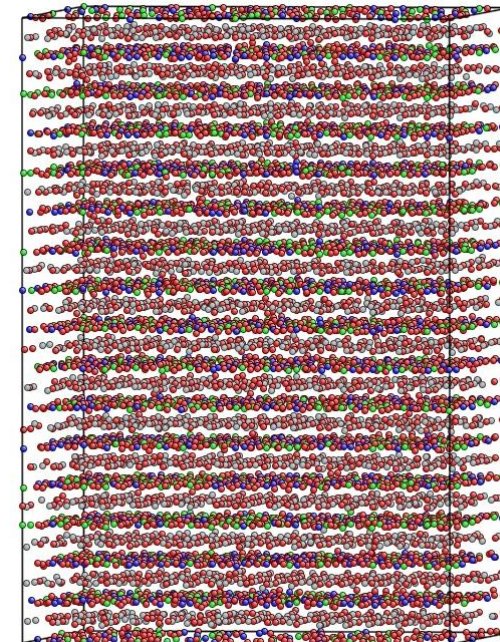
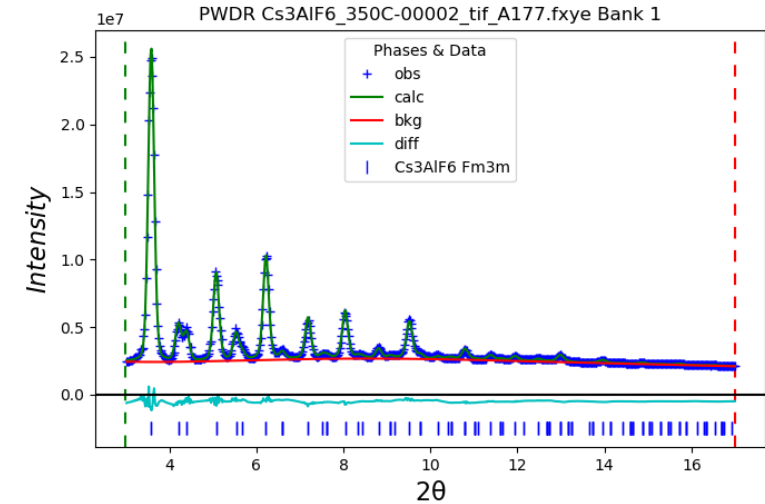


Fitting the PDF and the diffraction pattern ensures that the model is consistent with the structure on all length scales. EXAFS data can also be used as an additional constraint. There are plans to incorporate NMR, Raman, and TEM data as well. It is also possible to fit X-ray and neutron data simultaneously.

Starting Model for Crystalline Systems

Perform a Rietveld refinement, preferably on the same total scattering data the PDF was made from.

Expand the unit cell into a superstructure, usually about 20,000 atoms is a good number to aim for. The PDF can be fit up to half the configurations shortest dimension.



Weighting of Data and Constraints

- Choosing a good σ is critical for getting a reasonable result! For data sets a value between 0.02 and 0.06 seems to work well. If unsure I suggest starting with 0.03 and seeing how it goes.
- Remember, a smaller σ means less accepted moves!
- Aim for an acceptance rate between 30 and 60 %, preferably somewhere around 50% after it has been running at least half an hour. If the rate is too high or too low then it is just smearing out atoms or going straight to the local minima.
- If there are multiple data sets they need to be weighted so they both have about the same effect on the structure. You may need to try some values, see if both data sets are being fit, and then adjust.
- Same with using constraints, the weighting must be high enough to effect the results but not so high as to ignore the data.



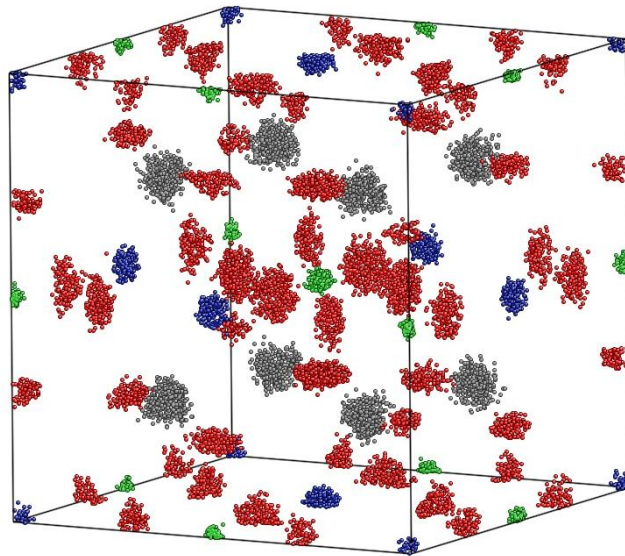
Types of Constraints

Constraints can keep the structure chemically reasonable when there is too much freedom or when certain atoms contribute very weakly to the PDF.

- Bond valence sums: One of the best, flexible in terms of bond lengths as long as the overall bonding is chemically reasonable.
- Distance window: Keeps atoms from getting too far apart.
- Inter-atomic potentials: Keeps reasonable distances, but forces a Gaussian like distribution.
- Bond angle: Useful for keeping polyhedra in tact. Can also use with very low weighting to be able to view the angles distribution.

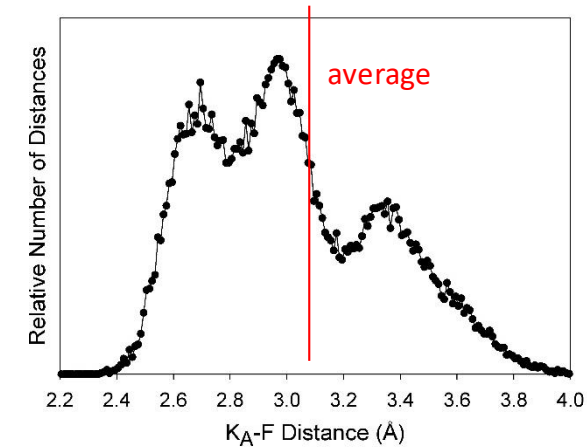
RMC Outputs

Example for the “cubic” HT form of perovskite K_3AlF_6

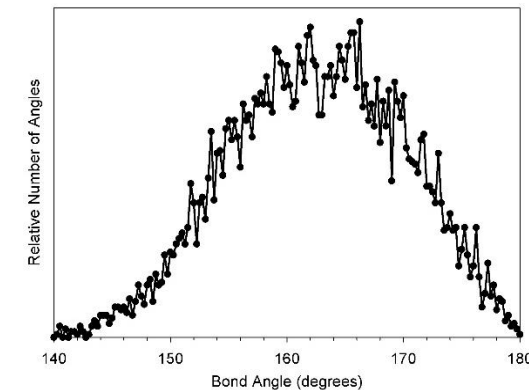


green = Al
blue = K(B-site)
gray = K(A-site)
red = F

The final RMC configuration folded back into a single unit cell to show “clouds” of atom distributions.



Distribution of distances is very different from the single Gaussian expected from the average structure.

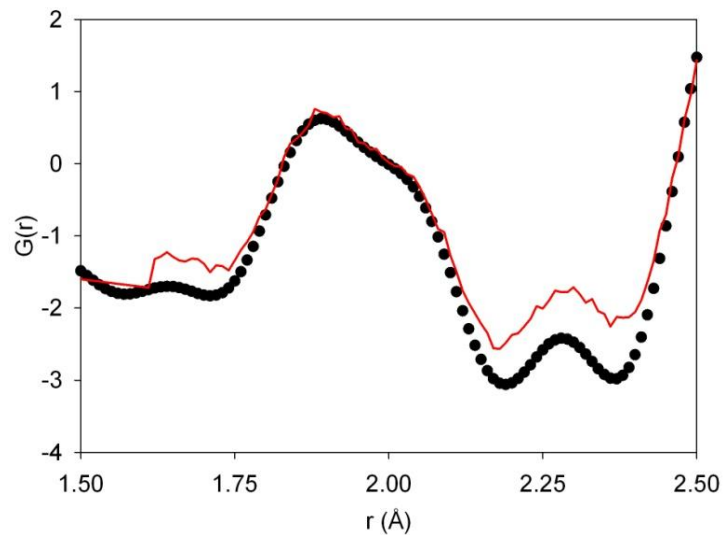


Bond angles differ significantly from the 180° expected from the average structure.



Crystalline Example: $\text{SrFe}_{1/2}\text{Mn}_{1/2}\text{O}_{2.5}$

A “cubic” perovskite with mixed B-site occupancy and random O vacancies.

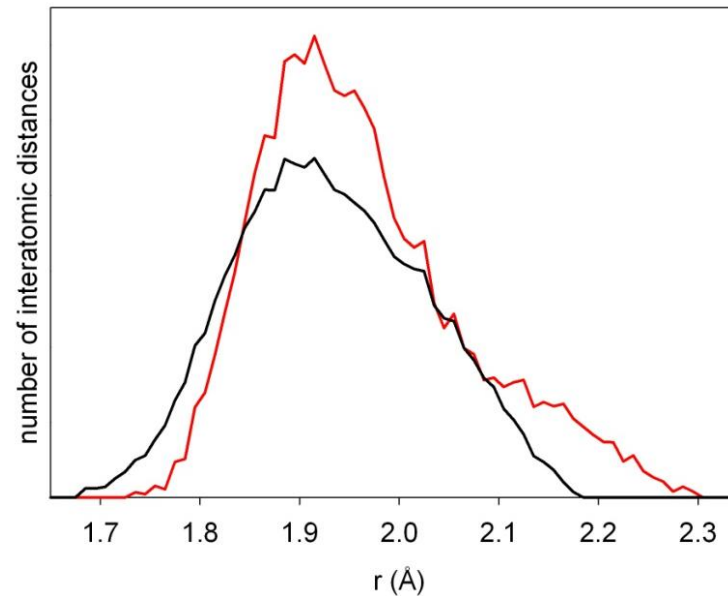


Fitting with a random distribution of Fe/Mn and O/vacancies always led to an overfitting of the first PDF peak.

Models where O vacancies are preferentially located between pairs of Fe gave a better fit.

Ave Fe CN = ~ 4.5

Ave Mn CN = ~ 5.5



Fe-O (black) and Mn-O (red) partials are shown.

Presence of long Mn-O bonds reveals that Mn^{3+} are Jahn-Teller distorted.



Some Notes on Liquid and Amorphous Materials

- Data collection criteria are different as high Q space resolution is not as important since there are no sharp peaks.
- Often coordination numbers are extracted without modeling by integrating the area under a peak in the $R(r)$ function.
- RMC modeling can be done and is quite useful but creating a starting model is more difficult.

How to get the CN of an atom from the area of a peak in $R(r)$:

$$R(r) = rG(r) + 4\pi r^2 \rho_0$$

$$CN = c_j(A)/w_{ij}$$

$$w_{ij} = c_i c_j f_i f_j / (\sum c_i c_f)^2 \quad \sum w_{ij} = 1$$

i is the central atom

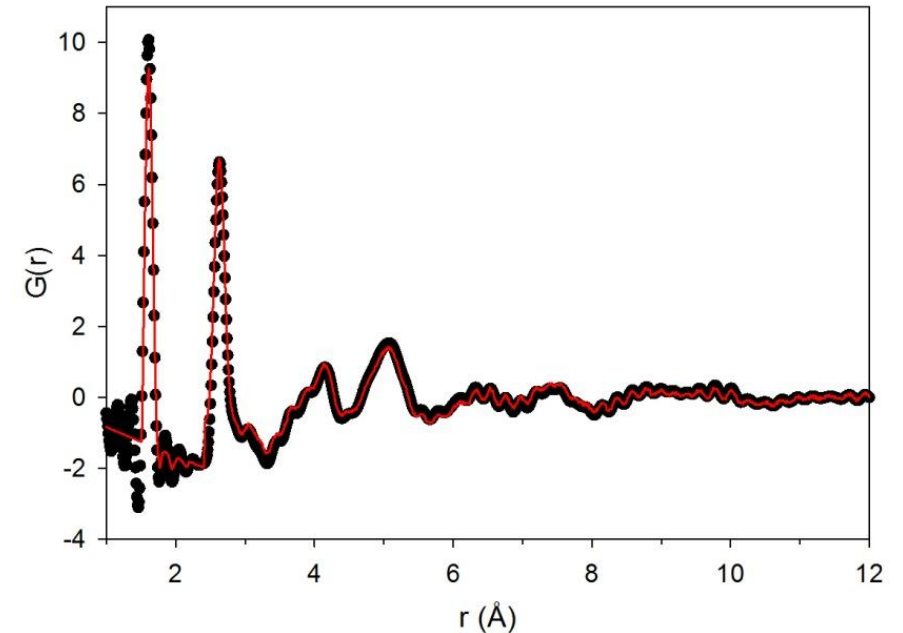
j is the coordinating atom

c is the concentration of the atom

f is the scattering factor

CN is the coordination number

A is the area under a peak



Neutron PDF of amorphous SiO_2 , one of the most abundant glasses, and the RMC fit.

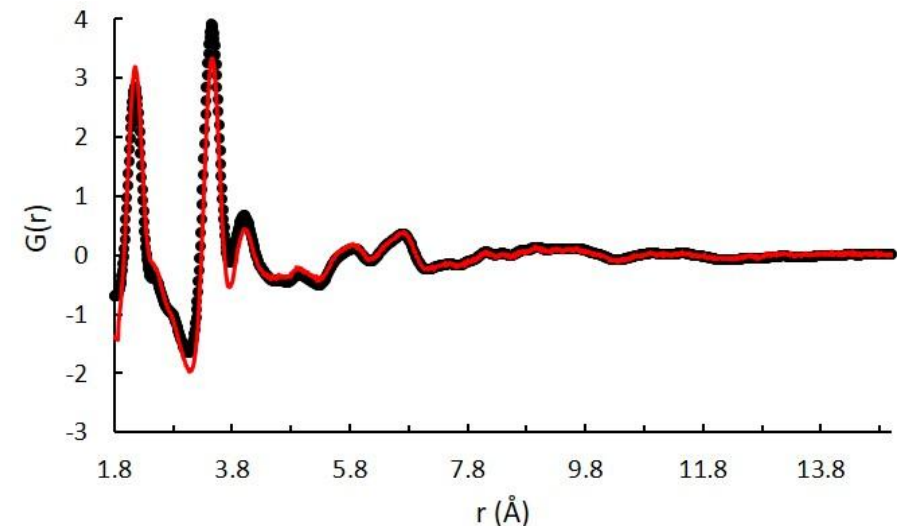
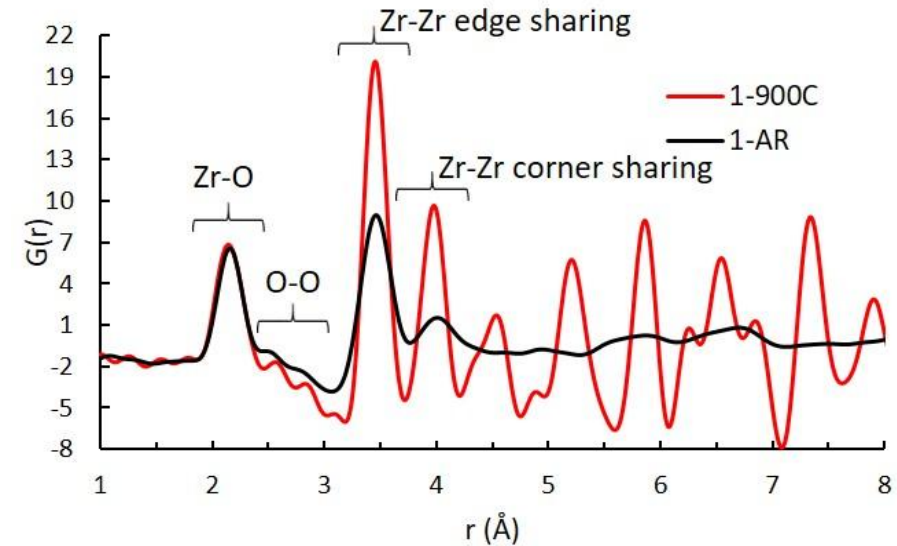
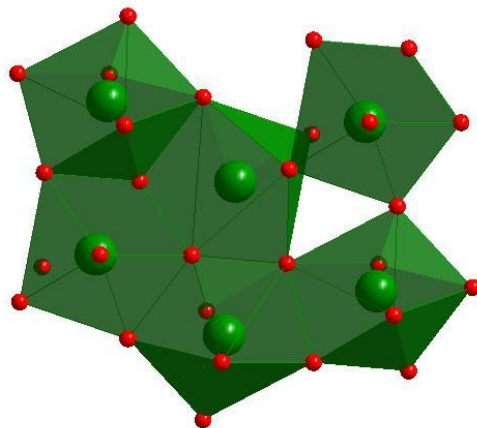


Amorphous Example: $\text{Zr}(\text{OH})_4$

High interest as an absorbent for chemical warfare agents.

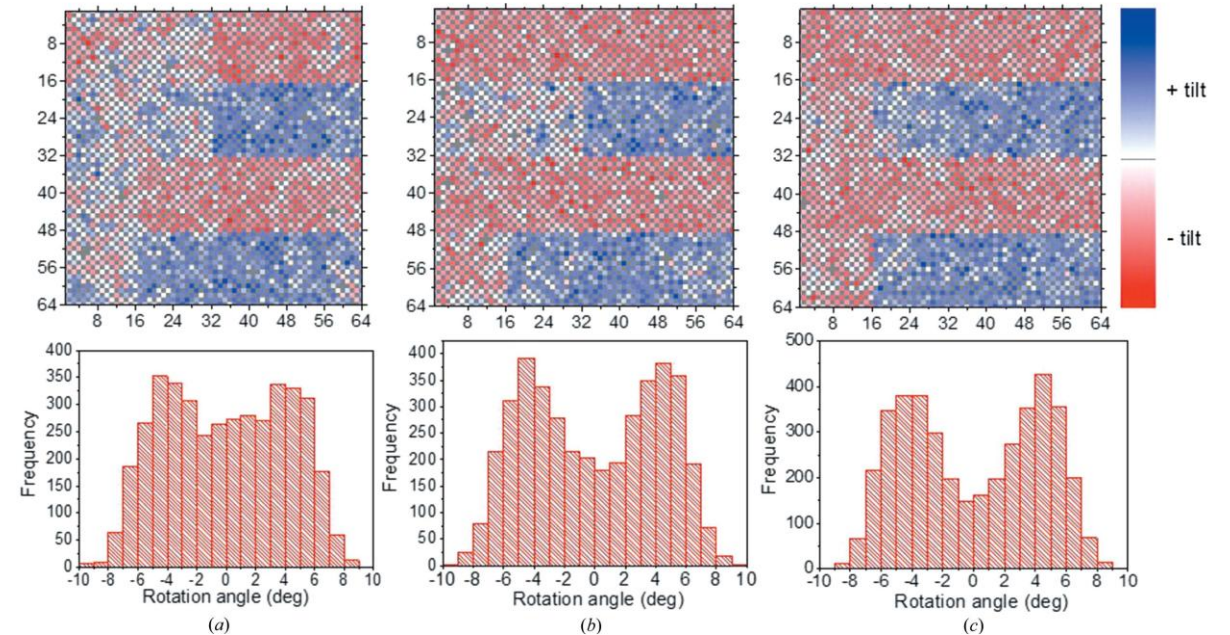
- Integration of area under the first peak in $R(r)$ gives a coordination # of Zr by O of 6.3.
- Starting model made by using large moves and setting a coordination number constraint.
- Then atoms fit against data with large moves and large σ .
- Finally smaller moves used with small σ .

A typical environment around a central ZrO_7 polyhedron.



Advances in RMC Analysis

- Parallel computing to speed up calculations.
- Direct handling of instrumental dampening and broadening.
- Ability to refine distance up to 10 nm.

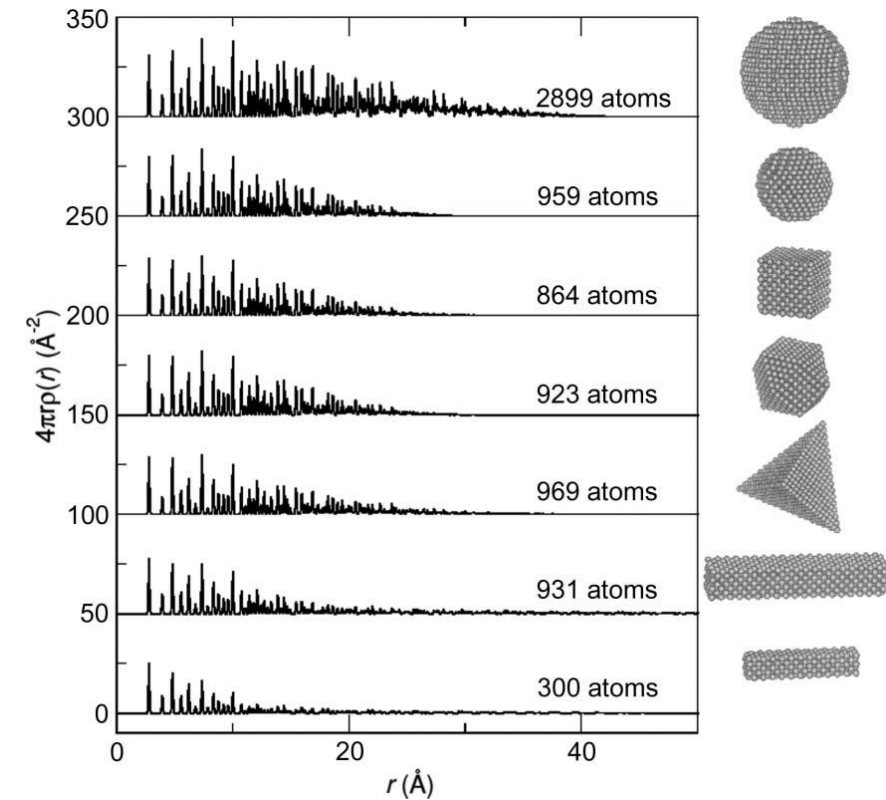
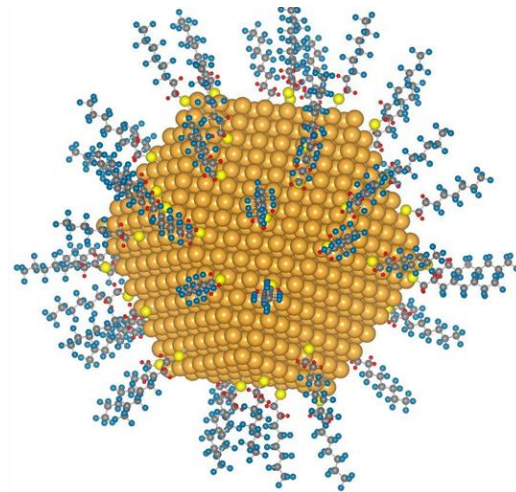


Recommended Reading: M. Eremenko, V. Krazyman, A. Gagin, I. Levin, Advancing reverse Monte Carlo structure refinements to the nanoscale, *J. Appl. Cryst.* (2017) 50, 1561-1570.



Whole Nanoparticle Simulation

- Software, such as the DISCUS program, is available to build and simulate PDFs for entire nanoparticles.
- Useful for modeling the decay of the PDF based on nanoparticle size and shape.
- Can also model effects of surface species, such as capping agents.



K. Page, et al., *J. Appl. Cryst.* (2011) 44, 327-336.

