



JOURNAL OF
APPLIED
CRYSTALLOGRAPHY

Volume 55 (2022)

Supporting information for article:

Stochastic atomic modeling and optimization with *fullrmc*

Bachir Aoun

Experimental data and constraints

In fullrmc, experimental data are incorporated via experimental constraints. Equations 1-4 in the article text define the available forms for the PDF and diffraction data. As we can see in equation 2 in the main article, $K(q)$ and $S(q)$ are functions of the sine Fourier transfer of $G(r)$ but $F(q)$ is function of the sine Fourier transform of $rG(r)$. We can also rewrite $F(q)$ as a function of $R(r)$ Given the following equality

$$\int r \sin(qr) dr = \frac{1}{q} (\sin(qr) - qr \cos(r))$$

$$\begin{aligned} F(Q) &= \frac{4\pi\rho_0}{q} \int_{r_{\min}}^{r_{\max}} rG(r) \sin(qr) dr \\ &= \frac{4\pi\rho_0}{q} \int_{r_{\min}}^{r_{\max}} (R(r) - 4\pi\rho_0 r) \sin(qr) dr \\ &= \frac{4\pi\rho_0}{q} \int_{r_{\min}}^{r_{\max}} R(r) \sin(qr) dr - \frac{(4\pi\rho_0)^2}{q} \int_{r_{\min}}^{r_{\max}} r \sin(qr) dr \\ &= \frac{4\pi\rho_0}{q} \int_{r_{\min}}^{r_{\max}} R(r) \sin(qr) dr - \frac{(4\pi\rho_0)^2}{q^2} (\sin(qr) - r \cos(r)) \Big|_{r_{\min}}^{r_{\max}} \\ &= - \left(\frac{4\pi\rho_0}{q} \right)^2 (\sin(qr) - r \cos(r)) \Big|_{r_{\min}}^{r_{\max}} + \frac{4\pi\rho_0}{q} \int_{r_{\min}}^{r_{\max}} R(r) \sin(qr) dr \end{aligned} \quad S1$$

From a simulation point of view, the computation of $g(r)$, as shown in equation S2 is at the core of all experimental constraints.

$$g(r) = \sum_{i,j} w_{i,j} \frac{\rho_{i,j}(r)}{\rho_0} = \sum_{i,j} w_{i,j} \frac{n_{i,j}(r)/v(r)}{N_{i,j}/V} \quad S2$$

Here r is the real space distance between two atoms, q is the reciprocal space distance or the momentum transfer, ρ_0 is the average number density of the system, $\rho_{i,j}(r)$ is the pair density function of atoms i and j , $w_{i,j}$ is the relative weighting of atom types i and j , N is the total number of atoms, V is the volume of the system, $n_{i,j}(r)$ is the number of atoms i neighboring j at a distance r , $v(r)$ is the annulus volume at distance r and of thickness dr , $N_{i,j}$ is the total number of atoms i and j in the system.

Corrections

When comparing results between measured and simulation computed data, certain normalization and corrections must be considered. As noted in the text we label $\hat{G}(r)$ as the corrected form of $G(r)$ and the experimental atom pair correlation is obtained by a Fourier transform of the measured structure factor. However, given the finite number of simulated atoms and the arbitrary nature of counting statistics in an

experiment, a simple constant scaling factor C_{sf} must be applied to the experimental $G(r)$ as the first step of correction.

$$\hat{G}(r) = C_{sf} \frac{2}{\pi} \int_{-\infty}^{\infty} q R(Q_c, Q_w, q) K(q) \sin(qr) dq \tag{S3}$$

In addition, the measured data accessible Q range is limited between $[Q_{min}, Q_{max}]$ because of the finite size of the detector used in the experimental setup. From a mathematical point of view, this is like multiplying a $[-\infty, \infty]$ unbound Q range structure factor by a unit rectangle function $R(Q_c, Q_w, q)$ centered around $Q_c = 0.5(Q_{max} + Q_{min})$ and $Q_w = Q_{max} - Q_{min}$ wide. Therefore $G(r)$ can be written as in Equation S4.

$$\hat{G}(r) = C_{sf} \frac{2}{\pi} \int_{q_{min}}^{q_{max}} q K(q) \sin(qr) dq = C_{sf} \frac{2}{\pi} \int_{-\infty}^{\infty} q R(Q_c, Q_w, q) K(q) \sin(qr) dq \tag{S4}$$

Given that the Fourier transform of the product is the convolution of the Fourier transforms, and shifting Q range to the origin by subtracting Q_{min} has no effect on the unbound Fourier transform of $R(Q_c, Q_w, q)$, and that the Fourier transform of a rectangle function of unit intensity and Q_w width is equal to ' $Q_w \text{sinc}(Q_w r)$ ' then Equation S4 can be re-written to the following:

$$\begin{aligned} \hat{G}(r) &= C_{sf} \int_{-\infty}^{\infty} R(Q_c, Q_w, q) dq \otimes \frac{2}{\pi} \int_{-\infty}^{\infty} q K(q) \sin(qr) dq \\ &= C_{sf} \int_{-\infty}^{\infty} R(Q_c, Q_w, q) dq \otimes G(r) \\ &= C_{sf} \int_{-\infty}^{\infty} R(0, Q_w, q) dq \otimes G(r) \\ &= C_{sf} Q_w \text{sinc}(Q_w, r) \otimes G(r) \end{aligned} \tag{S5}$$

Q_{max} is commonly misused in place of Q_w in the **sinc** function, but ideally one must consider the non-zero size of the direct beam stop. Nevertheless, fullrnc adopts the widely used Q_{max} correction notation instead of Q_w .

Moreover, another correction must be considered to account for the limited Q space resolution of the experiment. The outcome of all experimental artifacts leading to a finite resolution is a width widening of the measuring units or information leak between neighboring ones. Given the random sampling nature of the collected experimental data, the resolution error statistic can be modeled with a q dependent standard deviation $\sigma(q)$ normal distribution across the whole measure Q range. A good approximation of the resolution correction can be measured using standard material. It's also acceptable and numerically convenient to use a constant $\sigma = \frac{\sigma_q}{\sqrt{2\pi}}$ standard deviation Gaussian resolution function $\Gamma(\sigma_q, q) = \frac{1}{\sigma\sqrt{2\pi}} e^{-q^2/2\sigma^2} = \frac{1}{\sigma_q} e^{-\pi q^2/\sigma_q^2}$ convoluting the simulated structure factor. Using the Fourier transform

property of a Gaussian $\Gamma(\sigma_q, r) = \frac{1}{\sqrt{2\pi}\sigma_q} \int_{-\infty}^{\infty} \Gamma(\sigma_q, q) e^{-2\pi i r q} dq = e^{-\sigma_q^2 r^2}$, the corrected form of $\widehat{G}(r)$ can be now written as the following

$$\widehat{G}(r) = C_{sf} Q_w \text{sinc}(Q_w, r) \otimes \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma} \Gamma(\sigma_q, q) \otimes K(q) \sin(qr) dq = C_{sf} Q_w \text{sinc}(Q_w, r) \otimes e^{-\sigma_q^2 r^2} G(r) \quad S6$$

With Equation S6, we complete the corrections that must be made to $G(r)$ to account for the obvious experimental artifacts, but one additional correction that must be made to correct the simulated configuration $G(r)$ to account for atomic position indetermination and uncertainty. This correction takes the form of another $\Gamma(\sigma_{i,j}, r)$ Gaussian function that separately convolutes Equation S2 with the atom pair correlation $\rho_{i,j}(r)$.

$$\sigma_{i,j} = \sigma'_{i,j} \sqrt{1 - \frac{\delta_1}{r_{i,j}} - \frac{\delta_2}{r_{i,j}^2} + q_{\text{broad}}^2 r_{i,j}^2} \quad S7$$

Here $\sigma_{i,j}$ is the atom pair Gaussian function standard deviation that is composed of atomic pair specific factor $\sigma'_{i,j}$ and configuration factors δ_1 , δ_2 and q_{broad} modeling the r dependence of atomic vibration upon the atomic pair correlation. The simulated atomic structure is a static snapshot of the position of the atoms. In reality, atomic positions are not fixed in space, as they are subject both isotropic thermal vibrations and anisotropic vibrations such as phonons. Therefore, a simulated static position is considered the maximum position likelihood of a normally distributed probability distribution of positions that is r dependent, combining the isotropic and anisotropic position uncertainty an atom can undergo. A more detailed explanation of Equation S7 can be found in Jeong et al. [3] Thus, Equations 7 and 8 in the text are the full and exact formulae that fullrmc uses to compute $\widehat{G}(r)$ and all other pair correlation quantities for respectively PBC and IBC.

Bibliography

- [1] S. J. L. Billinge and T. Egami, "Short-range atomic structure of $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$ determined by real-space refinement of neutron-powder-diffraction data," *PHYSICAL REVIEW B*, no. 47, p. 14386, 1993.
- [2] D. Keen, "A comparison of various commonly used correlation functions for describing total scattering," *Journal of Applied Crystallography*, no. 34, pp. 172-177, 2001.

- [3] I.-K. Jeong, R. H. Heffner, M. J. Graf and S. J. L. Billinge, "Lattice dynamics and correlated atomic motion from the atomic pair distribution function," vol. 67, p. 104301, PHYSICAL REVIEW B .