

# 1 Diffuse Scatter in the Gaussian Approximation

## 1.1 Atomic deviations from their mean positions results in diffuse scattering

Consider the scattering of x-rays from a crystal in the far-field. A general form of the scattered intensity,  $I$ , at some wavevector  $\mathbf{q}$  in reciprocal space is:

$$I(\mathbf{q}) = \left| \sum_c e^{-i\mathbf{q} \cdot \mathbf{u}_c} \sum_i f_i e^{-i\mathbf{q} \cdot \mathbf{r}_i} e^{-i\mathbf{q} \cdot \boldsymbol{\delta}_{ci}} \right|^2 \quad (1)$$

where  $\mathbf{u}_c$  is the vector from the origin of the coordinate system to the origin of unit cell  $c$ ,  $f_i$  is the atomic form factor of atom  $i$ ,  $\mathbf{r}_i$  is the vector that describes the mean position of atom  $i$  relative to origin of the unit cell, and  $\boldsymbol{\delta}_{ci}(t)$  is the instantaneous displacement vector of atom  $i$  in unit cell  $c$ .

Adopting the compact notation  $x_{ij} = x_i - x_j$  for  $\mathbf{u}$ ,  $\mathbf{r}$ , and  $\boldsymbol{\delta}$ , we may write the time (or, equivalently, ensemble) averaged scattering

$$\langle I(\mathbf{q}) \rangle = \left\langle \sum_{c,d} e^{-i\mathbf{q} \cdot \mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} e^{-i\mathbf{q} \cdot \boldsymbol{\delta}_{ci}} e^{-i\mathbf{q} \cdot \boldsymbol{\delta}_{dj}} \right\rangle \quad (2)$$

We make three critical assumptions about the structure of the statistical ensemble of atomic displacements  $\boldsymbol{\delta}$ :

1. The correlation range of atomic displacements is small with respect to the size of the crystal. Equivalently, the correlation between atomic displacements in different unit cells are independent,

$$\langle \boldsymbol{\delta}_{ci}^T \boldsymbol{\delta}_{dj} \rangle = \mathbf{0} \text{ if } c \neq d$$

and therefore we may write the displacements simply using their atomic index, for instance  $\boldsymbol{\delta}_i$ . Note that if correlations exist between what might traditionally be considered a minimal unit cell, and one wishes to consider these correlations, it is possible to simply define a larger unit cell that encompasses the entire correlated region.

2. Atoms in different unit cells behave identically in a statistical fashion, so  $p(\boldsymbol{\delta}_{ci}) = p(\boldsymbol{\delta}_{di})$  for all  $i$ .
3. The atom displacements may be described by a pairwise multivariate normal distribution, with zero mean and covariance matrix  $V_{ij} = \langle \boldsymbol{\delta}_i^T \boldsymbol{\delta}_j \rangle \in \mathcal{R}^{3 \times 3}$

$$p(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) \sim \text{MVN}(\mathbf{0}, V_{ij})$$

this is the simplest model that takes into account anisotropic correlations between atoms.

Since the average is over pairwise probability distributions, we may re-write eq. 2 as

$$\langle I(\mathbf{q}) \rangle = \sum_{c,d} e^{-i\mathbf{q} \cdot \mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} \iint p(\boldsymbol{\delta}_{ci}, \boldsymbol{\delta}_{dj}) e^{-i\mathbf{q} \cdot (\boldsymbol{\delta}_{ci} - \boldsymbol{\delta}_{dj})} d\boldsymbol{\delta}_{ci} d\boldsymbol{\delta}_{dj} \quad (3)$$

in appendix A1 we show

$$\iint p(\boldsymbol{\delta}_{ci}, \boldsymbol{\delta}_{dj}) e^{-i\mathbf{q} \cdot (\boldsymbol{\delta}_{ci} - \boldsymbol{\delta}_{dj})} d\boldsymbol{\delta}_{ci} d\boldsymbol{\delta}_{dj} = \exp \left\{ -\frac{1}{2} \mathbf{q}^T V_{ci ci} \mathbf{q} - \frac{1}{2} \mathbf{q}^T V_{dj dj} \mathbf{q} + \mathbf{q}^T V_{ci dj} \mathbf{q} \right\} \quad (4)$$

under our assumptions,  $V_{ci dj} = \mathbf{0}$  if  $c \neq d$ , and  $V_{ci ci}$  is identical for all  $c$ , such that we may write  $V_{ci ci} = V_{ii}$  (and similarly  $V_{dj dj} = V_{jj}$ ). Thus, it is natural to split eq. 2 into two parts – one expressing interference *between* unit cells (where  $V_{ci dj} = \mathbf{0}$ ), and one expressing interference *within* repeats of a single cell,

$$\langle I(\mathbf{q}) \rangle = \sum_{c,d \neq c} e^{-i\mathbf{q} \cdot \mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} e^{-\frac{1}{2} \mathbf{q}^T V_{ii} \mathbf{q} - \frac{1}{2} \mathbf{q}^T V_{jj} \mathbf{q}} \quad (5)$$

$$+ N \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} e^{-\frac{1}{2} \mathbf{q}^T V_{ii} \mathbf{q} - \frac{1}{2} \mathbf{q}^T V_{jj} \mathbf{q} + \mathbf{q}^T V_{ij} \mathbf{q}} \quad (6)$$

$$= \sum_{c,d} e^{-i\mathbf{q} \cdot \mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} e^{-\frac{1}{2} \mathbf{q}^T V_{ii} \mathbf{q} - \frac{1}{2} \mathbf{q}^T V_{jj} \mathbf{q}} \quad (7)$$

$$+ N \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} e^{-\frac{1}{2} \mathbf{q}^T V_{ii} \mathbf{q} - \frac{1}{2} \mathbf{q}^T V_{jj} \mathbf{q}} \left[ e^{\mathbf{q}^T V_{ij} \mathbf{q}} - 1 \right] \quad (8)$$

with  $N$  being the number of unit cells.

The astute reader will notice the first term as the traditional crystallographic scattering

$$I(\mathbf{q})_{\text{Bragg}} = \sum_{c,d} e^{-i\mathbf{q}\cdot\mathbf{u}_{cd}} \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii} \mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj} \mathbf{q}} \quad (9)$$

$$= \left| \left( \sum_c e^{-i\mathbf{q}\cdot\mathbf{u}_c} \right) \left( \sum_i f_i e^{-i\mathbf{q}\cdot\mathbf{r}_i} e^{-\frac{1}{2}\mathbf{q}^T V_{ii} \mathbf{q}} \right) \right|^2 \quad (10)$$

$$= |S(\mathbf{q})|^2 |F(\mathbf{q})|^2 \quad (11)$$

where  $V_{ii}$  is an anisotropic B-factor (also called the Debye Waller factor).  $|S(\mathbf{q})|^2$  becomes a Dirac comb as the number of unit cells grows, showing this scattering is localized to discrete regions of  $\mathbf{q}$ .

The remaining scattering  $\langle I(\mathbf{q}) \rangle - I(\mathbf{q})_{\text{Bragg}}$  is typically termed the diffuse scattering

$$I_{\text{diffuse}}(\mathbf{q}) = N \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} e^{-\frac{1}{2}\mathbf{q}^T V_{ii} \mathbf{q} - \frac{1}{2}\mathbf{q}^T V_{jj} \mathbf{q}} \left[ e^{\mathbf{q}^T V_{ij} \mathbf{q}} - 1 \right]$$

There are two notable features of the diffuse scattering. First, lacking the lattice transform  $|S(\mathbf{q})|^2$  it is *not* localized in reciprocal space. Second, it is non-trivial only if there is correlated displacements between the atoms  $V_{ij} \neq \mathbf{0}$ .

## 1.2 Diffuse Scatter in Special Cases

### 1.2.1 No Cross-Correlation

In the case  $V_{ij}$  is  $\mathbf{0}$  for all  $\{i, j\}$ , then all  $i \neq j$  terms vanish, and

$$I_{\text{diffuse}}(\mathbf{q}) = N \sum_i f_i^2 \left[ 1 - e^{-\mathbf{q}^T V_{ii} \mathbf{q}} \right]$$

just a relatively unstructured background remains, proportional to the anisotropic B-factors. Note anisotropy is still possible if the individual atomic displacements are highly anisotropic. The diffuse scatter will, however, lack the “speckle” features that normally distinguish it from background scattering.

### 1.2.2 Correlated Bodies

In the case that an entire set of atoms moves as a unit, the covariance between all pairs of atoms  $V_{ij}$  can be written as a scalar  $\sigma^2$ , and the diffuse scatter becomes

$$\begin{aligned} I_{\text{diffuse}}(\mathbf{q}) &= N \sum_{i,j} f_i f_j e^{-i\mathbf{q}\cdot\mathbf{r}_{ij}} \left[ 1 - e^{-q^2 \sigma^2} \right] \\ &= N \left[ 1 - e^{-q^2 \sigma^2} \right] |F(q)|^2 \end{aligned} \quad (12)$$

The diffuse scatter in this case is a scaled version of the molecular transform intensity. This expression was derived by Moore and later Chapman.

### 1.2.3 Isotropic Real Space Correlation Length

Here, we reproduce the isotropic correlation model of Clarage. Our approach is equivalent but more direct, employing the expression for the scattered intensity rather than reasoning about the Patterson. Clarage makes the following assumptions about the atomic correlations:

1. Isotropic, distance dependent interatomic correlation:  $V_{ij} = \sigma^2 \Gamma(\mathbf{r}_{ij})$  with  $\Gamma$  a kernel describing the correlation length.
2. Isotropic B-factors:  $V_{ii} = \sigma^2$  for all  $i$ .

Using these simplifying approximations, we may write the diffuse scatter

$$I_{\text{diffuse}}(\mathbf{q}) = N \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} e^{-q^2 \sigma^2} \left[ e^{q^2 \sigma^2 \Gamma(\mathbf{r}_{ij})} - 1 \right]$$

Clarage suggests that so long as  $q^2 \sigma^2 \Gamma(\mathbf{r}_{ij})$  is small, we may Taylor expand the exponential  $\exp\{q^2 \sigma^2 \Gamma(\mathbf{r}_{ij})\} \approx 1 + q^2 \sigma^2 \Gamma(\mathbf{r}_{ij})$  and further simplify

$$I_{\text{diffuse}}(\mathbf{q}) = q^2 \sigma^2 e^{-q^2 \sigma^2} N \sum_{i,j} f_i f_j e^{-i\mathbf{q} \cdot \mathbf{r}_{ij}} \Gamma(\mathbf{r}_{ij})$$

For this approximation to hold,  $q^2 \sigma^2 \Gamma(\mathbf{r}_{ij}) \ll 1$ . Since  $q \sim 1\text{\AA}^{-1}$ ,  $\sigma \ll 1\text{\AA}$ , and  $\Gamma(\mathbf{r}_{ij}) \leq 1$  this is not a bad approximation. To assist in the analysis of this function, define a new kernel function

$$s(\mathbf{q}, \mathbf{r}) = \sum_{i,j} f_i f_j \delta(\mathbf{r} - \mathbf{r}_{ij})$$

with  $\delta(\cdot)$  the Dirac delta (not a displacement). Using this kernel, we may re-write the diffuse scattering as a Fourier transform, rather than a discrete sum

$$I_{\text{diffuse}}(\mathbf{q}) = q^2 \sigma^2 e^{-q^2 \sigma^2} N \int s(\mathbf{q}, \mathbf{r}) \Gamma(\mathbf{r}) e^{-i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r}$$

using the Fourier convolution theorem (specifically, here that  $\mathcal{F}[f \cdot g] = \mathcal{F}^{-1}[f] * \mathcal{F}^{-1}[g]$ , with  $*$  denoting convolution), we can split the Fourier integral

$$I_{\text{diffuse}}(\mathbf{q}) = q^2 \sigma^2 e^{-q^2 \sigma^2} \frac{N}{16\pi^6} \left[ \left( \int s(\mathbf{q}, \mathbf{r}) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \right) * \left( \int \Gamma(\mathbf{r}) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} \right) \right]$$

let  $\tilde{\Gamma}(\mathbf{q}) = \int \Gamma(\mathbf{r}_{ij}) e^{i\mathbf{q} \cdot \mathbf{r}_{ij}} d\mathbf{r}$  and notice

$$\int s(\mathbf{q}, \mathbf{r}) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} = \int \sum_{i,j} f_i f_j \delta(\mathbf{r} - \mathbf{r}_{ij}) e^{i\mathbf{q} \cdot \mathbf{r}} d\mathbf{r} = \sum_{i,j} f_i f_j e^{i\mathbf{q} \cdot \mathbf{r}_{ij}} = I_0$$

is the scattered intensity of the coherently diffracting volume. This could either be a number of asymmetric units or the entire crystal (note: poorly presented, need to clean this up). In summary, we may write the diffuse scatter as a convolution between this intensity function and the Fourier transform of the correlation kernel  $\Gamma$ ,

$$I_{\text{diffuse}}(\mathbf{q}) = q^2 \sigma^2 e^{-q^2 \sigma^2} \frac{N}{16\pi^6} \left[ I_0 * \tilde{\Gamma}(\mathbf{q}) \right] \quad (13)$$

### 1.3 Inference of the covariance matrix

Let us assume that we have many measurements of  $I_{\text{diffuse}}(\mathbf{q})$  and wish to infer a model of the form just discussed. To do this, we may minimize the least squares loss function

$$\mathcal{O}(\{V_{ij}\}) = \sum_{\{\mathbf{q}\}} |I_{\text{diff}}^{\text{obs}}(\mathbf{q}) - I_{\text{diff}}^{\text{calc}}(\mathbf{q}; V)|^2$$

where “obs” and “calc” indicate the observed and modeled values respectively. Our task is to vary  $V = \{V_{ij}\}$  to minimize  $\mathcal{O}$ . We will assume that each  $V_{ij}$  is *isotropic* and therefore can be represented by a scalar. The extension to the anisotropic case is straightforward, but as we will see, the number of free parameters in the isotropic model will already be quite large.

Let  $A_{ij}(\mathbf{q}) \equiv f_i f_j \exp\{i\mathbf{q} \cdot (\mathbf{r}_i - \mathbf{r}_j) - \frac{1}{2} q^2 (V_{ii} + V_{jj})\}$ , Then we may write

$$I_{\text{esb}}^{\text{calc}}(\mathbf{q}; V) = \sum_{ij} A_{ij}(\mathbf{q}) \left[ e^{q^2 V_{ij}} - 1 \right]$$

and the derivatives of the objective function take on a simple form. Employing the chain rule we can write

$$\frac{\partial \mathcal{O}}{\partial V_{ij}} = \sum_{\{\mathbf{q}\}} 2 [I_{\text{diff}}^{\text{obs}}(\mathbf{q}) - I_{\text{diff}}^{\text{calc}}(\mathbf{q}; V)] \frac{\partial}{\partial V_{ij}} I_{\text{diff}}^{\text{calc}}(\mathbf{q}; V) \quad (14)$$

$$= \sum_{\{\mathbf{q}\}} 2 [I_{\text{diff}}^{\text{obs}}(\mathbf{q}) - I_{\text{diff}}^{\text{calc}}(\mathbf{q}; V)] \cdot A_{ij}(\mathbf{q}) q^2 e^{q^2 V_{ij}} \quad (15)$$

No closed form for the minimum exists, but one can see this numerical derivative prescribes a straightforward descent algorithm with step size  $I_{\text{diff}}^{\text{obs}}(\mathbf{q}) - I_{\text{diff}}^{\text{calc}}(\mathbf{q}; V)$  in direction  $A_{ij}(\mathbf{q}) q^2 e^{q^2 V_{ij}}$  for each  $V_{ij}$ . The derivative is zero and the minimum is reached when  $I_{\text{diff}}^{\text{obs}}(\mathbf{q}) = I_{\text{diff}}^{\text{calc}}(\mathbf{q}; V)$ .

Computing second derivatives,

## 1.4 Appendix A1

We prove

$$\iint p(\boldsymbol{\delta}_{c_i}, \boldsymbol{\delta}_{d_j}) e^{-i\mathbf{q} \cdot (\boldsymbol{\delta}_{c_i} - \boldsymbol{\delta}_{d_j})} d\boldsymbol{\delta}_{c_i} d\boldsymbol{\delta}_{d_j} = \exp \left\{ -\frac{1}{2} \mathbf{q}^T V_{c_i c_i} \mathbf{q} - \frac{1}{2} \mathbf{q}^T V_{d_j d_j} \mathbf{q} + \mathbf{q}^T V_{c_i d_j} \mathbf{q} \right\} \quad (16)$$

First, recall the characteristic function of a multivariate normal distribution is

$$\int \cdots \int f_{\boldsymbol{\mu}, \Sigma}(\mathbf{x}) e^{-i\mathbf{u} \cdot \mathbf{x}} d\mathbf{x} = \exp \left\{ i\mathbf{u} \cdot \boldsymbol{\mu} - \frac{1}{2} \mathbf{u}^T \Sigma \mathbf{u} \right\} \quad (17)$$

this can be shown by performing the Gaussian integral on the lhs directly.

We will write (16) in canonical MVN form and then employ (17). To do this, introduce the augmented variables

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{\delta}_{c_i} \\ -\boldsymbol{\delta}_{d_j} \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \begin{bmatrix} \mathbf{q} \\ \mathbf{q} \end{bmatrix}$$

which are vectors in  $\mathcal{R}^6$ . The covariance matrix of  $\mathbf{x}$  may be written in block form

$$\Sigma = \langle \mathbf{x}^T \mathbf{x} \rangle = \begin{bmatrix} V_{c_i c_i} & -V_{c_i d_j} \\ -V_{d_j c_i} & V_{d_j d_j} \end{bmatrix}$$

simply using the definition  $V_{ab} = \langle \boldsymbol{\delta}_a^T \boldsymbol{\delta}_b \rangle$ . Recall also that  $\langle \mathbf{x} \rangle = \mathbf{0}$ . Now, we may write (16) in terms of  $\mathbf{x}$  and  $\mathbf{q}_2$ ,

$$= \int p(\mathbf{x}) e^{-i\mathbf{q}_2 \cdot \mathbf{x}} d\mathbf{x}$$

employing (17) and simplifying completes the proof.

## 2 Map Construction & Volumetric Corrections

Our goal is to take a set of pixel intensities from a rotation series and construct an accurate reciprocal space map. In order to do this, we need to account for the fact that each pixel integrates a different shaped volume of reciprocal space.

For a pixel  $i$  exposed to the scattering from a continuously rotated crystal, we may write measured intensity (up to a constant, arbitrary factor) at that pixel as

$$I_i = \frac{1}{\ell_i} \int_{\Omega_i} I(\mathbf{q}) d\mathbf{q} \quad (18)$$

where  $I(\mathbf{q})$  is the continuous scattered intensity,  $\ell_i$  is the arc length of the trajectory the pixel transverses due to crystal rotation, and  $\Omega_i$  is the 3D volume of reciprocal space the pixel integrates. We divide by the arc length  $\ell_i$  because during rotation, for an Ewald sphere of fixed width, the diffraction is “spread” along the rotation direction (NEED a much better way to explain this – currently my english thinking is murky).

Further, let  $V_i$  be the volume of the region  $\Omega_i$ . By Pappus' second theorem,  $V_i = A\ell_i$ , where  $A$  is the pixel's cross-sectional area in 3D reciprocal space. We assume that a solid angle correction has been applied to all the images; in which case, this value will be identical for every pixel.

Let  $M_\alpha$  be the estimated intensity for map voxel  $\alpha$ . If we construct the map by averaging pixel intensities,

$$M_\alpha = \frac{1}{n_\alpha} \sum_{i \in S_\alpha} I_i$$

where  $S_\alpha$  is the set of all pixels  $i$  mapped to voxel  $\alpha$ , and  $n_\alpha$  is just the cardinality of  $S_\alpha$ . Let  $\bar{I}_i$  be the average value of  $I(q)$  across pixel  $i$ . Then we may write

$$\begin{aligned} M_\alpha &= \frac{1}{n_\alpha} \sum_i \frac{1}{\ell_i} \int_{\Omega_i} I(\mathbf{q}) d\mathbf{q} \\ &= \frac{1}{n_\alpha} \sum_i \frac{\bar{I}_i V_i}{\ell_i} \\ &= A \langle \bar{I}_i \rangle_\alpha \end{aligned}$$

Note  $\langle \bar{I}_i \rangle_\alpha$  is the average value of  $I(\mathbf{q})$  across the volume of all pixels that we have assigned to a particular voxel. This is exactly what we want. We get it as long as (1) we average pixels in a voxel, (2) apply a solid angle correction to the images, and (3) equation (18) is correct. Note we have not needed to resort to any arguments about where the pixels are in reciprocal space.

Note there is some inherent error in this method – where the volume integrated by pixels falls outside the map voxel it is assigned to (as will happen to pixels on the boundaries of voxels), small errors will be introduced. This can be checked by looking at symmetry-equivalent voxels and ensuring their values differ only by small amounts.

**Comment:** One might wonder why we use a Lorentz correction when integrating Bragg data. Bragg data are processed in an entirely different way, summing pixels rather than generating a true reciprocal map. In that case, the volume of reciprocal space included in each integrated Bragg peak is different, and therefore it may be simpler to apply a Lorentz correction to account for the time that peak spends in the diffraction condition.