An effective introduction to structural crystallography using 1D Gaussian atoms

Emily Smith\textsuperscript{1}, Gwyndaf Evans\textsuperscript{2} and James Foadi\textsuperscript{2} \textsuperscript{©}

\textsuperscript{1}Gonville and Caius College, Trinity Street, Cambridge, United Kingdom
\textsuperscript{2}Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot, United Kingdom

E-mail: james.foadi@diamond.ac.uk

Received 30 December 2016, revised 14 June 2017
Accepted for publication 24 July 2017
Published 27 September 2017

Abstract

The most important quantitative aspects of computational structural crystallography can be introduced in a satisfactory way using 1D truncated and periodic Gaussian functions to represent the atoms in a crystal lattice. This paper describes in detail and demonstrates 1D structural crystallography starting with the definition of such truncated Gaussians. The availability of the computer programme CRONE makes possible the repetition of the examples provided in the paper as well as the creation of new ones.

Keywords: crystallography, Fourier series, Fourier transform, anomalous scattering, computational methods for structural crystallography

(Some figures may appear in colour only in the online journal)

1. Introduction

X-ray crystallography is the most widely used technique to determine the shape of molecules at atomic resolution. The complexity of structures found has increased considerably since William Lawrence Bragg used diffraction images from various crystals to calculate the position of individual atoms in a crystal cell for sodium chloride, zinc sulphide and diamond \cite{1}. Accordingly, the amount and difficulty of algorithms and calculations involved in data collection and structure solution have increased notably. This means that only researchers with strong quantitative and computational backgrounds are capable of designing and
implementing adequate algorithms for processing data from the diffraction of X-ray crystals. Traditionally, physicists have been the largest group of investigators providing both innovation and calculating power in structural crystallography. With time, scholars from other sectors, chemistry and biology, have devoted substantial resources and manpower to the creation of vast computer program collections that can be easily accessed and used by scientists without the habitual quantitative background [2, 3]. Maintaining such systems and, especially, making sure they are subject to appropriate rounds of innovation requires the ability to recruit skilled personnel with adequate preparation in all quantitative aspects of structural crystallography. With its wide spectrum of subjects taught within condensed matter theory, physics is still the most represented among developers of methods in structural crystallography. This paper is hence generally aimed at physicists, with a particular eye to those engaged or interested in condensed matter theory. The paper is especially going to be valuable to students with an interest in becoming developers of methods and software in structural crystallography.

Structural crystallography is indeed a well known and mature subject. But it is riddled with specific jargon and requires familiarity with the concept of symmetry and especially all mathematical and computational aspects of the Fourier series and transform. Also, probability theory and statistical calculations are indispensable when processed data are used to determine the atomic positions. The learning curve is, therefore, quite steep. Things are furthermore complicated by the need to visualize and carry forward all calculations in 3D. The present paper aims at demonstrating that it is possible to learn all major and important quantitative concepts of structural crystallography using one, rather than three dimensions. The key to this demonstration is the representation of individual atoms with truncated and periodical 1D Gaussians, suitable to mimic crystal structures in the unit cell. The truncation of Gaussian functions is necessary in this context to avoid infinite sums and integrals whose integrands are infinite sums in most calculations. A periodic and regular pattern with shifted Gaussians is, indeed, rigorously obtainable only if the number of Gaussians is infinite. Working with truncated Gaussians, in contrast, implies integrals with just one integrand, despite the crystal structure being represented by an infinite lattice.

1D crystallography is not, of course, a novel idea. It has recently been discussed and demonstrated in detail by Pinkerton [7] using carbon atoms, following an example similar to that of Stout and Jensen [8]. Such educational treatment of the subject is effective, but limited in scope. Challenges represented by most crystallographic calculations are barely addressed in these works. Quite differently, the original approach of the present paper lies in the extent to which all crystallographic ideas and formulas are developed, in the generality of 1D structures employed as examples (not simply made of carbon atoms), in the rigorous nature of the mathematics involved and, especially, in the creation of a specific computer program to carry out all calculations, plots and pictorial representations. An introduction to crystallography via 1D rather than 3D structures is valuable partly because it clarifies concepts and ideas, and because it makes mathematical formulas less cumbersome and diagrams more intuitive, thus allowing for a smoother initiation into a convoluted technical subject. Most importantly, integrals and Fourier series in 1D are well taught and received by students during the first two years of undergraduate studies. The greatest majority of students in quantitative subjects are certainly capable of understanding and interpreting the 1D mathematics of Fourier series and Fourier integrals. For this topic, the key passages in the calculations are all present in the 1D case. Moving up to 2D or 3D brings in an unnecessary pedagogical increase in complicated factorisations that only serve the purpose of making computations longer, heavier and, thus, less digestible. Another advantage of using a working model of a crystal structure and its Fourier counterparts in 1D, rather than 2D or 3D, is the speed and ease with which ideas and
concepts in real space can be translated algorithmically and visually into reciprocal space. This fact is well known to methods developers in x-ray structural crystallography, when testing the initial correctness and effectiveness of their novel algorithms on 1D models (e.g. the infinite chain of carbon atoms).

In what follows, details and calculations are sketched in the main text and expanded in the appendices. A computer package, CRONE [4], developed for the R statistical platform [5] and freely distributed in the CRAN repository [6], can be downloaded for free and employed to closely follow the details of the exposure.

2. The crystal structure of 1D Gaussians

The most important mathematical feature of a crystal structure is its translational periodicity, for which the repeating portion of space is called a unit cell. Gaussians are desirable functions to represent the core electron density of atoms in the unit cell because they peak around the centre of their distribution, thus representing well the cloud of core electrons around the nucleus, and because they can be handled reasonably well from an analytical point of view. Unfortunately they are not periodic functions. One way to regain the lost periodicity is through the convolution of the Gaussian with the lattice function,

$$L_u(x) \equiv \sum_{u=-\infty}^{+\infty} \delta(x - ua).$$

Indeed, a 1D crystal structure can be modelled as a function \(\psi(x)\) equal to the convolution of the Gaussian centred at 0 and the lattice function:

$$\psi(x) \equiv [L_u(x)] \ast \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$= \int_{-\infty}^{+\infty} \sum_{u=-\infty}^{+\infty} \delta(x' - ua) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x' - x)^2}{2\sigma^2}\right] \, dx'$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{u=-\infty}^{+\infty} \exp\left[-\frac{(x - ua)^2}{2\sigma^2}\right].$$

It is easy to prove that the obtained result, the infinite summation of regularly shifted Gaussians, is a periodic function. Other properties can be worked out starting from its definition. But every calculation requires the summation of converging series, most of them involving the use of special functions (see appendix A). It is more convenient, pedagogically, to leave unaltered the analytic form of the Gaussian within one unit cell and to impose some kind of ‘forced’ periodicity whereby Gaussian tails are truncated so as to have their domain exactly equal to one period, i.e. equal to the unit cell length, \(a\). In this section it will be shown how atoms and groups of atoms suitable for crystallographic calculations can be derived starting from the full non-periodic Gaussian function and ending with a truncated but periodic one. The resulting analytic formula is formally equivalent to a standard Gaussian function with a normalisation constant depending on cell’s size and atom’s width.
2.1. Gaussian atoms

A single 1D atom will be investigated first. As stated earlier, the following Gaussian function

$$\rho(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right]$$

extends from $-\infty$ to $+\infty$, is not periodical, but it is normalised, the area under its curve being 1. One way to transform it into a periodical function is to cut symmetrically its two tails and stick each end to the endpoints of two other identical Gaussians, truncated in the same way. In one dimension this is how to proceed (see figure 1). Both the left and right tails are truncated at a distance $a/2$ from the central peak, i.e. the two cuts occur at $x_0 - a/2$ and $x_0 + a/2$. Infinite copies of the truncated curve are, next, shifted to the left and to the right of $x_0$ in discrete amounts equal to multiples of the unit cell length, $a$. The curve thus obtained is continuous but its first derivative is not, its discontinuities falling at $x_n = (2n+1)a/2, n = 0, 1, 2, \ldots$. The correct analytic form for this curve, limited to interval $[0, a]$, depends on whether $x_0$ is in the first half or the second half of the unit cell:

1. If $0 \leq x_0 \leq a/2$,

$$\rho(x) = \begin{cases} K \exp\left[-(x-x_0)^2/(2\sigma^2)\right] & \text{for } 0 \leq x \leq x_0 + a/2, \\ K \exp\left[-(x - (x_0 + a))^2/(2\sigma^2)\right] & \text{for } x_0 + a/2 \leq x \leq a. \end{cases}$$

Figure 1. Procedure to turn a standard Gaussian into a periodic function. Initially the Gaussian is centred at $x_0$ and its tails extend to $-\infty$ and $+\infty$ (red, dashed curve). This is then truncated at $x_0 - a/2$ and $x_0 + a/2$ (the two points marked with the two short blue segments). Copies of the truncated curve are created and shifted to the left and to the right by $na$ for any integer $n$ (black curve). In this specific example $a = 20$. 
If \( a/2 \leq x_0 \leq a \),

\[
\rho(x) = K \exp\left\{ -\left[ x - (x_0 - a) \right]^2 / \left( 2\sigma^2 \right) \right\} \quad \text{for} \quad 0 \leq x \leq x_0 - a/2,
\]

\[
\rho(x) = K \exp\left\{ -(x - x_0)^2 / (2\sigma^2) \right\} \quad \text{for} \quad x_0 - a/2 \leq x \leq a.
\]

(5)

\( K \) is the normalization constant calculated in appendix B and different from the usual \( 1/\sqrt{2\pi\sigma^2} \) because the Gaussian is truncated and the integration area of interest goes from 0 to \( a \). It is, in fact, useful in 1D structural crystallography to require that the area under this curve be equal to \( Z \), the atomic number. An auxiliary normalization function, \( G \), depending on \( a \) and \( \sigma \), can be conveniently introduced at this point. It is defined as

\[
G(a, \sigma) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \left[ \text{erf}\left( \frac{a}{2\sqrt{2}\sigma} \right) \right]^{-1},
\]

where \( \text{erf} \) is the error function. The required normalization constant \( K \) is given by the following relation:

\[
K = ZG(a, \sigma).
\]

(7)

The 1D Gaussian atom so defined depends on the two parameters \( x_0 \) and \( \sigma \), on the atomic number \( Z \) and on the unit cell length \( a \). \( x_0 \) is the atom’s position, provided by the specific structure under investigation. \( \sigma \) is a measure of the size of the atomic number \( Z \) and the thermal vibration of the atom nucleus, more specifically the vibration of its centre. We will address both aspects shortly, but first it is important to understand that each atom needs to sit comfortably in the unit cell. This means that all atoms’ density should be contained inside the cell. Such condition is certainly met if the tail’s truncation is carried at a certain integer quantity of standard deviations from the peak position. This integer \( m \) is given a default value of 5, but can, obviously, be increased or shortened. The following formula can be used to make sure that all the atoms are included:

\[
a = D_M + 2m\sigma_M, \quad D_M \equiv \max_{ij}(|x_i - x_j|), \quad \sigma_M \equiv \max_i(\sigma_i),
\]

(8)

where \( D_M \) is the maximum distance between atoms in one unit cell; if only one atom is contained in the cell, \( D_M = 0 \). The default choice is, as said before, \( a = D_M + 10\sigma_M \).

2.2. Atoms’ width and thermal vibration

It is reasonable for the atomic width to depend on the number of electrons \( Z \). The width is also augmented if the atomic nucleus vibrates due to thermal motion because this corresponds to the stationary electron cloud being smeared over a larger region. From a quantitative point of view the ‘cold’ atom (no thermal vibrations) can be associated with a positive parameter \( \sigma_0 \). In the present article (and in the accompanying R package, CRONE) \( \sigma_0 \) is proportional to the square root of \( Z \):

\[
\sigma_0 = k_\sigma \sqrt{Z},
\]

(9)

with \( k_\sigma = 0.05 \) by default. This ensures that all Gaussians in the unit cell have a shape that preserves individuality (each atom is separate from the others) because each atom’s width scales up as the square root of the atomic number, rather than linearly with it. Atomic vibrations can be expressed as a probability of the nucleus being found around its position of equilibrium. An appropriate choice of probability is the normal distribution with variance traditionally indicated by the parameter \( U \). The resulting electron density is well modelled as
convolution between the ‘cold’ atom density and the probability distribution of the nucleus’ centre. In appendix C it is shown that the resulting width $\sigma$ satisfies the following relation:

$$\sigma^2 = \sigma_0^2 + U,$$

and the normalized thermal Gaussian is given by the following expression:

$$\rho(x) = ZG(a, \sqrt{\sigma_0^2 + U})\exp\left[-\frac{(x - \mu)^2}{2(\sigma_0^2 + U)}\right],$$

where $\mu$ can be either $x_0 - a$, or $x_0$ or $x_0 + a$, as shown in equations (4) and (5), and $G$ is the function defined by formula (6). The net effect of thermal vibration is clearly to broaden the Gaussian shapes, but the specific choice of normalization constant fixes the area under the curve equal to $Z$, as before.

### 2.3. Gaussian molecule and structure

A 1D crystal structure is obtained when one or more Gaussian atoms populate the unit cell. In the present paper we adopt a point of view similar to the one adopted in current structural crystallography, which approximates the electron density of the crystal structure with the sum of the electron density of all atoms included. If $\rho$ is the crystal structure density and $\rho_1$, $\rho_2$, ..., $\rho_n$ are the densities of the atoms composing the structure, then we will always assume that

$$\rho(x) = \sum_{j=1}^{n} \rho_j(x - x_j), \quad 0 \leq x \leq a,$$

where $x_j$ is the central peak position of atom $j$. This is not strictly true as interactions among valence electrons modify the region between nuclei. But relation (12) is approximately valid in the context of kinematic diffraction, where x-ray scattering is mostly affected by core electrons and where the time scales involved are much longer than the nuclei and electrons’ motion.

In general the unit cell can include one or more atoms. In addition, copies of the same atom can be the result of crystallographic symmetry operations. In three dimensions there exist 230 different types of symmetries. In 1D the scenario is greatly simplified, but the main features of symmetry are still retained as there are just two types called $P1$ and $P\bar{1}$. The way symmetry is used in crystallography starts with the partition of unit cells in equal regions called asymmetric units. The atomic content in one asymmetric unit is equivalent to the that of any other asymmetric unit. Symmetry operations consider the coordinates of all atoms populating one asymmetric unit and produce equivalent atomic coordinates in all other asymmetric units. In 1D crystallography, with only two possible symmetries, the initial asymmetric unit can be chosen as either the whole unit cell, or as the segment between 0 and $a/2$. The two symmetries are quantitatively summarised in table 1. For example, suppose a $P1$ crystal structure with unit cell of length $a = 10$ has two atoms in an asymmetric unit at $x_1 = 2.5$ and $x_2 = 4.3$. Due to symmetry there will be two more atoms in the unit cell, one at $x_3 = 7.5$ and the other at $x_4 = 5.7$. $x_3$ is the symmetry equivalent of $x_1$ because $-2.5$ is outside the $[0, a]$ of the reference unit cell, but its translationally equivalent atom resides at $a - 2.5 = 10 - 2.5 = 7.5$. Similarly, $x_4$ is the symmetry equivalent of $x_2$.

There exist certain patterns in 2D that exhibit a kind of linear symmetry described by the so-called Frieze groups. As the functions treated in our models are strictly 1D and have radial symmetry around atoms’ peaks, such symmetry patterns are not applicable in the present context. The description of symmetry in 1D can also be enriched when quasi-crystals are taken into account, but this topic is not essential for the introductory level of this paper.
2.4. Atom’s occupancy

In ideal crystal structures all atoms have translational equivalents in all unit cells. In general, though, crystal’s imperfections mean that this is not always the case, and that some unit cells have missing atoms. This is taken care of theoretically with the idea of occupancy, a number between 0 and 1. An atom with occupancy equal to 1 will be found in the translationally equivalent positions across all unit cells. A smaller value of the occupancy means that the specific atom cannot be found in all unit cells. For instance, an atom with occupancy equal to 0.85 is found in roughly 85% of the crystal’s unit cells.

The concept of occupancy is also useful when listing atoms located in the so-called special position of the cell, which is $a/2$ (in three dimensions there are more than one special position). An atom located at $x = a/2$ has a symmetry equivalent at exactly the same position; this is, obviously, the same atom. Thus, both atoms can be listed as a structure’s content in the unit cell, as long as their occupancy is fixed at 0.5.

3. Linear molecules as models for 1D crystallography

1D crystallography can be demonstrated effectively using the structure of some existing linear molecules, as all their atoms lie approximately on a segment. Six molecules have been chosen to be used as examples in this paper. For each of them the interatomic distances have been searched on the Internet as published average bond lengths, but the coordinates have been placed arbitrarily in unit cells calculated using relation (8). Those cases in which one of the atoms acted as the inversion centre have been assigned space group $\overline{P1}$. To characterise atomic thermal vibrations in these molecules we have used the so-called $B$ factors, $B$, rather than the variances $U$, as customary in structural crystallography. They are defined as

$$B = 8\pi^2 U.$$  \hspace{1cm} (13)

The choice of $B$ factors is for us arbitrary. We have, thus, decided to fix the variance for hydrogen at $U_H = 80/(8\pi^2) = 10/\pi^2 \AA^2$, which corresponds to an approximate vibration of 1 Å around the equilibrium centre. For all other atoms it is reasonable to assume that such oscillations will have smaller amplitudes. This is heuristically justified by the energy equipartition theorem that distributes thermal energies in inverse proportion to atomic masses (atomic number, $Z$, in our case). All $B$ factors assigned to the linear molecules can, thus, be calculated using the following equation:

$$U_j = \frac{U_H}{Z_j} = \frac{10}{\pi^2 Z_j}, \quad B_j = 8\pi^2 U_j = \frac{80}{Z_j}. \hspace{1cm} (14)$$

So the $B$ factors will be $80 \ \AA^2$ for hydrogens, $80/6 \approx 13.333 \ \AA^2$ for carbons, $80/8 = 10 \ \AA^2$ for oxygens, etc.

**Table 1.** Possible asymmetric units and symmetry operations for the two symmetry systems in 1D crystallography.

| Symmetry type | Asymmetric units | Symmetry-related positions |
|--------------|------------------|---------------------------|
| $P1$         | $0 \leq x \leq a$ | $x$                       |
| $\overline{P1}$ | $0 \leq x \leq a/2$ | $x, -x$                  |
|              | $a/2 \leq x \leq a$ |                           |
The six linear molecules are presented in Table 2, with details of atomic positions and $B$ factors. Plots of the density corresponding to carbon dioxide and thiocyanate in the table are shown in Figure 2, superposed for comparison to the same structures with ‘cold’ atoms (all $B$ factors equal to 0).

### Table 2. 1D structures used as examples in this paper.

| Molecule Name          | Symmetry | Unit Cell (Å) | Atomic elements | Position in asymmetric unit (Å) | $B$ factor ($\text{Å}^2$) | Occupancy |
|------------------------|----------|---------------|-----------------|---------------------------------|--------------------------|-----------|
| Beryllium fluoride     | $P\bar{1}$ | 7.660         | F               | 2.290                           | 8.889                    | 1.0       |
|                        |          |               | Be              | 3.830                           | 20.000                   | 0.5       |
| Carbon dioxide         | $P\bar{1}$ | 6.066         | O               | 1.870                           | 10.000                   | 1.0       |
|                        |          |               | C               | 3.033                           | 13.333                   | 0.5       |
| Xenon difluoride       | $P\bar{1}$ | 11.674        | F               | 3.837                           | 8.889                    | 1.0       |
|                        |          |               | Xe              | 5.837                           | 1.481                    | 0.5       |
| Nitronium              | $P\bar{1}$ | 6.014         | O               | 1.857                           | 10.000                   | 1.0       |
|                        |          |               | N               | 3.007                           | 11.429                   | 0.5       |
| Cyanate                | $P\bar{1}$ | 3.833         | O               | 0.707                           | 10.000                   | 1.0       |
|                        |          |               | C               | 1.870                           | 13.333                   | 1.0       |
|                        |          |               | N               | 3.126                           | 11.429                   | 1.0       |
| Thiocyanate            | $P\bar{1}$ | 4.969         | S               | 1.000                           | 5.000                    | 1.0       |
|                        |          |               | C               | 2.813                           | 13.333                   | 1.0       |
|                        |          |               | N               | 3.969                           | 11.429                   | 1.0       |

The six linear molecules are presented in Table 2, with details of atomic positions and $B$ factors. Plots of the density corresponding to carbon dioxide and thiocyanate in the table are shown in Figure 2, superposed for comparison to the same structures with ‘cold’ atoms (all $B$ factors equal to 0).

### 4. Fourier Series and Structure Factors

The electron density for the full crystal, $\psi(x)$, is a periodical function. It can, therefore, be expressed as an infinite Fourier series:

$$\psi(x) = \frac{1}{a} \sum_{h=-\infty}^{+\infty} F_h \exp\left(-2\pi i h \frac{x}{a}\right)$$

(15)

where the complex expansion coefficient $F_h$ in crystallography is called the structure factor. As explained by the theory of the Fourier series, this coefficient is calculated as an integral over one periodic unit or, in this case,

$$F_h = \int_0^a \psi(x) \exp\left(2\pi i h \frac{x}{a}\right) dx \equiv \int_0^a \rho(x) \exp\left(2\pi i h \frac{x}{a}\right) dx,$$

(16)

where the last passage is justified by the fact that in the interval $[0, a]$ function $\psi(x)$ coincides with $\rho(x)$, as defined by equations (4) and (5). The expression of the Fourier coefficient when there is only one atom in the unit cell is important and interesting because all other expressions can be built starting from it. We are, thus, looking to calculate integral (16) in which $\rho(x)$ is the piecewise function (4) or (5), with $K$ given by expression (7). The full and detailed calculation is carried out in appendices D and E. It yields, for one atom:
The Fourier coefficient, or equivalent structure factor, is a complex quantity, as expected. Its amplitude, in this specific case of just one atom, is known as the scattering factor and indicates with the symbol $f$: 

$$ F_h = Z \exp\left(-2\pi^2 \sigma^2 \frac{h^2}{a^2}\right) \exp\left(2\pi i h \frac{x_0}{a}\right). $$

The Fourier coefficient, or equivalent structure factor, is a complex quantity, as expected. Its amplitude, in this specific case of just one atom, is known as the scattering factor and indicated with the symbol $f$: 

$$ f \equiv Z \exp\left(-2\pi^2 \sigma^2 \frac{h^2}{a^2}\right). $$

Its phase (or argument), contains $x_0$ and is therefore connected to the atomic position. In the expression for the scattering factor $\sigma$ is a generic quantity. For cold atoms $\sigma = \sigma_0$; for thermal atoms, as seen in formula (10), $\sigma^2 = \sigma_0^2 + U$. We have, accordingly,

$$ f_0 \equiv Z \exp\left(-2\pi^2 \sigma_0^2 \frac{h^2}{a^2}\right). $$

$$ f = Z \exp\left(-2\pi^2 (\sigma_0^2 + U) \frac{h^2}{a^2}\right) $$

$$ = Z \exp\left(-2\pi^2 \sigma_0^2 \frac{h^2}{a^2}\right) \exp\left(-2\pi^2 U \frac{h^2}{a^2}\right). $$

Figure 2. Density for two of the models reported in table 2, carbon dioxide and thiocyanate. The black curves correspond to all atoms having $B$ factors as assigned in the table. The red dashed curve correspond to the same structures with ‘cold’ atoms, i.e. where all atoms have $B$ factors equal to 0.
This last expression can be re-written using $f_0$, equation (19), and $B$, definition (13):

$$f = f_0 \exp \left( -\frac{B h^2}{4 a^2} \right).$$

(21)

The same result could have been obtained starting from the expression for the density of a thermal atom as a convolution because the Fourier integral of a convolution is the product of the individual Fourier integrals. The scattering factors of thermal atoms decay more rapidly than those of cold atoms because the exponential factor in equation (21) is a number smaller than 1 and decreases with increasing resolution (increasing values of the Miller index $h$). Examples of scattering curves plotted against increasing resolution, $h/a$, are shown in figure 3.

4.1. Structure factors of multiple atoms in the cell

When molecules, rather than one atom, are contained in the unit cell, the electron density is represented by equation (12). Integrals defining the structure factors are, in this case, equal to the sum of integrals for individual atoms. Therefore, the structure factor when more than one atom populates the unit cell is is also the sum of individual contributions:

$$F_h = \sum_{j=1}^{n} f_j \exp \left( 2\pi i h \frac{x_j}{a} \right).$$

(22)

where considering equations (18), (19) and (21)

$$f_j = Z_j \exp \left( -2\pi^2 \sigma_j^2 \frac{h^2}{a^2} \right) \exp \left( -\frac{B_j h^2}{4 a^2} \right).$$

(23)
From the above expressions it can clearly be seen that the structure factors of crystallographic structures composed of several atoms in the unit cell are the sum of individual complex numbers, each one of them having a length equal to the atom’s scattering factor $f_i$ and phase built out of the atom’s position. A representation of this sum in the Argand plane for the thiocyanate structure in table 2 is shown in figure 4.

5. Anomalous scattering

The reason why most quantitative structural crystallography can be introduced using only one dimension is related to the fact that structure factors are defined in the complex plane irrespective of whether the underlying molecules are built in one, two or three dimensions. Once a method to calculate structure factors is available, a whole array of topics mimicking those in 3D crystallography can be explained and easily illustrated. In [7], for instance, a simple 1D structure is used to explain fundamental crystallographic concepts like unit-cell refinement, the Patterson function, E-values and direct methods, and structure refinement. These same concepts can, of course, be explained also with the general framework adopted to build 1D crystallographic structures using truncated Gaussian functions. In this paper these arguments are skipped and the remainder will instead be devoted to anomalous scattering, as this is of relevance to several techniques used in modern structural crystallography. What follows is a very cursory presentation of anomalous scattering; a very accessible but more in-depth introduction to the same subject can be found on the website maintained by Ethan Merritt at the Biomolecular Structure Centre, University of Washington [10].

Atoms scatter anomalously when the energy, and thus the wavelength, used is close to their resonant values. This means that in general crystals do not scatter anomalously or, better, that the magnitude of anomalous diffraction is negligible when compared to ordinary scattering. At specific wavelengths, though, certain atoms in the crystal resonate and the resulting diffracted beam includes significant anomalous contributions. Their particular nature is exploited for phasing, i.e. to find phases of the scattering factors. Ordinary diffraction is

Figure 4. Structure factor $F_3$ (Miller index $h = 3$) for thiocyanate, as sum of contributions by individual atoms in the complex Argand plane. The structure factor is mostly contributed by sulphur, as is clear from the figure.
mathematically reflected in the usual factor appearing in equations (22) and (23). The expression for anomalous diffraction makes use of imaginary scattering factors in the following way:

\[ F_h = \sum_{j=1}^{n} (f_j^r + f_j^i + if_j^n) \exp \left( 2\pi i h \frac{\lambda_j}{\alpha} \right). \]  

(24)

The new scattering factor is equal to the ordinary, real, component plus a complex one, \( f_j^r + if_j^n \). Both the real and imaginary parts of this additional anomalous factor are virtually independent of resolution and only depend on wavelength. Values of \( f_j^r \) and \( f_j^n \) for all atomic species and for several wavelengths have been calculated theoretically and can be easily tabulated. The values used in the CRONE package have been extracted from Ethan Merritt’s site [10]. An example of \( f_j^r \) and \( f_j^n \) as functions of wavelength for iron atoms is depicted at figure 5. The top curve describes \( f_j^r \), while the bottom curve describes \( f_j^n \). In order to phase structures using anomalous scattering it is necessary to choose wavelengths close to the resonant energies (close to the dip of \( f_j^r \)) because in its neighbourhood the \( f_j^n \) anomalous contribution will be the largest, causing structure factors to change appreciably from their ordinary values.

The most important consequence that anomalous scattering has on structure factors is the breaking of the so-called Friedel’s law, according to which amplitudes of structure factors with opposite Miller indices are identical, and phases are opposite of each other. For structures not scattering anomalously in any appreciable way, such amplitudes will not be identical, due to experimental errors, but very similar. When some atoms in the structure scatter anomalously, the same related amplitudes are expected to be systematically different. It is on such difference that methods for structure factors’ phase estimate are based, as illustrated in the rest of this section.

Given a Miller index, \( h \), and its opposite, \(-h\), Friedel’s law states that the structure factor associated with \( h \) is the complex conjugate of the structure factor associated with \(-h\):
When contributions from the anomalous scattering become important, though, Friedel’s Law is not valid any longer and in general the following will be true: 

\[ |F_{-h}| = |F_h| \quad \text{and} \quad \varphi_{-h} = -\varphi_h. \]  

(25)

In this case the pair corresponding to opposite Miller indices, called a Bijvoet pair, is separated into a plus and a minus part, indicated as \( F^+ \) and \( F^- \). One of the methods used to phase crystal structures with anomalous phasing employs Bijvoet pairs to create a special density map called an anomalous Fourier difference map (or anomalous difference in short). This density map has amplitudes equal to the absolute square difference of each Bijvoet pair, and all phases equal 0. If the anomalous difference map is indicated as \( \rho_{\text{ano}}(x) \), then

\[ \rho_{\text{ano}}(x) = \frac{1}{a} \sum_{h=-\infty}^{+\infty} \Delta^2_{\text{ano}}(h) \exp(-2\pi i h x / a). \]  

(27)

where

\[ \Delta_{\text{ano}}(h) = ||F^+|-|F^-||. \]  

(28)

It can be shown that the anomalous difference map is equivalent to a corrupted Patterson map in which some of the prominent peaks correspond to distances between the anomalous scatterers, e.g. peaks between heavy atoms \([11]\). Once these distances have been measured, an initial, approximate model of the structure, made of anomalous scatterers, can be suggested. An updated Fourier transform calculated with the experimental structure factors’ amplitudes and phases from the initial model, should show improved positions for the anomalous scatterers, but also peaks corresponding to atoms not yet determined. The updated density map will be closer to the true density map if the selected peaks from the anomalous Fourier difference are the correct ones. The updated density can, in turn, be used to confirm or reject the positions of the anomalous scatterers and to select new peaks for the atoms missing from the initial approximate model of the structure. Old and new peaks will form a new structural model used to carry out another Fourier cycle (observed amplitudes and calculated phases). This sort of Fourier recycling provides atomic positions converging to the correct structure, if the initial anomalous scatterers have been properly selected. A specific example will make the procedure clearer.

Consider a P1 structure made of one carbon, one oxygen and two iron atoms at positions:

\[ x_1 = 2.0, \quad x_2 = 3.5, \quad x_3 = 10.0, \quad x_4 = 13.0 \]

in a unit cell of side \( a = 30 \text{ Å} \). All atoms are kept at a very low temperature, corresponding to a \( B \) factor equal to 0.5 Å\(^2\). All atoms in this structure are represented in figure 6. Structure factors have been computed for positive and negative Miller indices up to \( h = 80 \). Their amplitudes for the first five Bijvoet pairs and corresponding averages, taken to be equal to the observed amplitudes, are tabulated in table 3. The anomalous difference map, calculated with the anomalous differences just obtained, is displayed as thick curve in figure 7, jointly with the Patterson map (thin curve). Some of the highest peaks in the anomalous difference map (with the exclusion of the origin peak, which is always the highest peak) correspond to the interatomic distance between anomalous scatterers. Of the three highest peaks only the one with a height of 10.23 and located at \( x = 3.00 \) coincides with a prominent Patterson peak: it is, therefore, the most likely candidate to indicate the distance between the two iron atoms. We can place two iron atoms anywhere in the unit cell (because of the arbitrariness of the cell’s origin), as long as their distance is 3.00 Å, i.e. the position of the chosen peak in the anomalous difference map. The two iron atoms arranged in this way form an initial model for
the complete structure. To make the comparison with the known structure easier, the origin of
the unit cell has been located so to have the two iron atoms placed at 10 Å and 13 Å. Phases
calculated from this model are used jointly with the observed amplitudes to generate the first
approximate electron density map. This is shown in figure 8(a), overlapped to the final correct
structure. The approximate density has, obviously, high peaks at the iron position because the
chosen interatomic distance from the anomalous Fourier was the correct one. More peaks are
present in the density, though. Among the four highest ones there should be peaks
corresponding to the carbon and oxygen atoms missing from the initial model. If we select the
two peaks at position \( x = 2.01 \) (carbon atom) and \( x = 3.48 \) (oxygen atom), the updated
electron density appears as in figure 8(b), basically matching the density for the correct structure. If the other two peaks (at positions \( x = 19.53 \) and \( x = 21.00 \)) are used, instead, the updated density does not resemble the correct one (figure 8(c)). If the peak selected in the anomalous difference map had been the highest one, at position \( x = 1.50 \), the initial approximate electron density would have been the same as the one depicted at figure 9. This electron density does not resemble the one for the correct structure, with the exception of the peak in correspondence with the only correctly placed atom at \( x = 10 \).

**Figure 6.** Atoms corresponding to the structure used for the anomalous phasing example in section 5.

**Table 3.** Miller indices, Bijvoet pairs amplitudes, observed amplitudes and anomalous
differences for the first five structure factors computed in relation to the structure used in
the anomalous phasing example (section 5).

| \( h \) | \( F^+ \) | \( F^- \) | \( F_{\text{obs}} \) | \( \Delta_{\text{ano}}(h) \) |
|------|------|------|--------|----------|
| 1    | 37.65| 41.65| 39.65  | 4.00     |
| 2    | 24.74| 22.11| 23.43  | 2.62     |
| 3    | 35.14| 33.14| 34.14  | 2.00     |
| 4    | 20.65| 22.30| 21.48  | 1.65     |
| 5    | 9.90 | 9.93 | 9.92   | 0.03     |
6. The CRONE package

A package for the R statistical platform [5], named CRONE (CRystallography in ONE dimension) [4], has been developed to carry out all calculations explained in this paper, and other operations useful for structural crystallography in 1D. CRONE is freely downloadable directly from any CRAN repository [6] or from the GitHub site [4]. This package is comprised of functions, datasets and tables sufficient to create and execute numerous examples of crystallographic calculations using 1D Gaussian atoms. A few common crystallography operations, explained and demonstrated with the help of CRONE, are included in appendix F. A variety of additional examples can be found in the package’s help pages as well as in an R-specific descriptive document available within the package and known as vignettes.

7. Conclusions

Amongst the most interesting and attractive aspects of structural crystallography, the mathematics of Fourier series and Fourier transforms play a dominant role in the community of physicists. Many of the most important features in this area can be explained and illustrated analytically making use of Gaussian functions. These concepts are very familiar to physics undergraduates and graduates, as they are taught in several modules. The 1D introduction to structural crystallography described in the present paper provides, therefore, an excellent starting point to captivate students’ attention and to steer their interest in this important topic across both the physical and biological sciences. Such an introduction makes use of truncated and periodical 1D Gaussians to represent atoms in a crystal lattice, both visually and quantitatively. Furthermore, the close parallel between the complex Fourier coefficients for both 1D and 3D structures facilitates the transition to modern crystallographic jargon. Students that learn 1D crystallography as explained in this paper will have an immediate grasp of both the
Figure 8. Approximate electron densities (red, full curves) compared with the final correct density (black, broken curve), during the process of phase determination. The first approximate density is shown in part (a) of the figure. The correct position of the two iron atoms is reflected in the height of the two largest peaks. Four more peaks are possible candidates for the carbon and oxygen in the final structure. The correct ones are indicated with C and O and their inclusion in the updated structural model yields the density in part (b) of the figure. Selection of the other two peaks, instead, leads to the incorrect structure, as shown in part (c).
qualitative and quantitative aspects of the discipline. The paper also introduces and demonstrates a new package in the R environment for statistics, CRONE, that makes any calculation within 1D crystallography possible and easy. Its use is a valid aid in classroom teaching and enables the generation of innumerable examples and exercises at various levels.

Acknowledgements and dedication

We would like to thank Professor Ethan Merritt for reading the manuscript and for providing thorough and extensive comments. We would also like to thank the referees whose observations, comments and critiques have helped to improve this article in its final version.

One of the authors, JF, is sponsored by grant number 653706 iNEXT-WP24 of the European Union. We would also like to thank the Summer Placement Students programme at Diamond for providing both money and resources for one of the authors (ES).

This paper is specially dedicated to the memory of the late Alessandro Nero, passionate about rowing and enthusiastic about physics and engineering, and whose premature and tragic death has left family and friends with heavy hearts.

Appendix A. Infinite crystal structures with full Gaussians: some results

Let us consider the simple case of a 1D structure made out of the regular repetition of one atom with atomic number represented by the integer $Z$, and where the spacing between atoms is $a$, equal to the cell length. Without losing generality we can place atomic centres at the lattice points, $-2a, -a, 0, +a, +2a, ...$. The structure is represented by the expression

$$\psi(x) = \frac{Z}{\sqrt{2\pi\sigma^2}} \sum_{u=-\infty}^{+\infty} \exp \left[ -\frac{(x - ua)^2}{2\sigma^2} \right]$$  \hspace{1cm} (A.1)

The following points help define the character of this function.
(i) $\psi(x)$ is periodical, with period $a$.
We can see this by computing $\psi(x + ka)$ and proving that this is equal to $\psi(x)$, where $k$ is any integer:

$$\psi(x + ka) = \frac{Z}{\sqrt{2\pi \sigma^2}} \sum_{u=-\infty}^{+\infty} \exp \left[ -\frac{(x + ka - ua)^2}{2\sigma^2} \right]$$

$$= \frac{Z}{\sqrt{2\pi \sigma^2}} \sum_{u=-\infty}^{+\infty} \exp \left\{ -\frac{[x - (u - k)a]^2}{2\sigma^2} \right\}.$$  

The quantity $u - k$ is an integer running from $-\infty$ to $+\infty$; therefore, calling $u' \equiv u - k$ we have:

$$\psi(x + ka) = \frac{Z}{\sqrt{2\pi \sigma^2}} \sum_{u'=-\infty}^{+\infty} \exp \left[ -\frac{(x - u'a)^2}{2\sigma^2} \right] \equiv \psi(x). \quad (A.2)$$

(ii) The value of $\psi(ka)$ is analytically given by an auxiliary Jacobi special function.

The value of $\psi(x)$ at any lattice point $ka$, where $k$ is an integer, can be approximated using the Jacobi auxiliary theta function $\theta_3$, which is a complex function defined in the context of elliptic integrals. In detail:

$$\psi(ka) = \psi(0 + ka) = \psi(0) \quad (A.3)$$

because the function is periodic. Therefore,

$$\psi(ka) = \frac{Z}{\sqrt{2\pi \sigma^2}} \sum_{u'=-\infty}^{+\infty} \exp \left( -\frac{a^2}{2\sigma^2} u'^2 \right). \quad (A.4)$$

The following property applies to infinite sums of Gaussian-like terms:

$$\sum_{u'=-\infty}^{+\infty} \exp (-Ku'^2) = \theta_3(0, \exp(-K)), \quad (A.5)$$

where $\theta_3(z, q)$ is the Jacobi third auxiliary function [9] with nome equal to $q$. Using property (A.5) in equation (A.4) brings us to the following result:

$$\psi(ka) = \frac{Z}{\sqrt{2\pi \sigma^2}} \theta_3 \left( 0, \exp \left( -\frac{a^2}{2\sigma^2} \right) \right). \quad (A.6)$$

It might be possible to compute $\psi(x)$ for values of $x$ different from $ka$, but such calculation necessarily involves a deeper and thorough use of Jacobi theta functions, a topic beyond the scope and aim of this paper.

(iii) $\psi(x)$ is normalised in the interval corresponding to the unit cell. The area under the curve represented by $\psi(x)$ is, obviously, infinite because such an area is the infinite summation of the finite and positive area in any interval corresponding to the unit cell. It is, though, important to verify that the area corresponding to one unit cell is finite and to calculate its value. The area to be computed is:

$$\text{area} = \int_{-a/2}^{a/2} \psi(x) \, dx$$

$$= \frac{1}{\sqrt{2\pi \sigma^2}} \sum_{u=-\infty}^{+\infty} \int_{-a/2}^{+a/2} \exp \left[ -\frac{(x - ua)^2}{2\sigma^2} \right] \, dx. \quad (A.7)$$

Using the substitution $(x - ua)/\left(\sqrt{2} \sigma\right) \equiv t$, expression (A.7) becomes
area = \frac{1}{\sqrt{\pi}} \sum_{u=\infty}^{+\infty} \left\{ \frac{(-u+1/2)(a/\sqrt{2}\sigma)}{\exp(-u^2)} \right\} \exp(-t^2) \, dt \\
\quad = \frac{1}{2} \sum_{u=\infty}^{+\infty} \left\{ \text{erf}\left[\left(-u + \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right] - \text{erf}\left[\left(-u - \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right]\right\}, \tag{A.8}

where \text{erf} is the error function defined as:

\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \exp(-t^2) \, dt. \tag{A.9}

The infinite summation can be re-written as the limit of a finite summation:

area = \frac{1}{2} \lim_{n \to +\infty} \sum_{u=-n}^{+n} \left\{ \text{erf}\left[\left(-n + \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right] - \text{erf}\left[\left(-n - \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right]\right\}, \tag{A.10}

After long, tedious, but not difficult simplifications, the expression for the area becomes:

area = \frac{1}{2} \lim_{n \to +\infty} \left\{ \text{erf}\left[\left(-n + \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right] - \text{erf}\left[\left(-n - \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right]\right\}, \tag{A.11}

and using the parity of the error function,

area = \frac{1}{2} \lim_{n \to +\infty} 2 \text{erf}\left[\left(n + \frac{1}{2}\right) \frac{a}{\sqrt{2}\sigma}\right] = 1. \tag{A.12}

Thus, the area of the infinite sum of Gaussians within a unit cell is finite and equal to 1. From a qualitative point of view this is sensible because, although there are infinite contributions that raise the curve upwards in a unit cell, such contributions are infinitesimally smaller and smaller, so that the final infinite sum converges.

**Appendix B. Normalization constant for the truncated Gaussian**

Given the following function,

\[ \rho(x) = K \exp\left[-\frac{(x-x_0)^2}{2\sigma^2}\right] \tag{B.1} \]

we want to compute that value of the constant \(K\) that makes the area under \(\rho(x)\), between 0 and \(a\), equal to \(Z\), the atomic number. Calculations are equivalent if the integral is computed between \(-a/2\) and \(+a/2\) for the Gaussian centred at the origin. This yields the following expression:

\[ K \int_{-a/2}^{+a/2} \exp[-x^2/(2\sigma^2)] \, dx = Z \tag{B.2} \]

or, equivalently,

\[ 2K \int_{0}^{+a/2} \exp[-x^2/(2\sigma^2)] \, dx = Z. \tag{B.3} \]

If \(x\) is replaced with the new integration variable \(t \equiv x/\sqrt{2\sigma^2}\), the expression becomes

\[ 2\sqrt{2\sigma^2} K \int_{0}^{a/(2\sqrt{2}\sigma)} \exp(-t^2) \, dt = Z. \tag{B.4} \]
Now, the error function, \( \text{erf}(x) \), is defined as
\[
\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) \, dt.
\] (B.5)

Thus, equation (B.4) can be re-written as
\[
\sqrt{2\pi\sigma^2} K \text{erf}\left(\frac{a}{2\sqrt{2}\sigma}\right) = Z,
\] (B.6)

which yields
\[
K = Z \frac{1}{\sqrt{2\pi\sigma^2}} \left[ \text{erf}\left(\frac{a}{2\sqrt{2}\sigma}\right) \right]^{-1}.
\] (B.7)

Considering definition (6) for function \( G \), \( K \) is therefore given by equation (7).

**Appendix C. Truncated Gaussian for a thermal atom**

The general form of a Gaussian for the ‘cold’ atom in equations (4) and (5) is
\[
\rho_0(x) = ZG(a, \sigma_0) \exp\left\{ -\frac{(x - \mu)^2}{2\sigma_0^2} \right\}, \quad 0 \leq x \leq a,
\] (C.1)

where \( \mu \) is either \( x_0 - a \), or \( x_0 \), or \( x_0 + a \). When the atomic nucleus vibrates due to an increase of energy (thermal vibration) the whole electron cloud is dragged along in this motion and the net macroscopic effect is a smearing of the cloud over a larger region. Mathematically the effect can be quantified through a convolution between density (C.1) and a normalized Gaussian probability with mean 0 and variance \( U \):
\[
p(x) = \frac{1}{\sqrt{2\pi U}} \exp\left\{ -\frac{x^2}{2U} \right\}, \quad -\infty < x < +\infty.
\] (C.2)

The Gaussian density of the thermal atom is, therefore, defined as
\[
\rho(x) = \int_{-\infty}^{+\infty} p(x - x') \rho_0(x') \, dx', \quad 0 \leq x \leq a.
\] (C.3)

The calculation is easier to illustrate if only one period of the atom, in the interval \([0, a]\), is used for \( \rho_0 \), while the probability is defined in the whole \((-\infty, +\infty)\) interval. The full set of periodical atoms can be used to derive the final result, albeit with some pedagogical difficulty. If expressions (C.1) and (C.2) are inserted into expression (C.3) for the convolution, the result is:
\[
\rho(x) = \frac{ZG(a, \sigma_0)}{\sqrt{2\pi U}} \int_{-\infty}^{+\infty} \exp\left\{ -\left[ \frac{(x' - x)^2}{2U} + \frac{(x' - \mu)^2}{2\sigma_0^2} \right] \right\} \, dx'.
\] (C.4)

Most of the algebraic manipulation needed now concerns the argument in square brackets. This should be transformed into the sum of two new quantities, one of which, including \((x - \mu)^2\), can be taken out of the integral. In fact, after a long, tedious but not difficult, series of algebraic passages, the argument inside the square brackets becomes
\[
\left( x' - \frac{x\sigma_0 + \mu U}{\sigma_0^2 + U} \right) \left\{ 2\frac{U\sigma_0}{\sigma_0^2 + U} + \frac{(x - \mu)^2}{2(\sigma_0^2 + U)} \right\}.
\] (C.5)

The first term contains \( x' \), the integration variable. After integration, this term gives rise to a number that multiplies all factors outside the integral in (C.4). The second term is untouched.
so that expression (C.4) becomes

$$\rho(x) = K' \exp \left[ -\frac{(x - \mu)^2}{2(\sigma_0^2 + U)} \right]$$  \hspace{1cm} (C.6)

$K'$ will not, in general, yield a properly normalized function. We have the freedom, though, to adopt a different constant in the definition of $\rho$. We can, in fact, use for $\rho$ the same normalization used in equation (7) because this guarantees the area under the curve to be $Z$.

The functional part of $\rho$ is identical to that of $\rho_0$ if $\sigma_0$ is replaced by $\sqrt{U + \sigma_0^2}$. Thus, the final, properly normalized, Gaussian density for thermal atoms is

$$\rho(x) = ZG(a, \sqrt{\sigma_0^2 + U}) \exp \left[ -\frac{(x - \mu)^2}{2(\sigma_0^2 + U)} \right], \quad 0 \leq x \leq a.$$  \hspace{1cm} (C.7)

For $U = 0$ this function correctly returns $\rho_0$.

**Appendix D. Fourier coefficient of a truncated Gaussian**

We would like to calculate

$$F_k = \int_{0}^{a} \rho(x) \exp\left(2\pi i h \frac{x}{a}\right) dx,$$  \hspace{1cm} (D.1)

where $\rho(x)$ is given by the piecewise functions (4) or (5), and where $K$ in such expressions is defined in equation (7). Even though there are four different analytic expressions in the definition of $\rho$, there is no need to carry out four separate integrations. Each branch of the piecewise density include terms of the following form:

$$K \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right].$$  \hspace{1cm} (D.2)

Furthermore, the important feature of the integration interval used to calculate the Fourier coefficients is its width, $a$. This means that such interval does not have to be centred necessarily on $x = a/2$. It is, in fact, convenient to centre it on $x = \mu$. The integral needed to compute the Fourier coefficients, thus, has the form

$$F_k = K \int_{\mu-a/2}^{\mu+a/2} \exp\left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] \exp\left(2\pi i h \frac{x}{a}\right) dx,$$  \hspace{1cm} (D.3)

with $\mu$ any of the values $x_0$, $x_0 + a$ or $x_0 - a$. To solve integral (D.3) the new variable $y \equiv x - \mu$ is used. The substitution yields

$$F_k = K \exp\left(2\pi i h \frac{\mu}{a}\right) \int_{-a/2}^{+a/2} \exp\left(-\frac{y^2}{2\sigma^2}\right) \exp\left(2\pi i h \frac{y}{a}\right) dy.$$  \hspace{1cm} (D.4)

To avoid using integration techniques in the complex plane and evaluating the error function for complex numbers it is, at this point, convenient to separate (D.4) into real and imaginary parts. The integral of the imaginary part involves integration within a symmetric interval of an odd function (product of a sine and a Gaussian); this integral is, therefore, zero. We are, thus, left with the following:
\[ F_\nu = K \exp \left( \frac{2\pi i h \mu}{a} \right) \int_{-a/2}^{+a/2} \cos \left( \frac{2\pi h y}{a} \right) \exp \left( -\frac{y^2}{2\sigma^2} \right) dy. \quad (D.5) \]

The integral in equation (D.5) is computed in appendix E and is equal to
\[ \sqrt{2\pi\sigma^2} \operatorname{erf} \left( \frac{a}{2\sqrt{2\sigma^2}} \right) \exp \left( -2\pi^2\sigma^2 \frac{h^2}{a^2} \right) \equiv \frac{1}{G(a, \sigma)} \exp \left( -2\pi^2\sigma^2 \frac{h^2}{a^2} \right). \quad (D.6) \]

So, given that \( K = ZG(a, \sigma) \), the Fourier coefficient becomes
\[ F_\nu = ZG(a, \sigma) \exp \left( \frac{2\pi i h \mu}{a} \right) \frac{1}{G(a, \sigma)} \exp \left( -2\pi^2\sigma^2 \frac{h^2}{a^2} \right) \]
\[ = Z \exp \left( -2\pi^2\sigma^2 \frac{h^2}{a^2} \right) \exp \left( 2\pi i h \frac{\mu}{a} \right). \quad (D.7) \]

When \( \mu \) is replaced with either of \( x_0, x_0 + a \) or \( x_0 - a \) equation (D.7) coincides with equation (17).

**Appendix E. Integrals of trigonometric and Gaussian functions**

The purpose of the present appendix is to calculate an integral of the form
\[ I(h) \equiv \int_{-a/2}^{+a/2} \cos \left( \frac{2\pi h y}{a} \right) \exp(-by^2) dy. \quad (E.1) \]

This is the integral contained in expression (D.5), where
\[ b = \frac{1}{2\sigma^2}. \quad (E.2) \]

In order to calculate a derivative inside the integral, let us assume \( h \) to be a continuous, real variable. Equation (E.2) can be seen as a function of \( h \), and its derivative calculated with respect to \( h \),
\[ \frac{dI(h)}{dh} = -\int_{-a/2}^{+a/2} \frac{2\pi}{a} \sin \left( \frac{2\pi h y}{a} \right) \exp(-by^2) dy. \quad (E.3) \]

After an integration by parts, the expression becomes
\[ \frac{dI(h)}{dh} = -\left[ -\frac{\pi}{ab} \exp(-by^2) \sin \left( \frac{2\pi h y}{a} \right) \right]_{-a/2}^{+a/2} \]
\[ + \int_{-a/2}^{+a/2} \frac{2\pi h}{ab} \cos \left( \frac{2\pi h y}{a} \right) \exp(-by^2) dy. \quad (E.4) \]

The integral in this expression is \( I(h) \) times a constant. The calculation leads to the following first-order differential equation:
\[ \frac{dI(h)}{dh} + \frac{2\pi^2}{ab^2} h I(h) = \frac{2\pi}{ab} \exp \left( -\frac{a^2 b}{4} \right) \sin(\pi h). \quad (E.6) \]

For values of \( h \) in an infinitesimal neighbourhood of any integer, the right-hand side of the equation is very close to zero and assumes the following form:
\[ \frac{dI(h)}{dh} + \left( \frac{2\pi^2}{a^2 b} \right) I(h) = 0, \quad (E.7) \]

whose solution is

\[ I(h) = C \exp \left( -\frac{\pi^2 h^2}{b a^2} \right). \quad (E.8) \]

The integration constant \( C \) can be found when \( h = 0 \) is used in (E.8), using also definition (E.1):

\[ C = I(0) = \int_{-a/2}^{+a/2} \exp(-by^2) \, dy = \frac{\pi}{\sqrt{b}} \exp \left( -\frac{a \sqrt{b}}{2} \right) \quad (E.9) \]

Therefore, the integral (E.1) is

\[ I(h) = \frac{\pi}{\sqrt{b}} \exp \left( -\frac{a \sqrt{b}}{2} \right) \exp \left( -\frac{\pi^2 h^2}{b a^2} \right). \quad (E.10) \]

If we use the value in (E.2) in place of \( b \) we obtain the result for the integral in expression (D.5), yielding the final result (D.6).

**Appendix F. Demonstration of the CRONE package**

Throughout this section knowledge of R is assumed. The reader wishing to learn more about this popular computational platform can find a substantial amount of introductory material on the R official website [5]. More tutorials can also be found in the website of one of the authors [12].

**F.1. Calculation and display of a 1D structure**

The six 1D structures available in CRONE, or any other existing or arbitrary linear molecules, can be calculated and displayed directly in direct space, or by making the transition via reciprocal space. The starting points for both calculations are:

- Length of unit cell, \( a \) (in angstroms);
- Space group symbol, \( SG \) (‘P1’ or ‘P-1’);
- Vector \( \mathbf{x} \) of atomic positions in the unit cell;
- Vector \( \mathbf{Z} \) of atomic numbers corresponding to all atoms in the cell;
- Vector \( \mathbf{B} \) of \( B \) factors for all atoms in the cell;
- Vector \( \mathbf{occ} \) of atomic occupancies

**F.1.1. Data preparation.** These quantities can be created \textit{ad hoc} in a completely arbitrary way, or loaded in the working directory from one of the six structures included in the package. Whichever way data are created, they can be arranged in the standard format used throughout the package, a named list, known as \texttt{sdata}, with its six elements equal to those displayed above. Here we will assume to be working with thiocyanate (see table 2). When data from this structure are loaded in the R working space, the following \texttt{sdata} object is created:

\[
> \texttt{sdata} \leftarrow \texttt{load_structure(‘thiocyanate’)}
\]
This 1D structure has three atoms in a unit cell of length $a = 4.969 \, \text{Å}$. They are sulphur (S, $Z = 16$) at position $x = 1.000$, carbon (C, $Z = 6$) at position $x = 2.813$ and nitrogen (N, $Z = 7$) at position $x = 3.969$. Their $B$ factors are, respectively, 5.000, 13.333 and 11.429. All have occupancies 1. The symmetry for this structure is P1, which means there is no symmetry.

**F.1.2. Calculation and display in direct space.** The CRONE function to calculate density corresponding to 1D structures is `structure_gauss`. The calculation is carried out at every point of a regular grid. The result is a list, `rho`, with two vectors, the regular grid $x$ and the density values $\rho$. Both vectors are jointly used to display the structure (see figure F1(a)):

```r
> print(sdata)
$sdata
[1] 4.969
SSG
[1] 'P1'
$x0
[1] 1.000 2.813 3.969
SZ
[1] 16 6 7
$B
[1] 5.000 13.333 11.429
$occ
[1] 1 1 1
```

This 1D structure has three atoms in a unit cell of length $a = 4.969 \, \text{Å}$. They are sulphur (S, $Z = 16$) at position $x = 1.000$, carbon (C, $Z = 6$) at position $x = 2.813$ and nitrogen (N, $Z = 7$) at position $x = 3.969$. Their $B$ factors are, respectively, 5.000, 13.333 and 11.429. All have occupancies 1. The symmetry for this structure is P1, which means there is no symmetry.

**Figure F1.** 1D density inside (a) one and (b) multiple unit cells, for thiocyanate. The three peaks correspond to the central position of sulphur, carbon and nitrogen.
F.1.3. Calculation and display in reciprocal space. The density displayed at figure F1(a) can also be calculated as Fourier series if the structure factors (i.e. the Fourier coefficients) are computed first. The function in charge of structure factors calculation in CRONE is called \texttt{strufac}. We will need to decide in advance how many Fourier components are required. The remaining part of the input is equivalent to the one used in the calculation for direct space, that is the information on the structure, \texttt{sdata}. The final result of the calculation is a named list with two elements, called \texttt{Fmod} and \texttt{Fpha}, which are vectors corresponding to the structure factors’ amplitudes and phases, respectively. Their location in both vectors corresponds to the location in vector \texttt{hidx} of input indices (Miller indices). Due to the way 1D truncated Gaussians are normalised, the fundamental component (\texttt{hidx}=0) of the Fourier series is always going to be equal to the sum of atomic numbers of all atoms in the unit cell (29 for this specific structure):

\begin{verbatim}
> print(names(ftmp))
[1] 'Fmod' 'Fpha'
> print(ftmp$Fmod [1:4])
[1] 29.000000 6.289007 12.500798 9.582935
> print(ftmp$Fpha [1:4])
[1] 0.00000 75.66567 147.71590 -149.88296
> sum(sdata$Z)
[1] 29
\end{verbatim}

The structure factors can be used to calculate a Fourier summation for all points of a grid \(x\) representing the unit cell. The function of CRONE available to carry out this so-called \textit{Fourier synthesis} is \texttt{fousynth}:

\begin{verbatim}
> N <- 1000  # Number of grid points in the unit cell
> rtmp <- fousynth(sdata$a, ftmp$Fmod, ftmp$Fpha, hidx, N)
> x <- rtmp$x
> rho <- rtmp$rr
\end{verbatim}

The vectors \(x\) and \(rho\) obtained are the same as those giving the density in figure F1(a).

F.2. Approximate phases and peak search

The main goal of structural crystallography is to find the positional coordinates of all atoms in a given structure. The quantity that can be calculated using experimental data is the electron density. Atomic coordinates correspond to the peaks in the electron density. Positional precision and accuracy depend, of course, on the calculated electron density quality, this, in turn, being dependent on the precision and accuracy of all estimated structure factors phases. Amplitudes and phases from the structure factors of thiocyanate (see section F.1.3) can be computed using \texttt{strufac}; we call the two vectors created as \texttt{Fn} and \texttt{Fph}, respectively. As phases are not obtained experimentally, but are estimated with some degree of uncertainty, we can simulate approximate ones, \texttt{Fapp}, starting from their correct value, \texttt{Fph} and adding a small random error:
Amplitudes and phases of correct coefficients

\[ F_m \leftarrow \text{ftmp}$Fmod \]
\[ F_{ph} \leftarrow \text{ftmp}$Fpha \]

Simulation of approximate phases

\[ F_{app}[1] \leftarrow F_{ph}[1] \# \text{The } h=0 \text{ component is untouched} \]
\[ \text{set.seed}(9195) \# \text{Fix simulated random numbers} \]
\[ F_{app}[2:21] \leftarrow F_{ph}[2:21] + \text{rnorm}(19, \text{mean}=0, \text{sd}=0.4) \]

Approximate density

\[ \rho_{app} \leftarrow \text{fousynth}(sdata$a, F_m, F_{app}, hidx, N) \]

The Fourier synthesis obtained using experimental amplitudes $F_m$ and approximated phases $F_{app}$ will be in general slightly different from the correct density (see figure F2).

Next, we need to find peak coordinates in the density obtained. In CRONE this is achieved using the function \texttt{local_maxima}. The peaks found in the correct density are shown in figure F3 and have values 0.999, 2.847 and 3.945, very close to the values in table 2. The second peak is not found in the approximate density. Maxima are displayed only for the first and third atom. They occur at 0.889 and 3.955. Starting with these values and cycling between structure factors and density, a convergence to correct values is eventually reached.

F.3. Errors in structure factors

Amplitudes from calculated structure factors are never exactly met in experiments. Experimental quantities are different from their theoretical values for many reasons. In CRONE function \texttt{sfobs} simulates errors of two types for the structure factor amplitudes,
Poissonian counting errors

Positional atomic coordinate errors

**Poissonian counting errors** are due to the fluctuations in the number of photons scattered along a given direction (thus corresponding to certain Miller indices). The net effect of these fluctuations is to alter the structure factors theoretical amplitude, producing slightly smaller or larger values, depending on the amount of photons. Fluctuations in the value of $s_{f_{\text{obs}}}$ are simulated as random generations of poissonian variates with mean equal to the theoretical structure factor’s amplitude. The final value is calculated as average of all simulated amplitudes and the error as the corresponding standard deviation. The number of simulated deviates is controlled by the parameter $n_{\text{trialP}}$ (default 100). As an example let us compare theoretical and simulated amplitudes for Miller indices $h_{\text{idx}}=1,...,5$, for thiocyanate: the **positional atomic coordinate errors** simply simulates errors of the atomic centres around their correct values as Gaussian errors with mean zero and standard deviation controlled by the parameter $v_{x0err}$. The underlying structure is only approximately equal to the correct structure and, accordingly, the corresponding structure factors’ amplitudes are slightly different from their calculated values. The example shown in the following simulate errors in the atomic coordinates for thiocyanate; the standard deviation used for the simulation is $v_{x0err}=0.4$ Å. Poisson counting errors have been cancelled by using $n_{\text{trialP}}=0$: the structure corresponding to structure factors calculated with errors just in the atomic coordinates is shown at figure F4. Coordinate errors are visible as peak shifts.

![Figure F4. Discrepancy between correct density for thiocyanate and density calculated with modified structure factors’ amplitudes affected by errors in the coordinates.](image)

```r
> G <- sfobs(hidx, sdata, ntrialP=10)
> # Calculated amplitudes
> Fm[2 : 6]
[1] 6.289007 12.500798 9.582935 3.117837 2.375137
> # Simulated amplitudes with errors
> Fobs <- G*S
> Fobs[2:6]
[1] 6.241704 12.768149 9.598217 3.036468 2.62306
> # Errors
> sFobs <- S*G > sFobs[2 : 6]
```

---

27
F4. The breaking of Friedel’s law

In section 5 it was mentioned that when a structure contains atoms which scatter anomalously in a significant way, then structure factors do not obey Friedel’s law anymore. This can be nicely demonstrated using the `strufac` function. In the set of 1D molecules available within CRONE there are no structures including atoms scattering anomalously in any appreciable way. We can artificially create such a structure by replacing the carbon of thiocyanate with an iron atom:

```r
> # Iron has Z=26
> sdata$Z[2] <- 26
> # Reduce thermal vibration
> sdata$B[2] <- 3.0
> print (sdata)
$s0
[1] 1.000 2.813 3.969
$Z
[1] 16 26 7
$B
[1] 5.000 3.000 11.429
$occ
[1] 1 1 1
```

Next, we can find out at which wavelength the anomalous scattering is expected to produce the largest differences in structure factors’ amplitudes, using function `fluorescent_scan`:

```r
> print (fluorescent_scan('Fe')$lambda)
[1] 1.742797
```

For wavelength close to this value the effect of anomalous scattering will be relatively large. We will, now, calculate structure factors where the anomalous scattering is switched off, so that the corresponding Friedel’s pairs are exactly equal:

```r
> # Positive and negative h to show Friedel’s pairs
```
Continued.

> hidx <- 10:10
> 
> # Exact structure factors: no anomalous effect
> ftmp1 <- strufac(hidx, sdata, lbda=1.0, anoflag=FALSE) > modsF1 <- ftmp1$Fmod
> 
> # Friedel’s pair h = -1, h=1
> print(paste(modsF1[10], modsF1[12]))
> [1] "15.682277494909 15.682277494909"
> 
> # Friedel’s pair h = -2, h=2
> print(paste(modsF1[9], modsF1[13]))
> [1] "18.0288655269983 18.0288655269983"

When the anomalous scattering is not switched off, but structure factors are calculated at a wavelength far from 1.74 Å, Friedel’s pairs will not be exactly equal, but will be approximately close. In fact, atoms always scatter anomalously, but the magnitude of such scattering is very small except at specific resonant wavelengths:

> ftmp2 <- strufac(hidx, sdata, lbda=1.0, anoflag=TRUE)
> 
> # Friedel’s pair h = -1, h=1
> print(paste(modsF2[10], modsF2[12]))
> [1] "15.8261334218521 15.6615216563831"
> 
> # Friedel’s pair h = -2, h=2
> print(paste(modsF2[9], modsF2[13]))
> [1] "17.8722603191626 18.3860646475774"

When structure factors’ calculation is carried out at 1.74 Å, the iron atom will resonate anomalously and increase the difference between Friedel’s pairs:

> ftmp3 <- strufac(hidx, sdata, lbda=1.74, anoflag=TRUE)
> 
> # Friedel’s pair h = -1, h=1
> print(paste(modsF3[10], modsF3[12]))
> [1] "14.1965995692363 13.7285463084373"
> 
> # Friedel’s pair h = -2, h=2
> print(paste(modsF3[9], modsF3[13]))
> [1] "16.1298629197846 17.5396455968006"

For real structure factors derived from collected data, the differences in amplitude for Friedel’s pairs can be concealed by experimental errors. In CRONE this effect can be explored with the help of function sfobs.
ORCID iDs

James Foadi © https://orcid.org/0000-0001-7681-1131

References

[1] Bragg W L 1913 Proc. Cambridge Phil Soc. 16 43–57
[2] Winn M D et al 2011 Acta Cryst. D 67 235–42
[3] Adams P D et al 2010 Acta Cryst. D 66 213–21
[4] CRONE 2017 https://github.com/jfoadi/crone
[5] R Core Team 2017 https://R-project.org/
[6] Comprehensive R Archive Network 2017 https://cran.r-project.org/
[7] Pinkerton A A 2015 J. Appl. Cryst. 48 901–5
[8] Sout G H and Jensen L H 1968 X-ray Structure Determination: A Practical Guide (New York: McMillan)
[9] Abramowitz M and Stegun I 1964 Handbook of Mathematical Functions (New York: Dover Publications)
[10] Merritt E A 1996–2012 X-ray anomalous scattering http://skuld.bmsc.washington.edu/scatter/AS_index.html
[11] Woolfson M M 1997 An Introduction to X-ray Crystallography (Cambridge: Cambridge University Press)
[12] A relatively-short introduction to R 2017 http://jfoadi.me.uk/statistics.html