X-Ray Diffraction

Yue Wu

Yusuf Hamied Department of Chemistry
Lensfield Road,
Cambridge, CB2 1EW

yw628@cam.ac.uk

Acknowledgements

Contents

Figure 1.1: A magnified image of the diffraction grating is formed using a convex lens to focus the diffraction pattern.

1 Diffraction

It might be familiar from physics that if you shine a beam of light on a diffraction grating, then a diffraction pattern will emerge. If you put a lens after the diffracted beam, they may be brought to a focus, and a magnified image of the diffraction grating will be seen, as shown in the figure.

Atoms in a crystal are, in some sense, very fine gratings. Can we use a similar method to image the atoms in a crystal directly? However, to form a diffraction pattern, we need the wavelength of the light to be in roughly the same length scale as the pattern we are trying to image. The atoms are on the order of Angstroms (10^{-10} m), therefore, we need to use X-rays for diffraction at atomic length scales. However, there is no practical conventional lens for X-rays, because lenses are made of atoms too, so the imaging process does not work. What we have to do is to measure the intensity of the diffracted beams at different directions, and infer the atomic structure mathematically.

This is, however, a hard thing to do. We will spend the vast majority of our lectures on how to translate the diffracted beam intensities into atomic structure.

1.1 Addition of Electromagnetic Waves

X-rays are electromagnetic waves, so they are described by Maxwell's equations, from which we can infer that the electromagnetic fields are essentially just oscillating electric and magnetic fields, travelling in some direction along the wavevector $\mathbf{k}$ (the magnitude of $\mathbf{k}$ also encodes the wavelength via $\|\mathbf{k}\| = 2\pi/\lambda$) with the speed of light c.[1] We label the coordinate x along the propagation direction

---

[1]In free space, $\rho = 0$ and $\mathbf{J} = 0$, so by Maxwell's equations

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{E}) = -\boldsymbol{\nabla} \times \frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial}{\partial t}(\boldsymbol{\nabla} \times \mathbf{B}) = -\frac{\partial}{\partial t}\left(\mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}\right) = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}\,. \tag{1.1}$$

By vector calculus identities, we also have

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{E}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E}\,, \tag{1.2}$$

and hence

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}\,. \tag{1.3}$$

This is the wave equation, with wave speed $c = 1/\sqrt{\mu_0 \epsilon_0}$, with general complex solution

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{1.4}$$

for $\mathbf{E} \in \mathbb{C}$ and $\mathbf{k}^2 c^2 = \omega^2$. By Maxwell's equations, we also have

$$\frac{\partial \mathbf{B}}{\partial t} = -\boldsymbol{\nabla} \times \mathbf{E} = -i\mathbf{k} \times \mathbf{E}\,, \tag{1.5}$$
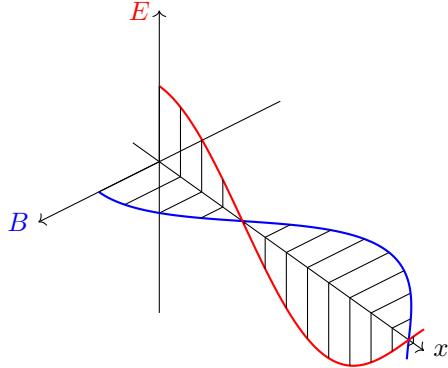
Figure 1.2: Electromagnetic wave propagating in space.
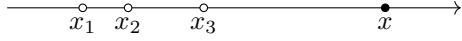
k, and consider the wave at some fixed time t. Then the electric (or equivalently the magnetic) field strength is given by

$$\psi(x) = A \sin\left(\frac{2\pi x}{\lambda} + \phi\right) , \tag{1.7}$$

where the real number A is the amplitude and ϕ is the phase offset. However, it turns out that when considering oscillations, it is always easier to use complex numbers, so we will alternatively see the electric field as the real part of the complex wave

$$\psi(x) = A \exp\left(\frac{2\pi \mathrm{i} x}{\lambda}\right) , \tag{1.8}$$

where $A = |A|\, \mathrm{e}^{\mathrm{i}\phi}$ is now complex and includes both the amplitude $|A|$ and the initial phase offset ϕ.

$$\begin{array}{ccccccc} & x_1 & x_2 & x_3 & & x & \end{array}$$

Now what happens if we have multiple sources, each sending off electromagnetic waves of the same frequencies but with different amplitudes and phases? Let's consider the simple 1D case, where n sources are located at x_j. We will let all n sources have a zero initial phase[2], so that the resulting wave at point x due to source j is given by

$$\psi_j(x) = A_j \exp\left(\frac{2\pi \mathrm{i}(x - x_j)}{\lambda}\right) , \tag{1.9}$$

where A_j is real because we have a zero initial phase. We will denote the phase of the wave due to source j at x as

$$\phi_j = \frac{2\pi(x - x_j)}{\lambda} . \tag{1.10}$$

Then to work out the total wave, we only need to sum up the contributions from each source to get

$$\Psi_{\mathrm{res}}(x) = \sum_j A_j \exp\left(\mathrm{i}\phi_j\right) . \tag{1.11}$$

It is easy to see that the phase angle of the resultant amplitude is

$$\Phi_{\mathrm{res}} = \tan^{-1}\left[\frac{\sum_j A_j \sin\phi_j}{\sum_j A_j \cos\phi_j}\right] . \tag{1.12}$$

---
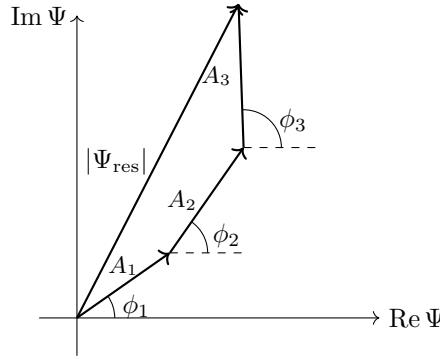
and so

$$\mathbf{B} = \frac{\mathbf{k}}{\omega} \times \mathbf{E} . \tag{1.6}$$

Therefore, **E** and **B** are both transverse waves that are in the same phase but perpendicular to each other.

[2]It is a trivial generalisation for the sources to have different non-zero initial phases. We make such assumption because this is the only situation we will need for later discussion, and it makes my diagram cleaner. It is actually also trivial to generalise it to 3D by replacing $x - x_j$ with $\|\mathbf{x} - \mathbf{x}_j\|$ in all the expressions below.

Now what if we move the measuring point x by an amount of δx? This will increase the phases of the wave from all the sources by

$$\delta\phi = \frac{2\pi\delta x}{\lambda}\,. \tag{1.13}$$

The net result is that the resultant magnitude of the combined wave does not change, but the phase changes by $\delta\phi$, as

$$\Psi'_{\mathrm{res}}(x) = \sum_j A_j \exp(\mathrm{i}(\phi_j + \delta\phi)) = \mathrm{e}^{\mathrm{i}\delta\phi} \sum_j A_j \exp(\mathrm{i}\phi_j)\,. \tag{1.14}$$
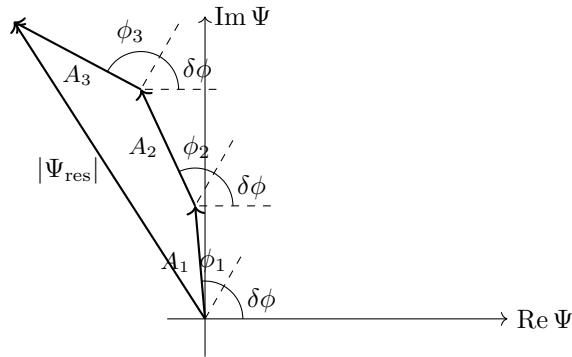
Figure 1.3: Movement of the measuring point results in a change of measured phase only.

What if we change the position of one of the sources x_k? This will in general lead to a change in the phase ϕ_k. Now both the phase and the amplitude of the resulting wave is different.
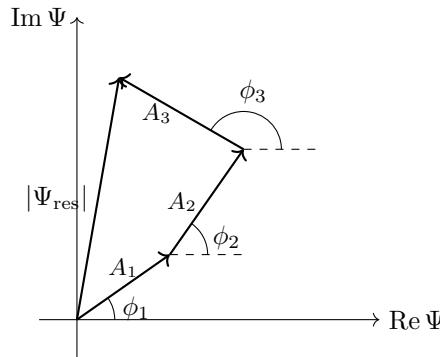
Figure 1.4: A change in the position of one of the sources leads to a change in both the magnitude and the phase measured.

3
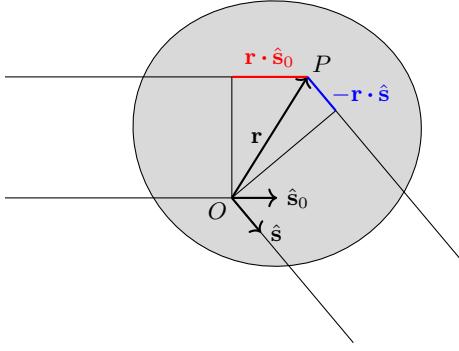
Figure 1.5: Path difference for diffractions in a general object.

1.2   Diffraction of an Object

We would now like to consider the diffraction of an extended object. Suppose we have parallel incident beams from the direction $\hat{\mathbf{s}}_0$ illuminating the whole object, and we are measuring the diffracted beam at some distant point Q. The dimension of this object is small compared to its distance to the measuring point Q so we can sensibly say that Q is at direction $\hat{\mathbf{s}}$ for all points in the object.[3] Moreover, we assume that the effect of diffraction is weak, so the diffraction only happens once — an already diffracted beam cannot be diffracted again. This is known as the *kinematic approximation.*[4]

To produce the diffraction pattern, we need to sum up the radiation scattered from all the points in the object. To do this, we first pick an arbitrary origin O in the object, and we need to figure out the intensity and phase difference between the beams diffracted in direction $\hat{\mathbf{s}}$ from O and from any other point P with position vector $\mathbf{r}$.

From figure 1.5, it is easy to see that the beam diffracted from P and O has a path difference

$$\Delta x = \mathbf{r} \cdot \hat{\mathbf{s}}_0 - \mathbf{r} \cdot \hat{\mathbf{s}} \,. \tag{1.15}$$

In particular, a positive path difference means that the beam from P falls behind the beam from O, leading to a negative phase difference, so

$$\Delta \phi = -\frac{2\pi \Delta x}{\lambda} = \frac{2\pi \mathbf{r} \cdot (\hat{\mathbf{s}} - \hat{\mathbf{s}}_0)}{\lambda} \,. \tag{1.16}$$

The amplitude of the beam diffracted from O and P may also be different, since the object may be inhomogeneous. Let $A(\mathbf{r})$ be a real function representing the diffraction amplitude, then by integrating over all points, the wave measured at point Q is

$$\Psi(\hat{\mathbf{s}}) = \int \mathrm{d}^3 \mathbf{r} \, A(\mathbf{r}) \exp\left(\frac{2\pi i \mathbf{r} \cdot (\hat{\mathbf{s}} - \hat{\mathbf{s}}_0)}{\lambda} \right) \,. \tag{1.17}$$

The integral is over the object volume.

We can define the *scattering vector* $\mathbf{S}$ by

$$\mathbf{S} := \frac{\hat{\mathbf{s}} - \hat{\mathbf{s}}_0}{\lambda} \,. \tag{1.18}$$

It is a vector bisecting the external angle between the incident and diffracted beam, as shown in the figure below. If we define the angle of deflection to be 2θ, then the length of the scattering vector is

$$\|\mathbf{S}\| = \frac{2 \sin \theta}{\lambda} \,. \tag{1.19}$$

---

[3]This type of diffraction is known as *Fraunhofer diffraction*. If you are observing near the diffracting object, then the diffraction is *Fresnel diffraction*.

[4]If this assumption is false (as it will be for electron diffractions), we need to consider a more complicated *dynamic scattering.*

The scattering vector has dimension $[\text{L}]^{-1}$, so it exists in the reciprocal space, which might be familiar from Part IB Chemistry A.[5]
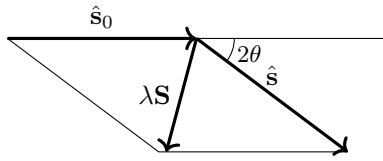

Figure 1.6: Construction of the scattering vector.

This allows us to simplify our expression of the measured wave as

$$\Psi(\mathbf{S}) = \int \mathrm{d}^3\mathbf{r}\, A(\mathbf{r}) \exp\left(2\pi i \mathbf{r} \cdot \mathbf{S}\right) . \tag{1.21}$$

1.3 Diffraction of X-rays from Electrons

We will now consider how X-rays interacts with matter. The interaction generally fall into two categories.

- *Elastic scattering.* This is our main focus. In an *elastic scattering*, or *Thomson scattering*, there is no net transfer of energy from the photon to the material. It can be thought of as some kind of resonance of the atomic electrons. When X-rays pass through the material, the oscillating electric field causes the atomic electrons to oscillate, which in turn emit X-rays in all directions. The incoming and outgoing rays retain a non-random phase relationship and are said to be *coherent.*

- *Inelastic processes. Inelastic processes* involve energy transfer from the incoming X-rays to electrons in atoms. The simplest example is photoelectron emission, where an X-ray photon providing energy for an atom to eject a core-shell electron. The photons are then re-emitted with lower energy in an incoherent fashion. This incoherence means that they cannot interfere with each other to produce diffraction patterns. Other processes of this type may result from interaction with valence shell electrons or vibrations, all being incoherent. We are therefore not interested in inelastic processes at this stage — we will return to this later.

The quantitative description of Thomson scattering strictly only applies to free electrons, but the results are found to be generally applicable to crystals under normal scattering conditions. Crucially, the amplitude of X-rays diffracted from some point is proportional to the local electron density $\rho(\mathbf{r})$. Therefore,

$$\Psi(\mathbf{S}) = \int \mathrm{d}^3\mathbf{r}\, \sigma\rho(\mathbf{r}) \exp(2\pi i \mathbf{r} \cdot \mathbf{S}) , \tag{1.22}$$

where σ is the scattering power of a single electron. We can divide through this constant of proportionality to obtain the *structure factor* $F(\mathbf{S})$ given by

$$F(\mathbf{S}) = \int \mathrm{d}^3\mathbf{r}\, \rho(\mathbf{r}) \exp\left(2\pi i \mathbf{r} \cdot \mathbf{S}\right) . \tag{1.23}$$

The limits of the two integrals above are both throughout the whole 3D space. Mathematically, $F(\mathbf{S})$ is the *Fourier transform* of the electron density $\rho(\mathbf{r})$.

---

[5]In some derivation, one uses *momentum transfer vector* $\mathbf{Q}$ defined by

$$\mathbf{Q} = \frac{2\pi(\hat{\mathbf{s}} - \hat{\mathbf{s}}_0)}{\lambda} . \tag{1.20}$$

instead of $\mathbf{S}$.

Definition 1.1. Let $f : \mathbb{R}^n \to \mathbb{C}$ be a function. The *Fourier transform* of f is $\mathcal{F}[f] \equiv \tilde{f} : \mathbb{R}^n \to \mathbb{C}$ given by[6]

$$\tilde{f}(\mathbf{k}) := \int \mathrm{d}^n \mathbf{x}\, f(\mathbf{x}) \mathrm{e}^{2\pi \mathrm{i} \mathbf{k} \cdot \mathbf{x}}. \tag{1.25}$$

Theorem 1.2. Let $\tilde{f}(\mathbf{k})$ be the Fourier transform of $f(\mathbf{x})$. The *inverse Fourier transform* that converts $\tilde{f}$ back to f is given by

$$f(\mathbf{x}) = \mathcal{F}^{-1}[\tilde{f}(\mathbf{k})] := \int \mathrm{d}^3 \mathbf{k}\, \tilde{f}(\mathbf{k}) \mathrm{e}^{-2\pi \mathrm{i} \mathbf{k} \cdot \mathbf{x}}. \tag{1.26}$$

Proof.

$$\begin{aligned}
\mathcal{F}^{-1}[\tilde{f}](\mathbf{x}) &= \int \mathrm{d}^n \mathbf{k}\, \mathrm{e}^{-2\pi \mathrm{i} \mathbf{k} \cdot \mathbf{x}} \int \mathrm{d}^3 \mathbf{s}\, \mathrm{e}^{2\pi \mathrm{i} \mathbf{k} \cdot \mathbf{s}} f(\mathbf{s}) \\
&= \int \mathrm{d}^n \mathbf{s}\, f(\mathbf{s}) \int \mathrm{d}^n \mathbf{k}\, \mathrm{e}^{2\pi \mathrm{i} k \cdot (\mathbf{s} - \mathbf{x})} \\
&= \int \mathrm{d}^n \mathbf{s}\, f(\mathbf{s}) \delta(\mathbf{s} - \mathbf{x}) \\
&= f(\mathbf{x}).
\end{aligned} \tag{1.27}$$

$\square$

Hence to convert the $F(\mathbf{S})$ back to $\rho(\mathbf{r})$, one simply needs to perform the inverse Fourier transform

$$\rho(\mathbf{r}) = \int \mathrm{d}^3 \mathbf{S}\, F(\mathbf{S}) \exp(-2\pi \mathrm{i} \mathbf{r} \cdot \mathbf{S}). \tag{1.28}$$

Again, the integral is throughout the whole 3D space, and it is omitted.

This means that if we can determine the $F(\mathbf{S})$, then we can easily reconstruct the electron density in a material, no matter how complex the structure is. (1) Measure the phases and the amplitudes of the diffracted beams. (2) Divide by single-electron scattering power σ to get $F(\mathbf{S})$. (3) Do the inverse Fourier transform to obtain $\rho(\mathbf{r})$. Simple as that! Three steps to solve any chemical structure in the world!

This should be the happy ending of our story on X-ray diffraction.

---

[6]You might be more familiar with the Fourier transform defined as

$$\tilde{f} = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}x\, f(x) \mathrm{e}^{-\mathrm{i}kx}, \tag{1.24}$$

where the normalisation factor of $1/\sqrt{2\pi}$ may or may not be there. It is just a matter of convention. However, this difference in convention is annoying since it results in slight tweaks in a lot of formulae.

2   The Phase Problem and the Patterson Function

2.1   The Phase Problem

If the problem is really that simple, then this would be a one-lecture course instead of a 12-lecture one. There is a huge issue. We cannot measure the phase of an X-ray.
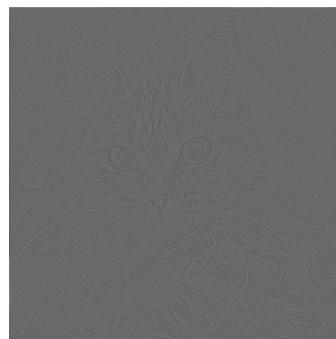
To measure the magnitude and the phase of an electromagnetic wave, we would use an antenna. However, this requires the antenna to have its dimensions comparable to the radiation wavelength. The X-rays we used for diffraction are at atomic length scale — that's why we can see the diffraction, but it is impossible to make an antenna at atomic length scale since antennas should be made from atoms! All we can measure in a diffraction experiment is the number of X-ray photons encountered in a given time, which is the diffracted *intensity*. This is proportional to $I = |F(\mathbf{S})|^2$. The phase information is completely lost. This is referred to as the *phase problem* in diffraction.

Now the question is: is getting the correct phase important? If using a wrong phase of the diffracted beam has little impact on the electron density calculated, then this would not be a huge problem. We can calculate the modulus of $F(\mathbf{S})$ by taking the square root of the intensity, and just plug in some random phases to calculate $\rho(\mathbf{r})$. However, it turns out that getting the correct phase is extremely important — it even plays a more important role than the magnitude to some extent. This can be illustrated by the following little experiment.
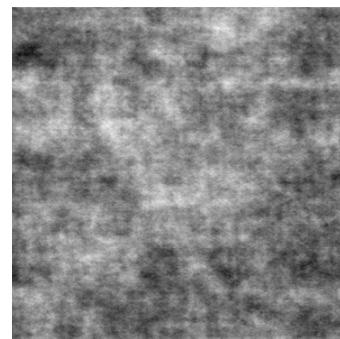
Instead of doing Fourier transform of a electron density function, we do a Fourier transform on a picture. A black and white picture is also a function of the pixels, where the values of the function is the gray scale. We can transform it into the frequency domain, and then transform it back, and we get the original image. However, if we use some randomly generated phase values to replace the correct phase values in the Fourier transform of the image, and we transform it back, we see a complete mess. This is what will happen if we use random phase values for the structure factor — we will get no valuable information. In fact, if we use some random value for the magnitudes of the Fourier transform, with the correct phase values, we can still vaguely see the shape of a cat.

(a) Correct magnitudes + correct phases
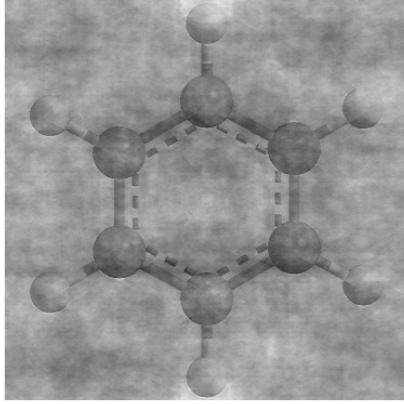
(b) Random magnitudes + correct phases

(c) Correct magnitudes + random phases

Figure 2.1: The effect of using the wrong phases and magnitudes in an inverse Fourier transform.

Just for fun, we can extend this experiment further. We have two images: one is a benzene molecule and the other is a cat. What will happen if we use the magnitude value of one figure with the phase value of the other? The result is shown in (2.2), and you can see it is the phase that determines what we will see. In general, we don't want to see a cat if we are investigating the structure of benzene using diffraction, so it's not a good idea just to use a random set of phases when doing inverse Fourier transform.

(a) Magnitudes of cat + phases of benzene   (b) Magnitudes of benzene + phases of cat

Figure 2.2: Phases govern what we see in an inverse Fourier transform. If you use the phases of a benzene, you see a benzene. If you use the phases of a cat, you see a cat.

Moral: phases are important, but we can't measure them. The main task of structure determination using X-ray diffraction is to deduce the phases of the structure factor. This is what we will spend the remaining 11 lectures on.

2.2   Patterson Function

Since all what we can get from measuring a diffraction pattern is the intensity $I(\mathbf{S}) = |F(\mathbf{S})|^2$, let's see what its inverse Fourier transform gives us anyway.

Before doing that, let's first introduce convolution and correlation.

Definition 2.1. For two functions $f, g : \mathbb{R}^n \to \mathbb{C}$, their *convolution* is

$$(f * g)(\mathbf{x}) := \int d^n\mathbf{s}\, f(\mathbf{s})g(\mathbf{x} - \mathbf{s})\,, \tag{2.1}$$

and their *correlation* is

$$(f \otimes g)(\mathbf{x}) := \int d^n\mathbf{s}\, f(\mathbf{s})^* g(\mathbf{x} + \mathbf{s})\,. \tag{2.2}$$

Theorem 2.2 (Convolution theorem). The Fourier transform of a convolution is the product of Fourier transforms.

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]\,. \tag{2.3}$$

Proof.

$$
\begin{aligned}
\mathcal{F}[f * g](\mathbf{k}) &= \int d^n\mathbf{x}\, e^{2\pi i \mathbf{k} \cdot \mathbf{x}} \int d^n\mathbf{s}\, f(\mathbf{s})g(\mathbf{x} - \mathbf{s}) \\
&= \int d^n\mathbf{s}\, f(\mathbf{s}) \int d^n\mathbf{x}\, e^{2\pi i \mathbf{k} \cdot \mathbf{x}} g(\mathbf{x} - \mathbf{s}) \\
&= \int d^n\mathbf{s}\, f(\mathbf{s}) \int d^n\mathbf{u}\, e^{2\pi i \mathbf{k} \cdot (\mathbf{u}+\mathbf{s})} g(\mathbf{u}) \qquad \text{define } \mathbf{u} = \mathbf{x} - \mathbf{s} \\
&= \int d^n\mathbf{s}\, e^{2\pi i \mathbf{k} \cdot \mathbf{s}} f(\mathbf{s}) \int d^n\mathbf{u}\, e^{2\pi i \mathbf{k} \cdot \mathbf{u}} g(\mathbf{u}) \\
&= \tilde{f}(\mathbf{k})\tilde{g}(\mathbf{k})\,. \tag{2.4}
\end{aligned}
$$

$\square$
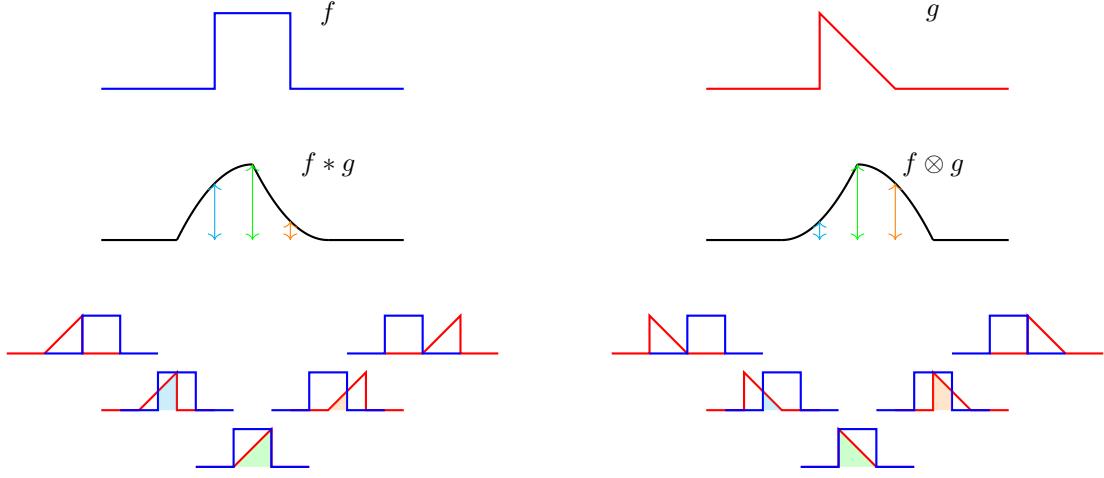
Figure 2.3: Graphic explanation of convolution and correlation.

Theorem 2.3. Conversely, the Fourier transform of a product fg is the convolution of their Fourier transforms

$$\mathcal{F}[fg] = \mathcal{F}[f] * \mathcal{F}[g]. \tag{2.5}$$

Proof.

$$
\begin{aligned}
\mathcal{F}[fg](\mathbf{k}) &= \int \mathrm{d}^n\mathbf{x}\, \mathrm{e}^{2\pi i \mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) g(\mathbf{x}) \\
&= \int \mathrm{d}^n\mathbf{x}\, \mathrm{e}^{2\pi i \mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}) \int \mathrm{d}^n\boldsymbol{\ell}\, \mathrm{e}^{-2\pi i \boldsymbol{\ell}\cdot\mathbf{x}} \tilde{g}(\boldsymbol{\ell}) \\
&= \int \mathrm{d}^n\boldsymbol{\ell}\, \tilde{g}(\boldsymbol{\ell}) \int \mathrm{d}^n\mathbf{x}\, \mathrm{e}^{2\pi i \mathbf{x}\cdot(\mathbf{k}-\boldsymbol{\ell})} f(\mathbf{x}) \\
&= \int \mathrm{d}^n\boldsymbol{\ell}\, \tilde{g}(\boldsymbol{\ell}) \tilde{f}(\mathbf{k}-\boldsymbol{\ell}) \\
&= \tilde{f} * \tilde{g}(\mathbf{k}). \tag{2.6}
\end{aligned}
$$

$\square$

We see a beautiful duality here. Product in the real domain is the convolution in the Fourier domain, and *vice versa*.

Theorem 2.4 (Wiener–Khinchin theorem). The Fourier transform of the autocorrelation of a function is its *power spectral intensity*.

$$\mathcal{F}[f \otimes f] = \left|\tilde{f}\right|^2. \tag{2.7}$$

Proof.

$$
\begin{aligned}
\mathcal{F}[f \otimes f](\mathbf{k}) &= \int \mathrm{d}^n\mathbf{x}\, \mathrm{e}^{2\pi i \mathbf{k}\cdot\mathbf{x}} \int \mathrm{d}^n\mathbf{s}\, f^*(\mathbf{s}) f(\mathbf{x}+\mathbf{s}) \\
&= \int \mathrm{d}^n\mathbf{s}\, f^*(\mathbf{s}) \int \mathrm{d}^n\mathbf{x}\, \mathrm{e}^{2\pi i \mathbf{k}\cdot\mathbf{x}} f(\mathbf{x}+\mathbf{s}) \\
&= \int \mathrm{d}^n\mathbf{s}\, f^*(\mathbf{s}) \int \mathrm{d}^n\mathbf{u}\, \mathrm{e}^{2\pi i \mathbf{k}\cdot(\mathbf{u}-\mathbf{s})} f(\mathbf{u}) \qquad \text{define } \mathbf{u} = \mathbf{x}+\mathbf{s} \\
&= \int \mathrm{d}^n\mathbf{s}\, \mathrm{e}^{-2\pi i \mathbf{k}\cdot\mathbf{s}} f^*(\mathbf{s}) \int \mathrm{d}^n\mathbf{u}\, \mathrm{e}^{2\pi i \mathbf{k}\cdot\mathbf{u}} f(\mathbf{u}) \\
&= \tilde{f}^*(\mathbf{k})\tilde{f}(\mathbf{k}) = \left|\tilde{f}(\mathbf{k})\right|^2. \tag{2.8}
\end{aligned}
$$

$\square$

Finally, if we apply both sides on the Wiener–Khinchin theorem, we get

$$\mathcal{F}^{-1}\left[|\tilde{f}|^2\right] = f \otimes f. \tag{2.9}$$
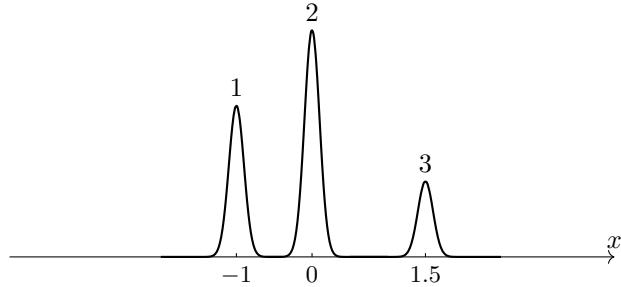
We can apply this to our diffraction intensity! We previously identified the structure factor to be the Fourier transform of the electron density, $F(\mathbf{S}) = \mathcal{F}[\rho(\mathbf{x})]$, and the diffraction intensity is exactly $I(\mathbf{S}) = |F(\mathbf{S})|^2$. Therefore, the inverse Fourier transform of the diffraction intensity is exactly the autocorrelation of the electron density!

$$\mathcal{F}^{-1}[I(\mathbf{S})] = \rho \otimes \rho(\mathbf{r}). \tag{2.10}$$

We often call this the *Patterson function*, denoted $P(\mathbf{r})$.

2.3   Illustrating the Patterson Function

The above derivation is a bit abstract. Let's do a simple 1D example of an electron density function with three peaks, modelling three atoms in a line, and see how its Patterson function look like.
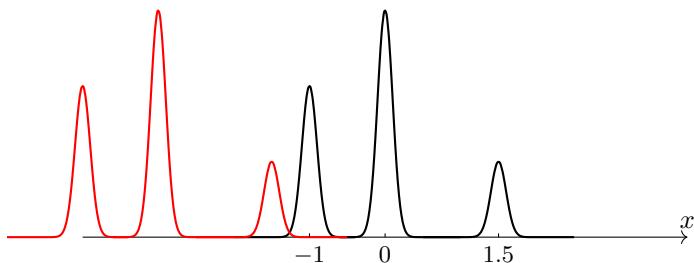

We put the three peaks at positions $x = -1$, $x = 0$ and $x = 1.5$ with weights 2, 3 and 1 respectively. From the definition of the Patterson function, which is the correlation of the electron density with itself
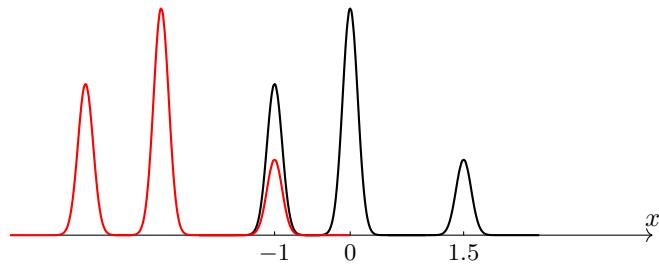
$$P(u) = \int \mathrm{d}x\, \rho(x)\rho(x+u), \tag{2.11}$$

we need to take another copy of electron density function, shift it by some distance u, and evaluate the integral of their product, which is roughly governed by how well do the peaks of the two functions overlap. Let's set u to a range of different values and see what happens.
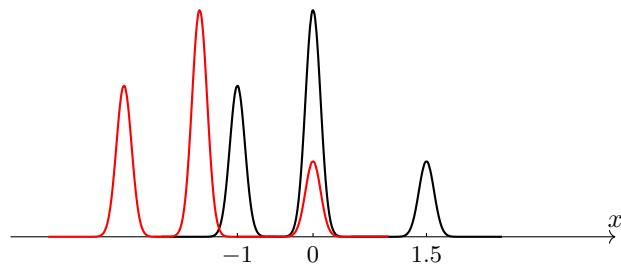
- $u = -3$: First, when u is very negative, the two function has almost no overlap, so at this value of u, $P(u) \approx 0$.
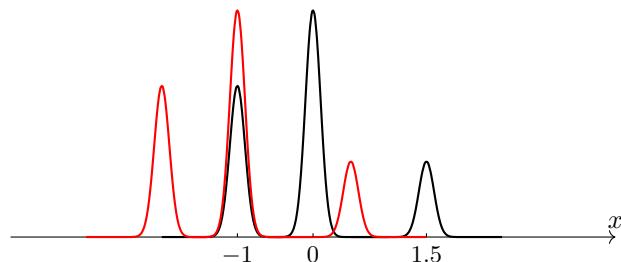
- $u = -2.5$: At this value of u, peak 1 and peak 3 just overlaps. Without the actual functional form of the peak, we can't evaluate this integral of their product, but since peak 1 has weight 2 and peak 3 has weight 1, we can roughly assign a relative weight of $2 \times 1 = 2$ to the Patterson function here.
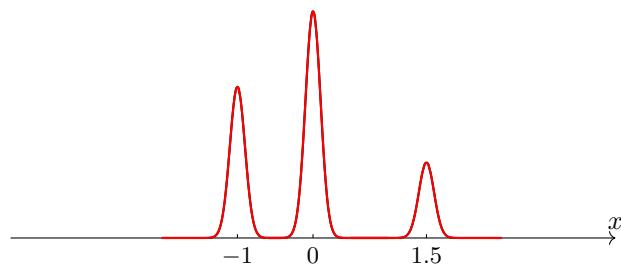
- $u = -1.5$: Now at this value of u, peak 3 and peak 2 overlaps, giving a peak in the Patterson function with relative weight $3 \times 1 = 3$.

- $u = -1$: Peak 1 and peak 2 overlap, giving a peak in the Patterson function with relative weight $2 \times 3 = 6$.

- $u = 0$: At this special value of u, the second function is not moved at all, so all the three peaks are perfectly aligned. This leads to a very large peak in the Patterson function with weight $2^2 + 3^2 + 1^2 = 14$.

You can go further to check other values of u, but we should have done enough to discover a pattern. The Patterson function will have peaks at the values when there are two peaks in the electron density function align, which are exactly the "inter-peak" distances in the electron density. The relative weight of the peak in the Patterson function is the product of the weight of the peaks in the electron density.

11
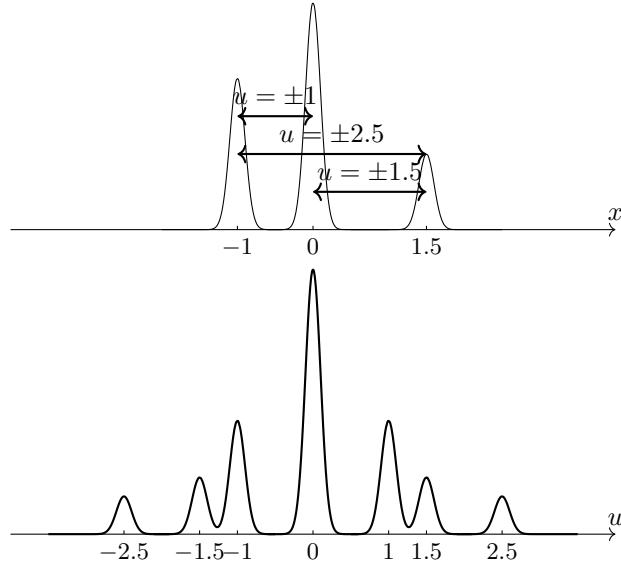
Figure 2.4: Peaks in the Patterson function correspond to the inter-peak vectors in the electron density function.

To conclude, let's summarise some features of the Patterson function:

1. The Patterson function is defined in the same space as $\rho(\mathbf{r})$ (as opposed to the structure factor defined in the reciprocal space), and it spans a range twice that spanned by $\rho(\mathbf{r})$.

2. The Patterson function is always centrosymmetric, even if $\rho(\mathbf{r})$ is not.

3. The maximum of the Patterson function occurs at $\mathbf{u} = \mathbf{0}$, and its value is

$$P(\mathbf{0}) = \int \mathrm{d}^3\mathbf{r}\,\rho^2(\mathbf{r}) \, . \tag{2.12}$$

4. Local maxima in $P(\mathbf{u})$ occur when peaks in $\rho(\mathbf{r})$ overlaps with those in $\rho(\mathbf{u}+\mathbf{r})$. This happens if the offset vector $\mathbf{u}$ is some peak↔peak vector in $\rho(\mathbf{r})$. If we take an atomic picture of the electron density, in which electron densities are concentrated on the atomic sites (we will formalise this latter), then peaks in $P(\mathbf{u})$ occur when $\mathbf{u}$ is an *interatomic vector*.

5. If there are N peaks in $\rho(\mathbf{r})$, there are N^2 interatomic vectors. Of these, N will occur at $\mathbf{u} = \mathbf{0}$, corresponding to vectors of an atom pointing itself. Consequently, there will in general be $N^2 - N$ non-origin peaks in $P(\mathbf{u})$ (some of them may overlap as well).

6. For a function $\rho(\mathbf{r})$ consists of resolved peaks, the total weight (area) of the peak produced in $P(\mathbf{u})$ is proportional to the product of the weights of the peaks in $\rho(\mathbf{r})$ and $\rho(\mathbf{r}+\mathbf{u})$ producing it. If atoms in $\rho(\mathbf{r})$ and $\rho(\mathbf{r}+\mathbf{u})$ have atomic numbers Z_1 and Z_2, then the weight of the peaks in $P(\mathbf{u})$ will be proportional to $Z_1 Z_2$. Consequently, Patterson peaks involving heavy atoms (where Z_1 and/or Z_2 is large) will show up most clearly.

7. Peaks in $P(\mathbf{u})$ will be broader than in $\rho(\mathbf{r})$, and peaks corresponding to vectors from light atoms may merge into a poorly-defined background.

Hence, as we will see later, the Patterson function will be helpful in locating heavy atoms in a structure.

2.4 Patterson Function of a Molecule

Now consider a SO_3 molecule. For simplicity, we will make this problem 2 dimensional. The atomic number of S and O are 16 and 8 respectively, so the electron density function has one peak of weight 16, which we set at the origin, and three peaks of weight 8 at the same distance from the origin, separated by 120°.
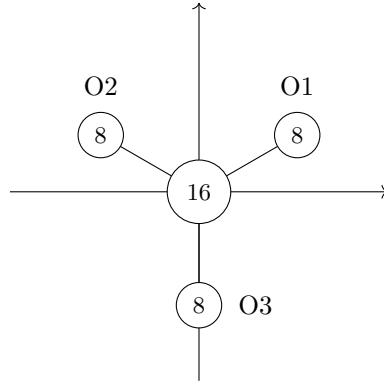
Figure 2.5: The electron density of a SO_3 molecule.
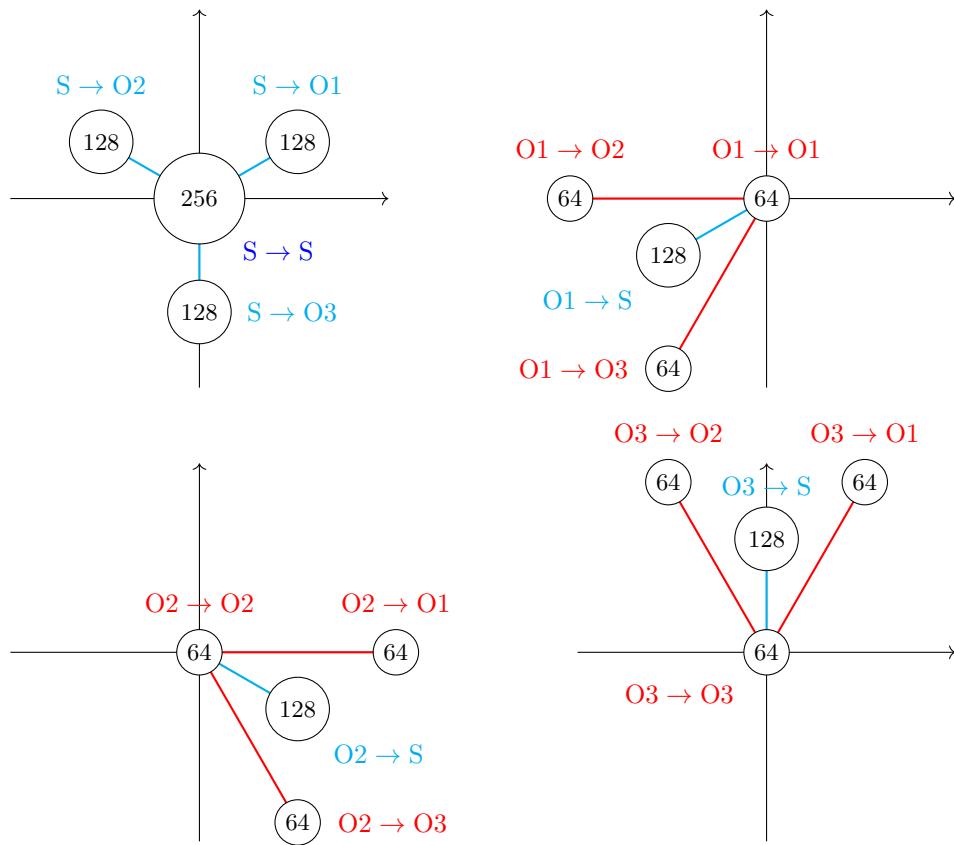
Figure 2.6: Interatomic vectors of SO_3.

To evaluate the Patterson function, we need to find all the interatomic vectors. Let's first consider vectors from the central S atom. There will be a self-vector $(S \rightarrow S)$ at the origin, with weight $16 \times 16 = 256$, and there will be three $S \rightarrow O$ vectors arranged symmetrically around origin, with weight $16 \times 8 = 128$ each. Their distance to the origin is the same as the $S - O$ bond length, which we will denote as r_1. Note that this looks like the molecule itself, with the S atom sitting at the origin.

13

Then consider the vectors from the atom O1. There will be a self vector (O1 → O1) at the origin, with weight $8 \times 8 = 64$. There are two other O → O vectors, O1 → O2 and O1 → O2, at $60°$ to each other, with weight 64 and length $\sqrt{3}r_1$. Bisecting these two is a S → O1 vector with weight 128 and length r_1. Note this again look like the molecule itself, but now with O1 sitting at the origin. Similarly, we can consider vectors from atoms O2 and O3 as well.

Then to construct the Patterson function, we only need to add these contributions up. The result is shown in the figure below.
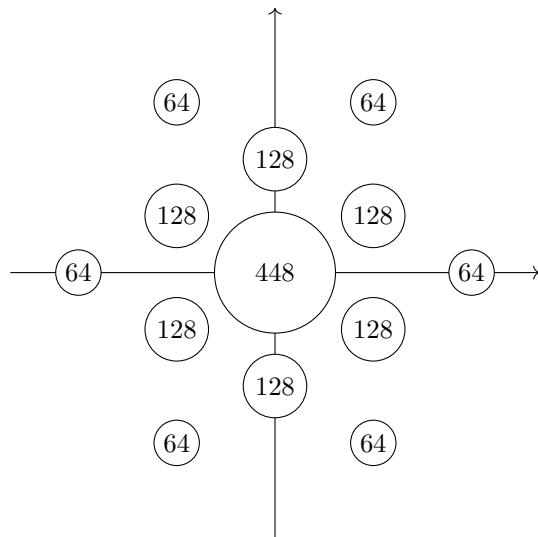
Figure 2.7: The Patterson function of a SO_3 molecule.

Some remarks:

- The weight of the origin is the sum of the squares the atomic numbers of all atoms.

- The Patterson function includes vectors between all atoms, not just between the chemically bonded ones.

- $P(\mathbf{u})$ must have a centre of symmetry at the origin, even though SO_3 does not have one. The point group of the SO_3 molecule is D_{3h}, and the centre of symmetry have upgraded the point group of the Patterson function to D_{6h}.

One can use the same principle to construct the Patterson of a 3D molecule. Below an example of the Patterson function of SO_4^{2-} ion, with T_d symmetry. The Patterson function has an extra inversion centre, so its point group is O_h.

2.5   The Patterson Function of Ideal Gas

We now move on to a more realistic case where there is a whole collection of molecules, and we will consider two extremes — an ideal gas and a crystalline solid.

First, let's consider an ideal gas. There are no intermolecular forces, negligible molecular volumes, and we will assume that all molecules are perfectly rigid and geometrically identical. The Patterson peaks will have contribution from all molecules: both intermolecular vectors and intramolecular vectors. For the vectors pointing within a molecule, the interatomic vectors should be identical to that for a single molecule, multiplied by the total number of molecules that are present. But as we
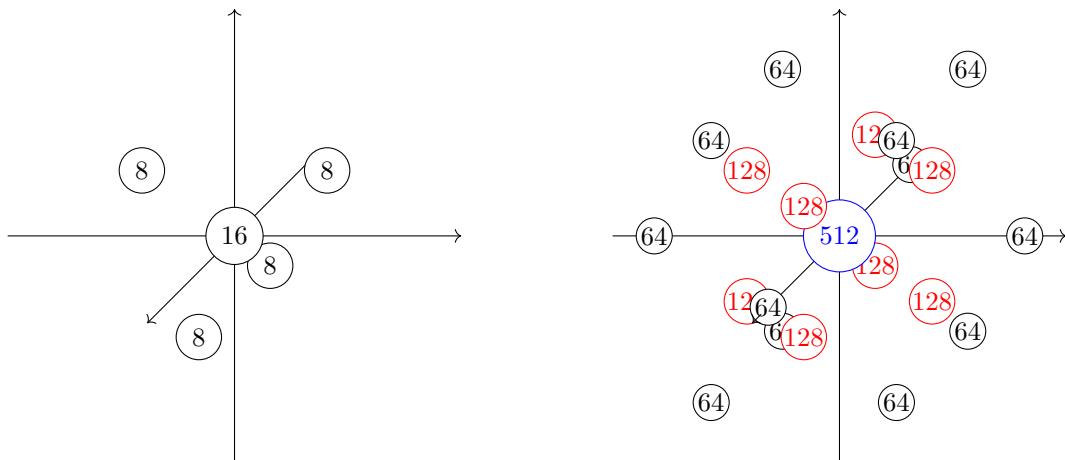
Figure 2.8: The electron density and the Patterson function of SO_4^{2-}.

are in the gas phase, we have another complexity. Due to the random molecular orientation, the Patterson function of an ideal gas appears as a spherically averaged version of the 3D Patterson that we constructed for one molecule. Thus, only the length of the interatomic vector is important, and we will plot the resulting Patterson function as a radial vector map. For example, for the SO_4^{2-} ideal gas (which apparently will not be ideal in reality), the radial Patterson function is shown in the figure below.[7]
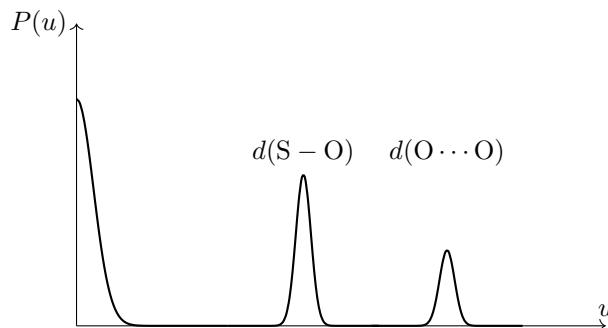
Figure 2.9: Intramolecular contribution to the radial Patterson function of SO_4^{2-}.

We then need to consider the intermolecular vectors. If the gas is indeed ideal with no intermolecular forces, then the positions and orientations of all molecules should be completely random at any time. Hence, all distances between molecules are equally possible, and so the intermolecular vectors will sum over time to form a constant background. If we are going to inject a little bit of reality into this crude model, we can say that the molecules cannot be too close from each other, so this background is zero at small u.

As a result, in an ideal gas, the radial Patterson function has discernible peaks corresponding solely to the spherically averaged Patterson function of one molecule, superimposed on a featureless background. Non-ideal gases, liquid and non-crystalline solid will show similar radial Patterson function due their disorderedness. However, they may show local coordination peaks.

---

[7]As for the atomic wavefunction, you can choose to plot a radial Patterson function or a radial distribution function — we are being rather loose here since we are doing everything schematically. Anyway, the true weight of a peak is the integral of the 3D Patterson function.
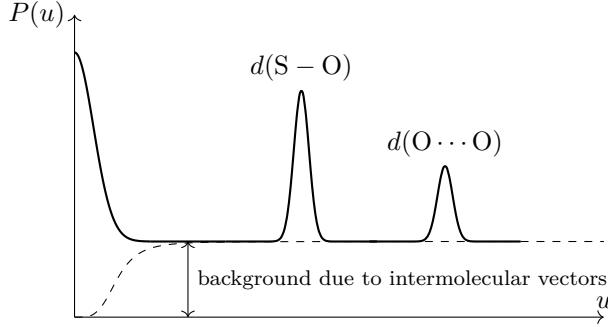
Figure 2.10: Radial Patterson function of SO_4^{2-} with both intermolecular and intramolecular contributions.

3 Diffraction from a Single Crystal

In the remainder of the course, we will focus almost exclusively on single crystals, since as we will see, we can develop robust methods to interpret single-crystal diffraction patterns.

3.1 Diffracted Intensity for a Single Crystal

We will first infer how the Patterson function of a crystal would look like — this shouldn't be too difficult due to the order in a crystal, and then we will perform the Fourier transform to try to figure out the form of its diffracted intensity.

We will consider a single crystal with unit cell vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, with $N = N_a \times N_b \times N_c$ unit cells in total, where N_a, N_b, N_c are the number of cells along each direction.

Suppose the Patterson function of a single unit cell is $P_{\text{cell}}(\mathbf{u})$, then its Fourier transform will give the diffracted intensity from one unit cell, which we will call $|F_{\text{cell}}(\mathbf{S})|^2$.

$$|F_{\text{cell}}|^2 = \int \mathrm{d}^3\mathbf{u}\, e^{2\pi i \mathbf{u} \cdot \mathbf{S}} P_{\text{cell}}(\mathbf{u}) \,. \tag{3.1}$$

Then for our single crystal, there will be N such contributions to the diffracted intensities, so we get the intra-cell contribution $N\, |F_{\text{cell}}(\mathbf{S})|^2$.

Next consider the contribution to the Patterson function by inter-cell vectors pointing from each cell to the cell next to it along the $+\mathbf{a}$ direction. For each intra-cell vector $\mathbf{u}$, there must be such a inter-cell vector $\mathbf{u} + \mathbf{a}$, so the contribution to the Patterson function from this type of inter-cell vectors is $P_{\text{cell}}(\mathbf{u} + \mathbf{a})$. This contributes

$$\int \mathrm{d}^3\mathbf{u}\, P_{\text{cell}}(\mathbf{u} + \mathbf{a}) e^{2\pi i \mathbf{u} \cdot \mathbf{S}} \tag{3.2}$$

to the diffraction intensity. We can simplify this a bit further. By the change of variable $\mathbf{v} = \mathbf{u} + \mathbf{a}$, we have

$$\int \mathrm{d}^3\mathbf{u}\, P_{\text{cell}}(\mathbf{u} + \mathbf{a}) e^{2\pi i \mathbf{u} \cdot \mathbf{S}} = \int \mathrm{d}^3\mathbf{v}\, P_{\text{cell}}(\mathbf{v}) e^{2\pi i \mathbf{v} \cdot \mathbf{S}} e^{-2\pi i \mathbf{a} \cdot \mathbf{S}}$$
$$= |F_{\text{cell}}|^2\, e^{-2\pi i \mathbf{a} \cdot \mathbf{S}} \,. \tag{3.3}$$

However, there are only $(N_a - 1)N_b N_c$ cells have such a neighbouring cell one unit along $+\mathbf{a}$ direction, so its contribution to the diffraction intensity is

$$(N_a - 1)N_b N_c\, |F_{\text{cell}}|^2\, e^{-2\pi i \mathbf{a} \cdot \mathbf{S}} \,. \tag{3.4}$$

Similarly, there are also contributions from inter-cell vectors pointing from a cell to the neighbouring cell along $-\mathbf{a}$ direction, which is

$$(N_a - 1)N_b N_c \left|F_{\text{cell}}\right|^2 e^{+2\pi i \mathbf{a} \cdot \mathbf{S}}. \tag{3.5}$$

The rest of the process is similar, there will be inter-cell vectors pointing from one cell to the cell $n_a \mathbf{a} + n_b \mathbf{b} + n_c \mathbf{c}$ away for any $-N_a < n_a < N_a$, $-N_b < n_b < N_b$, $-N_c < n_c < N_c$, contributing

$$(N_a - |n_a|)(N_b - |n_b|)(N_c - |n_c|)\left|F_{\text{cell}}\right|^2 e^{-2\pi i (n_a \mathbf{a} + n_b \mathbf{b} + n_c \mathbf{c}) \cdot \mathbf{S}} \tag{3.6}$$

to the total diffraction intensity. We need to sum all of these up, and we get a total diffraction intensity

$$
\begin{aligned}
I(\mathbf{S}) &= \left|F_{\text{cell}}\right|^2 \sum_{n_a = -N_a}^{N_a} \sum_{n_b = -N_b}^{N_b} \sum_{n_c = -N_c}^{N_c} (N_a - |n_a|)(N_b - |n_b|)(N_c - |n_c|) e^{-2\pi i (n_a \mathbf{a} + n_b \mathbf{b} + n_c \mathbf{c}) \cdot \mathbf{S}} \\
&= \left|F_{\text{cell}}\right|^2 \prod_{j \in \{a,b,c\}} \sum_{n_j = -N_j + 1}^{N_j - 1} (N_j - |n_j|) e^{-2\pi i n_j \mathbf{j} \cdot \mathbf{S}}.
\end{aligned} \tag{3.7}
$$

The latter form is rather more compact, but less intuitive.[8]

A more Mathematical View

Consider the convolution of some function $f(\mathbf{r})$ with a Dirac delta $\delta(\mathbf{r} - \mathbf{a})$ with its peak at $\mathbf{r} = \mathbf{a}$.

$$f(\mathbf{r}) * \delta(\mathbf{r} - \mathbf{a}) = \int_{-\infty}^{\infty} d^3\mathbf{u}\, f(\mathbf{u})\delta(\mathbf{u} - (\mathbf{r} - \mathbf{a})) = f(\mathbf{r} - \mathbf{a}). \tag{3.8}$$

We see that its effect is to shift $f(\mathbf{r})$ by $\mathbf{a}$ — or in other words, it put a copy of $f(\mathbf{r})$ at position $\mathbf{a}$. A crystal is essentially a copy of ρ_{cell} at each lattice point[9]. This allows us to view the electron density of a crystal as the convolution of the lattice of Dirac delta functions with the electron density of a single unit cell:

$$\rho(\mathbf{r}) = \rho_{\text{cell}}(\mathbf{r}) * \rho_{\text{lattice}}(\mathbf{r}), \tag{3.9}$$

where

$$\rho_{\text{lattice}}(\mathbf{r}) = \sum_{n_a=0}^{N_a-1} \sum_{n_b=0}^{N_b-1} \sum_{n_c=0}^{N_c-1} \delta(\mathbf{r} - n_a \mathbf{a} - n_b \mathbf{b} - n_c \mathbf{c}) \tag{3.10}$$

is the lattice of delta functions (It is obviously not a electron density, but we will denote it as ρ anyway). A rather useful properties is the Patterson function of a convolution is the convolution of the Patterson functions,[10] and so

$$P(\mathbf{u}) = P_{\text{cell}}(\mathbf{u}) * P_{\text{lattice}}(\mathbf{u}). \tag{3.12}$$

---

[8]You might have noticed that the limits of the sum should be $\pm(N_j - 1)$ instead of N_j. We have included $n_j = \pm N_j$ into our sum since they contribute nothing and can make our formula slightly more compact.

[9]This is not quite accurate as a unit cell may contain more than one lattice point in non-primitive cell. This is not a crystallographic course so we will be rather loose here.

[10]This is perhaps not too obvious to prove using the integral definitions of these quantities. However, an important fact of electron density is that it is a real function, so $P(\mathbf{r}) = \rho(\mathbf{r}) \otimes \rho(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r})$. Then

$$
\begin{aligned}
P(\mathbf{r}) &= \rho(\mathbf{r}) * \rho(-\mathbf{r}) \\
&= [\rho_{\text{cell}}(\mathbf{r}) * \rho_{\text{lattice}}(\mathbf{r})] * [\rho_{\text{cell}}(-\mathbf{r}) * \rho_{\text{lattice}}(-\mathbf{r})] \\
&= [\rho_{\text{cell}}(\mathbf{r}) * \rho_{\text{cell}}(-\mathbf{r})] * [\rho_{\text{lattice}}(\mathbf{r}) * \rho_{\text{lattice}}(-\mathbf{r})] \\
&= P_{\text{cell}}(\mathbf{r}) * P_{\text{lattice}}(\mathbf{r})
\end{aligned} \tag{3.11}
$$

by the commutativity and associativity of convolution (which are not that difficult to prove).

The Patterson function (autocorrelation) of the lattice of delta functions is

$$P_{\text{lattice}}(\mathbf{u}) = \sum_{n_a=-N_a}^{N_a} \sum_{n_b=-N_b}^{N_b} \sum_{n_c=-N_c}^{N_c} (N_a - |n_a|)(N_b - |n_b|)(N_c - |n_c|)\delta(\mathbf{u} - n_a\mathbf{a} - n_b\mathbf{b} - n_c\mathbf{c}) \,. \quad (3.13)$$

Then by the convolution theorem, the diffracted intensity, which is the Fourier transform of the Patterson function, must be the product of the diffracted intensity of a unit cell $|F_{\text{cell}}(\mathbf{r})|^2$ with the Fourier transform of this Patterson function of the lattice, which is given by

$$I(\mathbf{S}) = |F_{\text{cell}}|^2 \sum_{n_a=-N_a}^{N_a} \sum_{n_b=-N_b}^{N_b} \sum_{n_c=-N_c}^{N_c} (N_a - |n_a|)(N_b - |n_b|)(N_c - |n_c|)\mathrm{e}^{-2\pi\mathrm{i}(n_a\mathbf{a}+n_b\mathbf{b}+n_c\mathbf{c})\cdot\mathbf{S}} \,. \quad (3.14)$$

We have arrived at the same result.

3.2 The Reciprocal Lattice

Although we can't measure the structure factor, we can still work it out assuming we know $\rho(\mathbf{r})$ to see how it looks like for a single crystal. Using the above expression of the electron density as a convolution (3.9), we can directly take its Fourier transform and get the structure factor

$$
\begin{aligned}
F(\mathbf{S}) &= F_{\text{cell}}(\mathbf{S})\mathcal{F}\left[\sum_{n_a=0}^{N_a-1} \sum_{n_b=0}^{N_b-1} \sum_{n_c=0}^{N_c-1} \delta(\mathbf{r} - n_a\mathbf{a} - n_b\mathbf{b} - n_c\mathbf{c})\right] \\
&= F_{\text{cell}}(\mathbf{S})\left[\sum_{n_a=0}^{N_a-1} \sum_{n_b=0}^{N_b-1} \sum_{n_c=0}^{N_c-1} \mathrm{e}^{-2\pi\mathrm{i}(n_a\mathbf{a}+n_b\mathbf{b}+n_c\mathbf{c})\cdot\mathbf{S}}\right] \\
&= F_{\text{cell}}(\mathbf{S}) \prod_{j\in\{a,b,c\}} \sum_{n_j=0}^{N_j-1} \mathrm{e}^{-2\pi\mathrm{i}n_j\mathbf{j}\cdot\mathbf{S}} \,.
\end{aligned}
\quad (3.15)
$$

We see a cute geometric series here! This allows to simplify it further
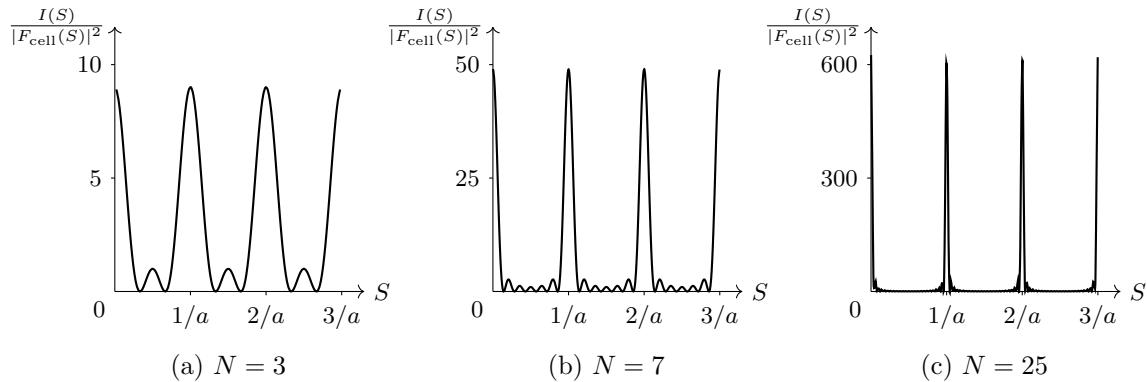
$$
\begin{aligned}
F(\mathbf{S}) &= F_{\text{cell}}(\mathbf{S}) \prod_{j\in\{a,b,c\}} \frac{1 - \exp[-2\pi\mathrm{i}N_j\mathbf{j}\cdot\mathbf{S}]}{1 - \exp[-2\pi\mathrm{i}\mathbf{j}\cdot\mathbf{S}]} \\
&= F_{\text{cell}}(\mathbf{S}) \prod_{j\in\{a,b,c\}} \frac{\exp[-\pi\mathrm{i}N_j\mathbf{j}\cdot\mathbf{S}]}{\exp[-\pi\mathrm{i}\mathbf{j}\cdot\mathbf{S}]} \frac{\exp[\pi\mathrm{i}N_j\mathbf{j}\cdot\mathbf{S}] - \exp[-\pi\mathrm{i}N_j\mathbf{j}\cdot\mathbf{S}]}{\exp[\pi\mathrm{i}\mathbf{j}\cdot\mathbf{S}] - \exp[-\pi\mathrm{i}\mathbf{j}\cdot\mathbf{S}]} \\
&= F_{\text{cell}}(\mathbf{S}) \prod_{j\in\{a,b,c\}} \exp[-\pi\mathrm{i}(N_j-1)\mathbf{j}\cdot\mathbf{S}]\frac{\sin(\pi N_j\mathbf{j}\cdot\mathbf{S})}{\sin(\pi\mathbf{j}\cdot\mathbf{S})}
\end{aligned}
\quad (3.16)
$$

Having the structure factor in this form, we can work out another expression of the diffraction intensity of single crystal

$$
\begin{aligned}
I(\mathbf{S}) &= F(\mathbf{S})F^*(\mathbf{S}) \\
&= |F_{\text{cell}}(\mathbf{S})|^2 \prod_{j\in\{a,b,c\}} \frac{\sin^2(\pi N_j\mathbf{j}\cdot\mathbf{S})}{\sin^2(\pi\mathbf{j}\cdot\mathbf{S})} \\
&= |F_{\text{cell}}(\mathbf{S})|^2 \frac{\sin^2(\pi N_a\mathbf{a}\cdot\mathbf{S})}{\sin^2(\pi\mathbf{a}\cdot\mathbf{S})} \frac{\sin^2(\pi N_b\mathbf{b}\cdot\mathbf{S})}{\sin^2(\pi\mathbf{b}\cdot\mathbf{S})} \frac{\sin^2(\pi N_c\mathbf{c}\cdot\mathbf{S})}{\sin^2(\pi\mathbf{c}\cdot\mathbf{S})} \,.
\end{aligned}
\quad (3.17)
$$

This is somewhat easier to plot. To make the plotting easier, we will consider a 1D crystal of N units cell, so

$$\frac{I(S)}{|F_{\text{cell}}(S)|^2} = \frac{\sin^2(\pi NaS)}{\sin^2(\pi aS)} \,. \quad (3.18)$$

18

(a) $N = 3$ (b) $N = 7$ (c) $N = 25$

Let's plot this function at different values of N and see how they look like.

We see that the functions have principal maxima at $S = k/a$ for integer values of k, and there are $N - 2$ subsidiary maxima between the principal maxima. As we increase N, the principal maxima becomes too strong that we can no longer see the subsidiary maxima. We can see this by taking the limit

$$\lim_{S \to k/a} \frac{\sin^2(\pi N a S)}{\sin^2(\pi a S)} = \lim_{u \to k} \left(\frac{\sin N u \pi}{\sin u \pi} \right)^2 = \lim_{u \to k} \left(\frac{N \pi \cos N u \pi}{\pi \cos u \pi} \right)^2 = N^2 \,, \tag{3.19}$$

so the height of the principal maxima is N^2. Hence, in the limit of large N (in real-life crystals, $N \sim 10^{23}$), we can well-approximate the diffraction pattern of single crystals as discrete beams, separated by $1/a$.

Similarly, in a 3D crystals, the diffraction intensity will only be observable at discrete values of $\mathbf{S}$ separated by $1/\|\mathbf{a}\|$, $1/\|\mathbf{b}\|$ and $1/\|\mathbf{c}\|$. This forms what is called a *reciprocal lattice*, with lattice vectors $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ given by

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{V} \qquad\qquad \mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{V} \qquad\qquad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{V} \,, \tag{3.20}$$

where $V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ is the volume of the unit cell. We will only observe diffracted beams with

$$\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \,, \tag{3.21}$$

i.e. on the reciprocal lattice points. We will show later that this corresponds to the same h, k, l as in the Miller planes in Bragg's description of diffraction which you might have seen if you take Part IA Material Sciences.

It is called the reciprocal lattice because it lives in the reciprocal space (Fourier domain). In fact, the reciprocal lattice is the Fourier transform of the real lattice. The unit of the reciprocal lattice vectors are length^{-1}, and their magnitudes are given by

$$\|\mathbf{a}^*\| = \frac{\|\mathbf{b}\| \, \|\mathbf{c}\| \sin \alpha}{V} \qquad \|\mathbf{b}^*\| = \frac{\|\mathbf{c}\| \, \|\mathbf{a}\| \sin \beta}{V} \qquad \|\mathbf{c}^*\| = \frac{\|\mathbf{a}\| \, \|\mathbf{b}\| \sin \gamma}{V} \,, \tag{3.22}$$

where α is the angle between $\mathbf{b}$ and $\mathbf{c}$, *etc.* This means that we can view the reciprocal lattice as "the lattice of periodicities in a crystal". To see this, let's consider all the possible set of repeated parallel planes that are compatible with the periodicity of the crystal. By being compatible with the periodicity, we require it to look identical in different unit cells, which requires the planes to intersect each edge of the lattice an integer number of times. This allows us to define the *Miller indices* for repeated planes in crystals by specifying how many times the plane intersect with the edge of the cell in each direction. If we put the first plane at the origin and the second plane intersects the axes at $1/h$, $1/k$ and $1/l$ for $h, k, l \in \mathbb{Z}$, then we would label these planes as (hkl). Then this set of planes will intersect $\mathbf{a}$ axis h times, $\mathbf{b}$ axis k times and $\mathbf{c}$ axis l times.
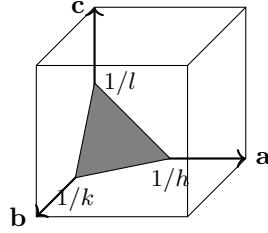
Figure 3.1: A (hkl) plane in a unit cell. There will be another $\max(h, k, l)$ planes parallel to this in this cell, intersecting **a** axis at $1/h, 2/h, \ldots, 1$, **b** axis at $1/k, 2/k, \ldots, 1$ and **c** axis at $1/l, 2/l, \ldots, 1$.
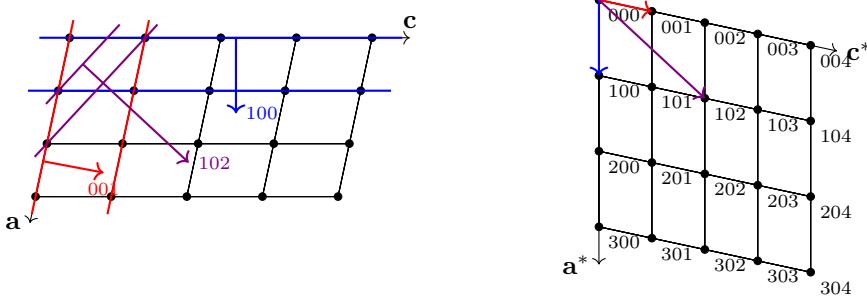

Figure 3.2: Constructing the reciprocal lattice using planes.

We can use these planes to construct the reciprocal lattice. Consider viewing the lattice along the **b** direction, and constructing planes of $(h0l)$ types. For each set of plane, if we draw a vector perpendicular to these planes and magnitude inversely proportional to the interplanar spacing, then the end of these vectors will form the $(h0l)$ section of the reciprocal lattice.

By the definition of the reciprocal lattice vectors, it is easy to see that

$$\mathbf{i} \cdot \mathbf{j}^* = \delta_{ij} \,, \tag{3.23}$$

$i, j \in \{a, b, c\}$. Hence, by expanding $\mathbf{r} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$ and $\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$, we can see that

$$\begin{aligned}
\mathbf{r} \cdot \mathbf{S} = {} & xh\mathbf{a} \cdot \mathbf{a}^* + xk\mathbf{a} \cdot \mathbf{b}^* + xl\mathbf{a} \cdot \mathbf{c}^* \\
& yh\mathbf{b} \cdot \mathbf{a}^* + yk\mathbf{b} \cdot \mathbf{b}^* + yl\mathbf{b} \cdot \mathbf{c}^* \\
& zh\mathbf{c} \cdot \mathbf{a}^* + zk\mathbf{c} \cdot \mathbf{b}^* + zl\mathbf{c} \cdot \mathbf{c}^* \\
= {} & xh + yk + zl \,.
\end{aligned} \tag{3.24}$$

Therefore for a single crystal, the structure factor can be rewritten as

$$F(h, k, l) = \int \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, \rho(x, y, z) \mathrm{e}^{2\pi \mathrm{i}(hx + ky + lz)} \,. \tag{3.25}$$

3.3 Bragg's Law

Here we present an alternative view of diffraction in a single crystal, which might be familiar from Part IA Material Sciences. We can think of the diffracted beam as being reflected by the planes in crystals, which are known as the Bragg's planes. The scattering vector is bisecting the incoming and outgoing rays, so it is perpendicular to those imaginary mirror planes. Now for a constructive interference to occur, we need the path difference of light reflected by consecutive Bragg planes to be an integer multiple of the wavelength, so

$$n\lambda = 2d \sin \theta \,, \tag{3.26}$$
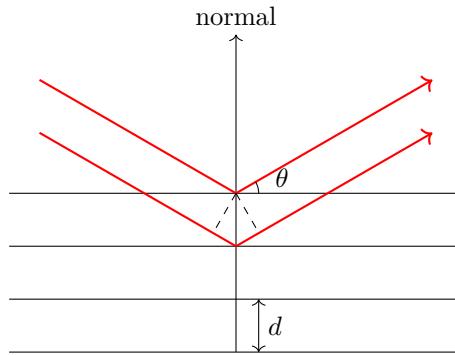
Figure 3.3: Diffraction using Bragg planes.

where θ is the angle between the planes and the beams, and d is the interplanar spacing. This factor of n can be omitted since diffraction from with $n > 1$ from a set of planes $(h\,k\,l)$ with spacing d is equivalent to diffraction with $n = 1$ from planes with spacing d/n, so we can instead think of them as being reflected from a high-order plane $(nh\,nk\,nl)$. For example, a $n = 4$ reflection from 100 planes is equivalent to a $n = 1$ reflection from 400 planes. Therefore, we have the Bragg's law

$$\lambda = 2d\sin\theta \tag{3.27}$$

for diffraction to be observed. We have defined $\|\mathbf{S}\| = 2\sin\theta/\lambda$, and so $\|\mathbf{S}\| = 1/d$. This again shows why diffraction is only observed on reciprocal lattice points.

3.4   Practical Measurements of Single-Crystal X-Ray Diffraction

Suppose we are measuring the diffraction from a particular single crystal. Since a crystal has a defined orientation, a certain scattering vector $\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ will appear in a well-defined direction relative to the crystal (perpendicular to the hkl planes). We can think the reciprocal lattice is attached to the crystal, and hence to measure a particular hkl diffraction, we need to arrange the incident beam to a particular direction relative to the crystal given by the Bragg's law. This requires us to be able to flexibly adjust the relative orientation of the incident beam, the detector and the crystal.
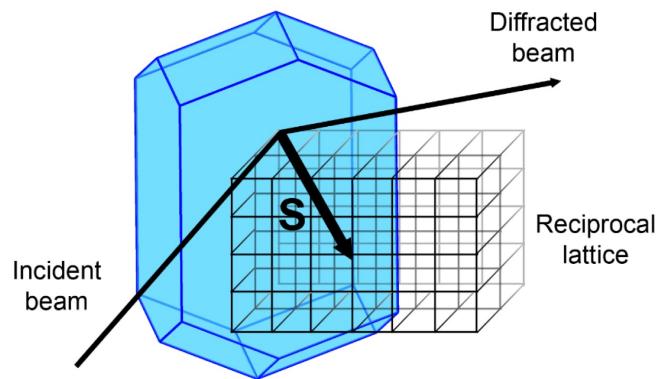

Figure 3.4: A schematic illustration of diffraction in a single crystal.

Practically, it is very difficult to move the X-ray source, so we have to move the crystal and the detector. If the source and the detector are making an angle 2θ, then this will correspond to a scattering vector of length $2\sin\theta/\lambda$ and direction bisecting the line from the crystal to the source
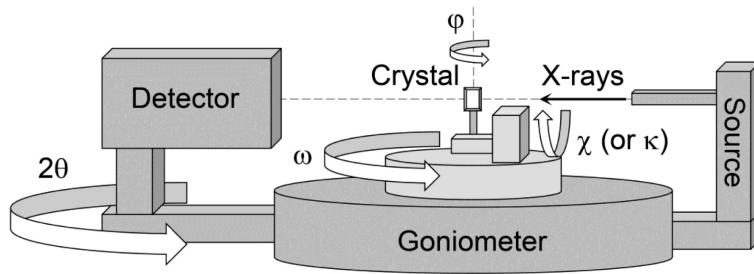
Figure 3.5: Apparatus for measuring the X-ray diffraction of a single crystal practically.

and to the detector. If this scattering vector happens to hit one of the reciprocal lattice point, then we will see a non-zero diffracted intensity in the detector. To make that happen, we have to be able to freely rotate the crystal around all 3 directions, so that we are effectively moving the reciprocal lattice attached to the crystal. This is achieved using a *goniometer.*

The outcome of a single-crystal X-ray diffraction measurement can be conceptually divided into two parts.

1. *Geometrical information.* From the value of $\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ we measured non-zero diffraction intensity, we can infer the reciprocal lattice vectors, and hence the real lattice vectors of the crystal. This can be done using only a small proportion of all the diffraction data, so only costs a few minutes.

2. *A list of diffracted intensity.* Once the geometry of the crystal in known, we can point the crystal and the detector only to where we know that there will be diffraction occurring. Then we can get a list of *hkl* values of measured diffracted beams and the corresponding diffraction intensities $|F(hkl)|^2$. To calculate the Patterson function (or the electron density if we somehow figure out the phases of the beams using methods that will be introduced later), we need the structure factor of the whole reciprocal space. This is not practical to do since there will be some maximum $\|\mathbf{S}\|$ value we are able to reach, and the diffraction will be weak for high *hkl* values. What we can do is to measure as many diffracted peaks as possible. This generally take hours, and hence X-ray diffraction is a slow technique, and the resulting structures are usually time-averaged.

3.5   The Independent Atom Model

Although $\rho(\mathbf{r})$ is a continuous function, we are generally only interested in the positions and types of atoms. We will see how can we make further progress if we introduce an atomic picture.

We will make the following assumptions.

1. The electron density is not continuous, but instead gathered into isolated, individual atoms at fixed sites. The atoms are spherical.

2. All electrons are located within atoms. Delocalised electrons or electrons in chemical bonds that are between atoms are neglected.

This formalism is known as the *independent atom model* (IAM). The second assumption seems like a fairly suspicious one. As chemists, we all know that chemical structures are held together by chemical bonds, and covalent bonds definitely involve the sharing of atoms. However for most elements, the majority of electrons are core electrons, which are little involved in bonding and are very localised. An obvious exception is hydrogen — we will discuss more on hydrogen later.

Now, if we describe the electron density of an atom labelled n as a spherically symmetric function $\rho_{\text{atom},n}(\mathbf{r})$, then the total electron density in the crystal is

$$\rho(\mathbf{r}) = \sum_n \rho_{\text{atom},n}(\mathbf{r}) * \delta(\mathbf{r} - x_n\mathbf{a} - y_n\mathbf{b} - z_n\mathbf{c}), \tag{3.28}$$

where (x_n, y_n, z_n) is the fractional coordinate of atom n and the sum is taken over all atoms in the crystal. Now from the convolution theorem, the structure factor of the cell is

$$
\begin{aligned}
F(\mathbf{S}) &= \mathcal{F}[\rho(\mathbf{r})]\\
&= \sum_n \mathcal{F}[\rho_{\text{atom},n}(\mathbf{r})]\mathcal{F}[\delta(\mathbf{r} - x_n\mathbf{a} - y_n\mathbf{b} - z_n\mathbf{c})]\\
&=: \sum_n f_n(\mathbf{S}) \exp(2\pi i \mathbf{r}_n \cdot \mathbf{S})\\
&= \sum_n f_n(\mathbf{S}) \exp(2\pi i (hx_n + ky_n + lz_n)),
\end{aligned}
\tag{3.29}
$$

where the f_n is the Fourier transform of $\rho_{\text{atom},n}(\mathbf{r})$ called the *atomic scattering factor*. It can be seen as the diffraction of a single atom, and has unit of electrons. Since we assume $\rho_{\text{atom},n}(\mathbf{r})$ is spherically symmetric, $f(\mathbf{S})$ is spherically symmetric as well. Although some other more complicated model may apply (e.g. multipole expansion) to make both $\rho_{\text{atom},n}(\mathbf{r})$ and $f(\mathbf{S})$ not spherically symmetric, we shall not concern with these.

3.5.1 Atomic Scattering Factor

For the special case of $\mathbf{S} = \mathbf{0}$,

$$F(\mathbf{0}) = \sum f_n(\mathbf{0}) = \int \mathrm{d}V\, \rho(\mathbf{r}). \tag{3.30}$$

For this to hold, we must have $f_n(\mathbf{0})$ equal to the total number of electrons in atom n. As $\mathbf{S}$ moves away from zero, the value of f_n must change to reflect the interference occurring between X-rays scattered from different positions within the atom's volume. Under the assumption that $f(\mathbf{S})$ is spherically symmetric, it is only dependent on $\|\mathbf{S}\| = (2\sin\theta)/\lambda$, so the scattering factor is usually fitted to parameterised function of $(\sin\theta)/\lambda$ as

$$f\left(\frac{\sin\theta}{\lambda}\right) = c + \sum_{i=1}^{4} a_i \exp\left[-b_i\left(\frac{\sin\theta}{\lambda}\right)^2\right] \tag{3.31}$$

with nine tabulated parameters for each atom. These parameters are known as Cromer–Mann coefficients.

Some examples of spherically symmetric atomic scattering factors are shown in figure 3.6. We can see that $f(0)$ is equal to the number of electrons in the atom/ion. Heavier elements generally have larger atomic scattering factors. For species having the same number of electrons, if the atom/ion is larger, than the scattering factors drop off more quickly. This is because a bigger atom would induce a larger path difference in diffraction, and hence leads to lower f_n.[11]

These factors are derived quantum mechanically from atomic wavefunctions, which are effectively at 0 K. In practice, atoms in crystals are typically slightly displaced from their average position — this may be dynamic (due to thermal vibration) or static. This effectively makes the atom look

---

[11]You can understand this by modelling the electron density in an atom as a Gaussian of width σ. Its Fourier transform (the scattering factor) will be a Gaussian of width σ^{-1}. Hence, if the atom is larger, then the scattering factor has smaller width. If we have a "point" atom, whose electron density is a delta function, then its Fourier transform is a constant, so the scattering factor will not drop off.

(a) Scattering factors of different species with different numbers of electrons.

(b) Scattering factors of different species with the same number of electrons.
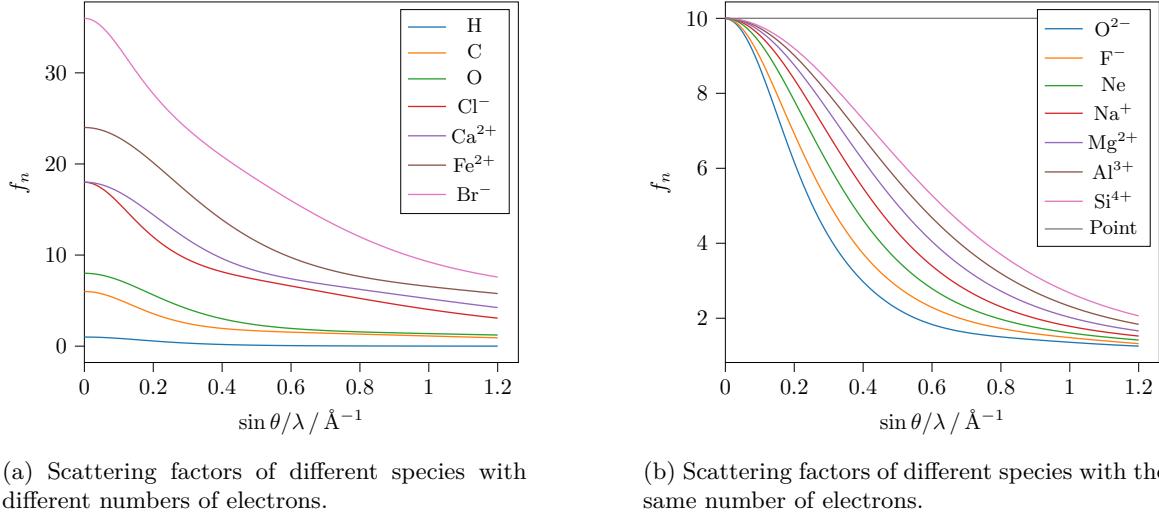
Figure 3.6: Atomic scattering factors of different species.
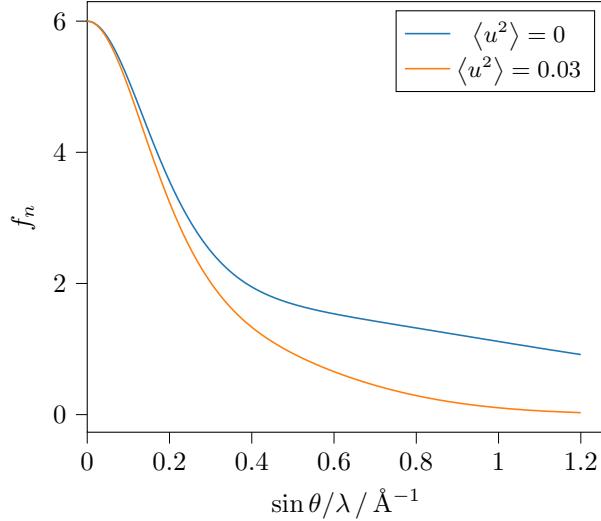
Figure 3.7: The atomic scattering factor of C with and without isotropic displacement.

larger, and hence the atomic scattering factor will drop off more quickly. This will put an extra *Debye–Waller factor*

$$\exp\left(-\frac{8\pi^2 \langle u^2 \rangle \sin^2 \theta}{\lambda^2}\right) \tag{3.32}$$

onto the atomic scattering factor at 0 K.

To conclude, under the independent atomic model, the structure factor of an atomic single crystal is

$$F(h, k, l) = \sum_n f_n \exp[2\pi i(hx_n + ky_n + lz_n)], \tag{3.33}$$

where n is taken over all atoms in the crystals. However, we have seen that $|F_{hkl}|^2 = N^2 |F_{\text{cell}}(hkl)|^2$, so the relative intensities of the diffracted beams are proportional to $|F_{\text{cell}}(hkl)|^2$. To solve for this, we only need to sum over all atoms in a unit cell

$$F_{\text{cell}}(h, k, l) = \sum_n f_n \exp[2\pi i(hx_n + ky_n + lz_n)]. \tag{3.34}$$

From now on we will drop the "cell" subscript, and we will always refer to F_{cell}. We see that this has now become a Fourier series. To figure out the electron density, we apply the inverse Fourier transform and we get

$$\rho(x, y, z) = \frac{1}{V_{\text{cell}}} \sum_{h} \sum_{k} \sum_{l} F(h, k, l) \exp[-2\pi\mathrm{i}(hx + ky + lz)] \,, \tag{3.35}$$

where $1/V_{\text{cell}}$ is volume of the unit cell for normalisation. This is also known as a *Fourier synthesis*, because it 'synthesises' the electron density by its frequency components using a complex Fourier series.

4 Symmetry in Crystals

In the previous chapter, we have already exploited the translation symmetry of crystals, which is their defining property. Typically, crystals also display other types of symmetries, such as rotations, mirror planes, and inversions (although they need not). This leads us to something that looks like the *point groups* that describe the symmetry of molecules, but in crystals, we have additional translational symmetry, so they are instead known as *space groups*. Rather remarkably, there are only 230 possible 3D space groups — the number is finite, unlike for point groups where there can be an infinite number of them. This is because not all symmetry elements are compatible with the translational symmetry, leaving us with a limited possible combination of symmetry elements. The most famous example is probably that a crystal can never have a five-fold rotation axis[12].
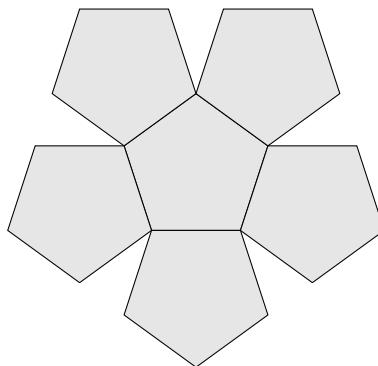

Figure 4.1: Five-fold rotational symmetry is not compatible with translational symmetry.

Let's first see how various symmetry operators affect the diffraction patterns.

4.1 Symmetry Operations

4.1.1 Inversion Centre

Suppose there is an inversion centre in the crystal, then by translational symmetry there must be one in each unit cell. We set it to be the origin of the unit cell, then for each atom at (x_n, y_n, z_n), there must be the same atom at $(-x_n, -y_n, -z_n)$. Therefore, when calculating the structure factor, instead of summing over all N atoms in the unit cell, we can sum over the $N/2$ pairs:

$$
\begin{aligned}
F(hkl) &= \sum_{n=1}^{N/2} f_n \left[\exp[2\pi i(hx_n + ky_n + lz_n)] + \exp[-2\pi i(hx_n + ky_n + lz_n)] \right] \\
&= \sum_{n=1}^{N/2} 2f_n \cos[2\pi(hx_n + ky_n + lz_n)] . \\
&= \sum_{n=1}^{N} f_n \cos[2\pi(hx_n + ky_n + lz_n)] .
\end{aligned}
\tag{4.1}
$$

Two immediate consequences are

- The structure factor is always real (phase is 0 or π).

- $F(hkl) = F(\bar{h}\bar{k}\bar{l})$.

---

[12]Something known as a *quasi-crystal* can, but it has no translational symmetry.

Therefore, the structure factor also shows an inversion symmetry, with point group symmetry (at least) C_i.

However, this is not something we can really exploit. This is because we can only measure $I(hkl) = |F(hkl)|^2$, or equivalently the modulus of the structure factor.

Theorem 4.1 (Friedel's law). For any crystal,

$$|F(hkl)| = |F(\bar{h}\bar{k}\bar{l})| \,. \tag{4.2}$$

Proof. This is because

$$
\begin{aligned}
F(\bar{h}\bar{k}\bar{l}) &= \sum_n f_n \exp[-2\pi\mathrm{i}(hx_n + ky_n + lz_n)] \\
&= \left[\sum_n f_n \exp[2\pi\mathrm{i}(hx_n + ky_n + lz_n)]\right]^* \\
&= F(hkl)^* \,.
\end{aligned}
\tag{4.3}
$$

They are complex conjugate, so they have the same modulus and diffraction intensity. $\square$

Therefore, the diffraction intensities of all crystals exhibit centres of inversion.

In practice, when describing a diffraction pattern, we often picture a reciprocal lattice with its lattice points scaled by the intensity of diffraction. This construction is called the *intensity-weighted reciprocal lattice*. The intensity-weighted reciprocal lattices of all crystals show a C_i point symmetry.

Aside: Intensity Statistics

However, it is possible to infer the presence of an inversion centre from the diffraction pattern based on the intensities. If a centre of symmetry is not present, then generally $F(hkl)$ would be a complex number, while if there is a centre of symmetry present, $F(hkl)$ can only be a real number. A real $F(hkl)$ has a higher probability to have a modulus close to zero than a complex one, and so we have a higher chance to observe weak $F(hkl)$ values if an inversion centre is present.

The above argument sounds a bit dubious. We can formalise it a little bit. Suppose the f_n values in a crystal follow some distribution with mean μ and variance σ^2, and the phase of each atom ϕ_n is somehow random (uniformly distributed) between 0 and 2π.

If there is a centre of inversion, then

$$F = \sum_n f_n \cos \phi_n \,. \tag{4.4}$$

$\cos \phi_n$ has a mean 0 and variance $\frac{1}{2}$. By central limit theorem, if N is large, then F is a Gaussian with expectation value of

$$\mathbb{E}[F] = N\mathbb{E}[f_n \cos \phi_n] = N\mathbb{E}[f_n]\mathbb{E}[\cos \phi_n] = 0 \tag{4.5}$$

and variance

$$
\begin{aligned}
\mathrm{Var}[F] &= N\,\mathrm{Var}[f_n \cos \phi_n] \\
&= N\left[(\mathrm{Var}[f_n] + \mathbb{E}[f_n]^2)(\mathrm{Var}[\cos \phi_n] + \mathbb{E}[\cos \phi_n]^2) - \mathbb{E}[f_n]^2\mathbb{E}[\cos \phi_n]^2\right] \\
&= \frac{N}{2}(\sigma^2 + \mu^2) \,.
\end{aligned}
\tag{4.6}
$$

Let's denote this variance σ_F^2, then for the centrosymmetric case, the probability distribution function of $|F|$ is

$$g(x) = \frac{2}{\sigma_F \sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma_F^2}\right], \qquad x \geq 0. \tag{4.7}$$
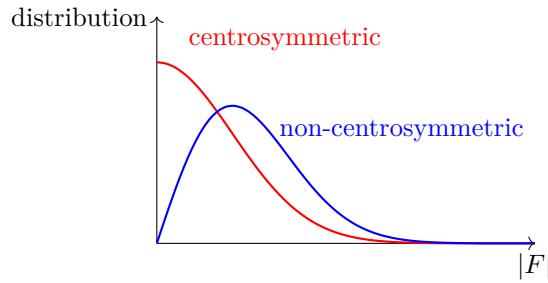
If there is not a symmetry of inversion, however, then

$$F = \sum_n f_n (\cos\phi_n + \mathrm{i}\sin\phi_n). \tag{4.8}$$

The real and imaginary part of F then follows the same Gaussian distribution as the centrosymmetric case, with mean 0 and variance σ_F^2. However, this is a 2D Gaussian. The probability density function of $|F|$ is then

$$
\begin{aligned}
h(x) &= \int_0^{2\pi} \mathrm{d}\theta\, x \frac{1}{2\pi\sigma_F^2} \exp\left[-\frac{x^2}{2\sigma_F^2}\right] \\
&= \frac{x}{\sigma_F^2} \exp\left[-\frac{x^2}{2\sigma_F^2}\right].
\end{aligned}
\tag{4.9}
$$

Its probability density at 0 vanishes.

4.1.2 Mirror Plane

Consider a mirror plane perpendicular to the **b** axis passing through the origin. This relates an atom at (x, y, z) to another at $(x, -y, z)$. The structure factor can be simplified to

$$
\begin{aligned}
F(hkl) &= \sum_{n=1}^{N/2} f_n [\exp[2\pi\mathrm{i}(hx_n + ky_n + lz_n)] + \exp[2\pi\mathrm{i}(hx_n - ky_n + lz_n)]] \\
&= \sum_{n=1}^{N} f_n \cos(2\pi ky_n) \exp[2\pi\mathrm{i}(hx_n + lz_n)].
\end{aligned}
\tag{4.10}
$$

Cosine is an even function, so

$$F(hkl) = F(h\bar{k}l), \tag{4.11}$$

and $|F(hkl)|^2 = |F(h\bar{k}l)|^2$. Therefore intensity-weighted reciprocal lattice will also have a mirror plane perpendicular to the $\mathbf{b}^*$ axis. Combined with the i symmetry of Friedel pairs, we have

$$|F(hkl)|^2 = |F(h\bar{k}l)|^2 = |F(\bar{h}k\bar{l})|^2 = |F(\bar{h}\bar{k}\bar{l})|^2, \tag{4.12}$$

and so the intensity-weighted reciprocal lattice will have a C_{2h} symmetry. Note that i and σ_h generate the C_2 symmetry.
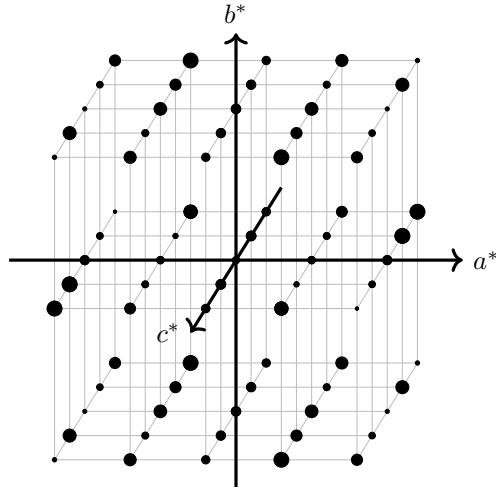
Figure 4.2: An intensity-weighted reciprocal lattice with C_{2h} symmetry.

4.1.3 Rotation Axes

Now let's consider the effect of a rotation axis. A m-fold rotation axis will rotate atoms by $2\pi/m$ radians around the defined axis. For simplicity, we will only consider the 2-fold axis. We set the axis such that the rotation axis passes through the origin and is parallel to the **b** axis of the unit cell, so that for each atom at (x_n, y_n, z_n), there is also a same atom at $(-x, y, -z)$. In this case, we may simplify the structure factor to

$$F(hkl) = \sum_n^{N/2} f_n[\exp[2\pi i(hx_n + ky_n + lz_n)] + \exp[2\pi i(-hx_n + ky_n - lz_n)]]$$

$$= \sum_n^N f_n \cos[2\pi(hx_n + lz_n)] \exp[2\pi i ky_n]. \tag{4.13}$$

Therefore,

$$F(hkl) = F(\bar{h}k\bar{l}). \tag{4.14}$$

The intensity-weighted reciprocal lattice therefore also has a 2-fold rotation axis along $\mathbf{b}^*$. Again, the Friedel's law still applies, so

$$|F(hkl)|^2 = \left|F(h\bar{k}l)\right|^2 = \left|F(\bar{h}k\bar{l})\right|^2 = \left|F(\bar{h}\bar{k}\bar{l})\right|^2. \tag{4.15}$$

Therefore the intensity-weighted reciprocal lattice has a C_{2h} point group symmetry (formally known as the *Laue group*), as i and C_2 further generates a σ_h symmetry.

Finally, note that if a crystal structure exhibits a 2-fold rotation axis along one direction, or a mirror plane perpendicular to some direction, this axis must be perpendicular to two of the lattice vectors, otherwise the rotated cell will have a different orientation, and the translational symmetry will be broken. For example, if a 2-fold axis is parallel to **b**, then **b** must be perpendicular to both **a** and **c**. This means that the structure is at least *monoclinic*, meaning that two of the unit-cell angles must be 90°. By convention, we usually choose $\alpha = \gamma = 90°$ and $\beta \neq 90°$ (although it could be that $\beta = 90°$, but then we will classify the crystal to have a higher symmetry e.g. *orthorhombic*).

4.1.4 Screw Axes

The fact that crystals are infinitely extended through the space leads us to some extra symmetry elements that are not present in point groups. Specifically, we can combine rotations or reflections with translations.

A n_m-screw axis stands for a rotation of $2\pi/n$ around the axis, and then a translation of m/n times the length of the lattice vector in that direction. We will only consider the 2-fold screw axis 2_1, and we will make the axis coincide with the **b** vector. This is a rotation of $180°$ along **b**, and then translate by $\mathbf{b}/2$, relating (x, y, z) to $(-x, \frac{1}{2} + y, -z)$. The resulting structure factor is

$$F(hkl) = \sum_{n=1}^{N/2} f_n \exp(2\pi \mathrm{i}ky_n)[\exp[2\pi \mathrm{i}(hx_n + lz_n)] + (-1)^k \exp[-2\pi \mathrm{i}(hx_n + lz_n)]]$$

$$= \sum_{n=1}^{N} f_n \exp(2\pi \mathrm{i}ky_n) \times \begin{cases} \cos[2\pi(hx_n + lz_n)] & \text{if } k \text{ is even} \\ \mathrm{i}\sin[2\pi(hx_n + lz_n)] & \text{if } k \text{ is odd.} \end{cases} \tag{4.16}$$

Consequently,

$$F(hkl) = \begin{cases} F(\bar{h}k\bar{l}) & \text{if } k \text{ is even} \\ -F(\bar{h}k\bar{l}) & \text{if } k \text{ is odd,} \end{cases} \tag{4.17}$$

but in either case,

$$\left|F(hkl)\right|^2 = \left|F(\bar{h}k\bar{l})\right|^2 . \tag{4.18}$$

Combining with the Friedel pair relationship, we see that the 2_1 screw axis has exactly the same effect on the symmetry of the diffracted intensity as the 2-fold rotation axis, both having C_{2h}.

Moreover, if we consider the special case of $h = l = 0$, then the structure factor for $0k0$ reduces to

$$F(0k0) = \sum_{n}^{N/2} f_n \exp(2\pi \mathrm{i}ky_n)[1 + (-1)^k] = \begin{cases} \sum_{n}^{N} f_n \exp(2\pi \mathrm{i}ky_n) & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd.} \end{cases} \tag{4.19}$$

The structure factors for all $0k0$ beams are zero if k is odd — this result is independent of the atomic coordinates and is purely originates from symmetry. This is known as a *systematic absence*. The pattern of the systematic absence depends on the direction of the screw axis:

- 2_1 axis parallel to **a**: $(h00)$ absent for odd h.

- 2_1 axis parallel to **b**: $(0k0)$ absent for odd k.

- 2_1 axis parallel to **c**: $(00l)$ absent for odd l.

We have a good explanation for this. Consider a 2_1 axis along **b**. If we project the 3D crystal structure into one dimension along **b**, i.e. we ignore the x and z coordinates, the presence of a 2_1 screw axis doubles the periodicity along **b** direction, and hence the periodicity in the reciprocal lattice will be halved. This is exactly what a $0k0$ beam is detecting: as reflections from a $(0k0)$ Bragg's plane has contributions from all x and z coordinates for a certain y value, it is only able to tell the periodicity along y direction, and it finds that things are repeating twice as often.

4.1.5 Glide Planes

A glide plane is a mirror plane reflection followed by a translation one half of the lattice. We will consider a c-glide perpendicular to b, which is a reflection about a mirror plane perpendicular to b followed by a translation of $\frac{1}{2}\mathbf{c}$, relating atoms at (x, y, z) and $(x, -y, \frac{1}{2} + z)$. The structure factor equation becomes

$$F(hkl) = \sum_{n}^{N/2} f_n \exp[2\pi \mathrm{i}(hx_n + lz_n)][\exp(2\pi \mathrm{i}ky_n) + (-1)^l \exp(-2\pi \mathrm{i}ky_n)]$$

$$= \sum^{N} f_n \exp[2\pi \mathrm{i}(hx_n + lz_n)] \times \begin{cases} \cos(2\pi ky_n) & \text{if } l \text{ is even} \\ \mathrm{i}\sin(2\pi ky_n) & \text{if } l \text{ is odd.} \end{cases} \tag{4.20}$$

This means that

$$F(hkl) = \begin{cases} F(h\bar{k}l) & \text{if } l \text{ is even} \\ -F(h\bar{k}l) & \text{if } l \text{ is odd,} \end{cases} \tag{4.21}$$

but in either case,

$$|F(hkl)|^2 = |F(h\bar{k}l)|^2 . \tag{4.22}$$

The glide plane produces a mirror plane in the diffraction pattern, just as a normal mirror plane would do, and combining with Friedel pair relation, the diffraction pattern again has a C_{2h} symmetry.

Consider $k = 0$, the structure factor reduces to

$$F(h0l) = \begin{cases} \sum_n^N f_n \exp[2\pi i(hx_n + lz_n)] & \text{if } l \text{ is even} \\ 0 & \text{if } l \text{ is odd.} \end{cases} \tag{4.23}$$

Therefore, a c-glide perpendicular to b produces systematic absences of $F(h0l)$ for odd l. We again have a good explanation for it: we set k to 0 so the beam is ignorant of us changing the y coordinate by reflection, and now the translation of $\frac{1}{2}\mathbf{c}$ doubles the periodicity along the c direction, so the reciprocal lattice is doubled in length in that direction.

To conclude the systematic absence for glide planes:
Glide plane perpendicular to a:

- b-glide: $(0kl)$ absent if k odd.

- c-glide: $(0kl)$ absent if l odd.

- n-glide: $(0kl)$ absent if $k+l$ odd. (*n*-glide means translation by a half in both y and z directions)

Glide plane perpendicular to b:

- a-glide: $(h0l)$ absent if h odd.

- c-glide: $(h0l)$ absent if l odd.

- n-glide: $(h0l)$ absent if $h + l$ odd.

Glide plane perpendicular to c:

- a-glide: $(hk0)$ absent if h odd.

- b-glide: $(hk0)$ absent if k odd.

- n-glide: $(hk0)$ absent if $h + k$ odd.

4.2 Lattice Types

In some cases, it is more preferable to define a unit cell that is not primitive, i.e. containing more than one lattice point. This is possibly because the non-primitive unit cell has a higher symmetry, or it allows the symmetry elements to align with the axes of the unit cell.

There are in total four types of 3D unit cells. The first type is *primitive*, denoted P, where the lattice points are at the corners of the cells only, so that there is only one lattice point in the unit cell. It is also possible for a cell to also contain lattice points in the centre of two opposite faces. This is conventionally chosen to be the ab face, so this type of cell is known as *C-centred*. Another possibility is to contain a second lattice point at the centre of the cell, and this is known as *body-centred* or *I-centred*. Finally, we can have lattice points at the centres of all unit-cell faces, which is known as *face-centred* or *F-centred*.
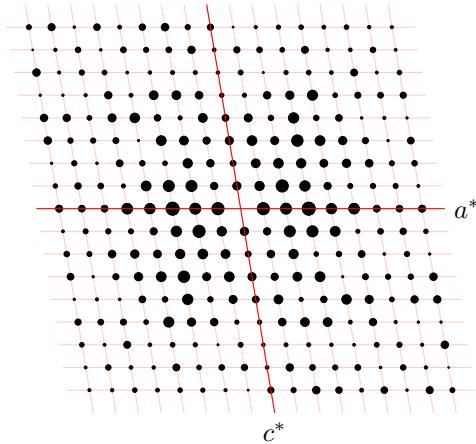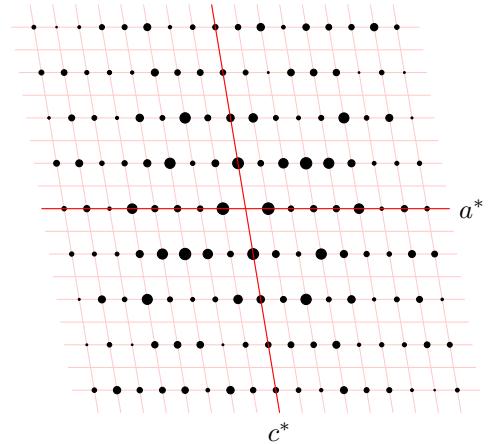
(a) mirror plane perpendicular to *b*.


(b) *c*-glide perpendicular to *b*.

Figure 4.3: Schematic (*h0l*) slices of the intensity-weighted reciprocal lattice. We see that glide planes create systematic absences, while mirror planes do not.
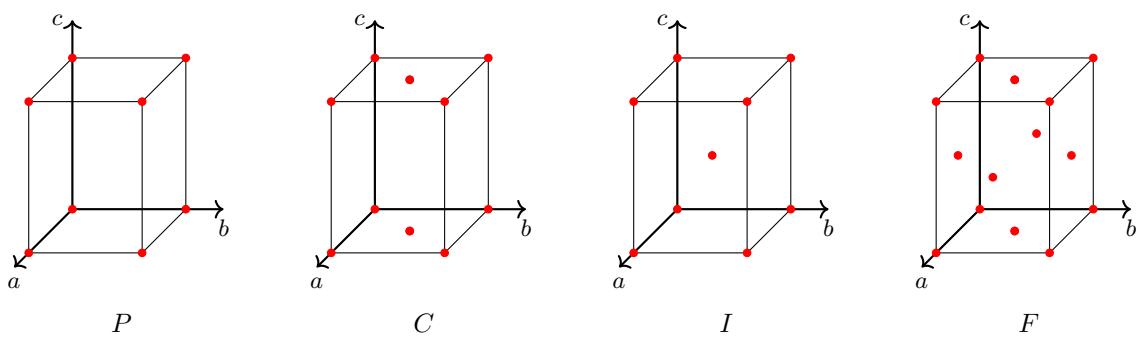

Figure 4.4: Types of unit cell centring.

Since non-primitive unit cells contain more than one lattice points, they also have systematic absences, which can be worked out using the same way as we did before.

- P: No systematic absences.

- C: (hkl) absent if $h + k = 2n + 1$.

- I: (hkl) absent if $h + k + l = 2n + 1$.

- F: (hkl) absent unless h, k, l are all odd or all even.

Heuristically, the reciprocal lattice of a P and a C lattices are themselves, the reciprocal lattice of an I lattice is a larger F lattice, and the reciprocal lattice of an F lattice is a larger I lattice.

By using a non-primitive cell, we are choosing to describe the real lattice with longer lattice vectors, so the reciprocal lattice vectors are correspondingly shorter. We are therefore using smaller reciprocal unit cells than a primitive reciprocal unit cell. Hence, systematic absences must arise.

4.3   Space Groups and Equivalent Positions

An identification of the lattice type together with a collection of symmetry operators defines a *space group*. It describes the symmetry of a crystal, just as a point group describes the symmetry of a molecule, but we have an extra translational symmetry described by the lattice type.

For example, let's consider the most common space group $P2_1/c$, where its name stands for primitive lattice, a 2_1 screw axis $\parallel b$ and a c-glide $\perp b$. These symmetry elements also create a centre of inversion. By convention, we set the inversion centre to be the origin, so that the 2_1 axis is at $(0, y, 1/4)$ and the c-glide is at $(x, 1/4, z)$. This collection of symmetry operators produces a set of coordinates that are related to each other, known as the *general equivalent positions* (GEPs). The GEPs for $P2_1/c$ are shown in the figure below.
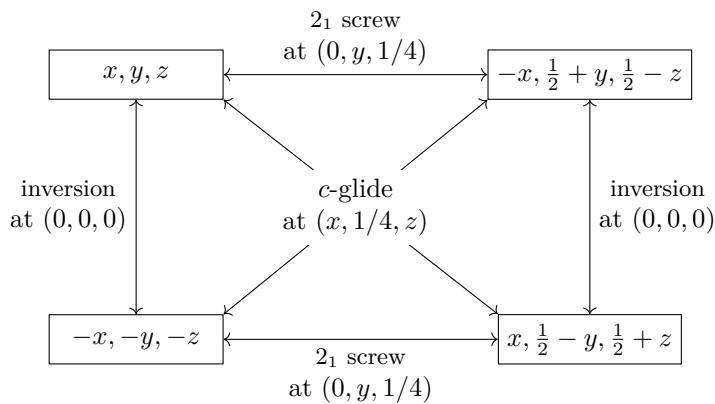

Figure 4.5: General equivalent positions in the $P2_1/c$ space group.

Therefore, there exists a minimal set of atomic coordinates that need to be specified to describe the whole unit cell (and hence the entire crystal structure), which is known as the *asymmetric unit*. The coordinates of the other atoms in the cell can be obtained by applying the symmetry elements in the space group.

However, notice that there are some special coordinates in the $P2_1/c$ group that do not produce a group of four distinct coordinates when the symmetry operators are applied. An example is $(0, 0, 0)$, for which you will only obtain two distinct coordinates $(0, 0, 0)$ and $(0, 1/2, 1/2)$ when applying the symmetry elements. Such positions are called the *special equivalent positions*. They arise because the coordinates coincide with the point group elements (centre of symmetry in this case), so it is not
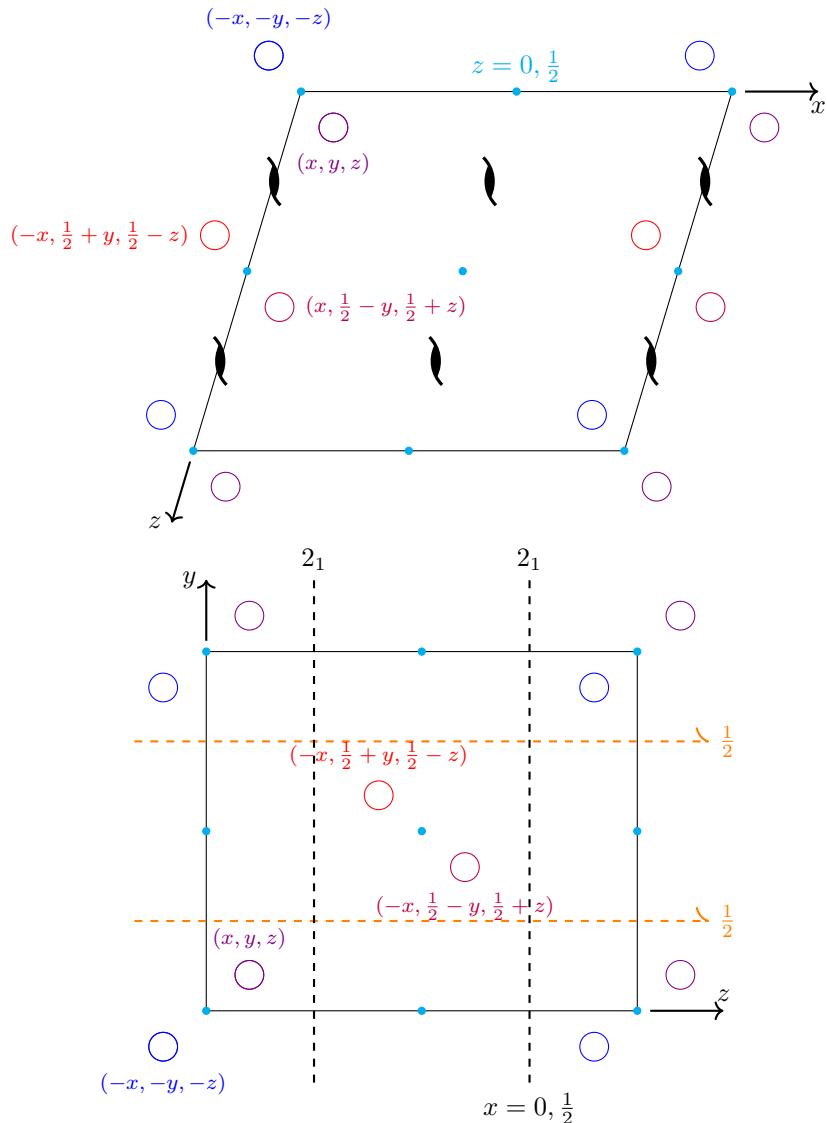
Figure 4.6: View of $P2_1/c$ space group unit cell along y axis and x axis. Note that extra symmetry elements arise automatically such that they appear twice as often as the lattice points.
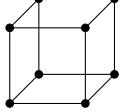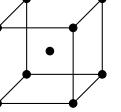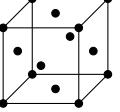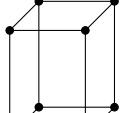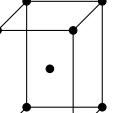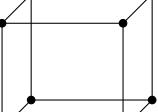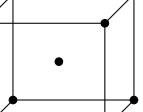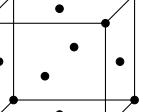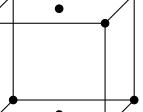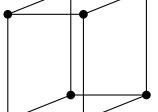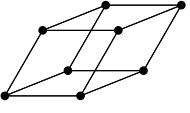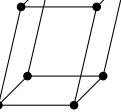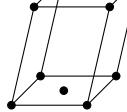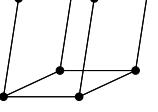
moved by those symmetry elements. Similarly, if a rotation axis or a mirror plane is present in some space group, there will be a whole line and a whole plane of special equivalent positions, respectively.

4.4 Determining Space Group

Having obtained a diffraction pattern from the experiment, we should be able to easily determine the space group. First, by observing the reciprocal cell geometry, we should be able to infer the lattice vectors a, b, c and their angles α, β and γ. Then by observing the general (hkl) systematic absences, we can infer the lattice type (P, I, F or C). This will give us one of the 14 possible types of 3D lattice, known as the *Bravais lattices*.

Having determined the lattice type, we can then identify the symmetry elements in the crystal. If the intensity-weighted reciprocal lattice exhibits a C_{2h} symmetry, then there will be a C_2 axis or a mirror plane in the structure. If there are other systematic absences, then there will be glide planes

Table 1: The 14 Bravais lattices.

Crystal system	*P*	*I*	*F*	*C*
Cubic $a = b = c$ $\alpha = \beta = \gamma = 90°$				
Tetragonal $a = b \neq c$ $\alpha = \beta = \gamma = 90°$				
Orthorhombic $a \neq b \neq c$ $\alpha = \beta = \gamma = 90°$				
Hexagonal $a = b \neq c$ $\alpha = \beta = 90°, \gamma = 120°$				
Trigonal $a = b = c$ $\alpha = \beta = \gamma \neq 90°$				
Monoclinic $a \neq b \neq c$ $\alpha = \gamma = 90°, \beta \neq 90°$				
Triclinic $a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90°$				

or screw axes. We can identify the presence of inversion centres from intensity statistics.

In practice, this can be easily done using automated programs. However, for a genuine experiment, there will be measurement errors, or there might be ambiguity on whether a diffraction is too small or it is really absent. Hence, all of these assessment indicate probable space groups rather than a concrete assessment.

4.5 Symmetry in Patterson Functions

Symmetry also allows us to identify atomic coordinates directly from Patterson function. Suppose we have confirmed that a crystal has a space group $P2_1/c$ from diffraction pattern, we then know that

the crystal must have the same atom at the following general equivalent positions:

$$(a): (x, y, z) \qquad (b): (-x, \tfrac{1}{2} + y, \tfrac{1}{2} - z) \qquad (c): (x, \tfrac{1}{2} - y, \tfrac{1}{2} + z) \qquad (d): (-x, -y, -z).$$

This results in 16 interatomic vectors in the Patterson function: four of them are self-vectors at the origin, and the remaining twelve are non-origin vectors. Since each $A \to B$ vector will be accompanied by a $A \leftarrow B$ in reverse, these will comprise six centrosymmetric pairs:

(i) $(a) \leftrightarrow (b)$: $\pm(2x, -\tfrac{1}{2}, -\tfrac{1}{2} + 2z)$,

(ii) $(a) \leftrightarrow (c)$: $\pm(0, -\tfrac{1}{2} + 2y, -\tfrac{1}{2})$,

(iii) $(a) \leftrightarrow (d)$: $\pm(2x, 2y, 2z)$,

(iv) $(b) \leftrightarrow (c)$: $\pm(-2x, 2y, -2z)$,

(v) $(b) \leftrightarrow (d)$: $\pm(0, \tfrac{1}{2} + 2y, \tfrac{1}{2})$,

(vi) $(c) \leftrightarrow (d)$: $\pm(2x, \tfrac{1}{2}, \tfrac{1}{2} + 2z)$.

As the structure is periodic, there will be the same atom 1 unit away along any coordinate, so we can shift any coordinate by ± 1. Consequently, vector (i) is actually identical to (vi), and (ii) is identical to (v). Therefore, for the space group $P2_1/c$, we actually have two pairs of doubly-weighted peaks at $\pm(2x, \tfrac{1}{2}, \tfrac{1}{2} + 2z)$, $\pm(0, \tfrac{1}{2} + 2y, \tfrac{1}{2})$ and two pairs of singly-weighted peaks at $\pm(2x, 2y, 2z)$ and $\pm(-2x, 2y, -2z)$. Those peaks in the Patterson functions are easy to locate because of their special forms (0 or $\tfrac{1}{2}$ in some coordinates). By identifying the coordinates of these peaks in the Patterson function, we can directly work out the coordinates of these atoms.

Note that since the Patterson function only gives distinct peaks for interatomic vectors between heavy atoms, this usually only allow us to work out the coordinates of the heavy atoms in the structure. However, this is usually a very important first step for a complete structure determination. Having known this initial information, we have three complementary methods to determine the coordinates of the rest of the atoms. They are the *heavy atom method*, *isomorphous replacement* and *anomalous scattering*.

4.6 The Heavy Atom Method

Suppose of the N atoms in the unit cell, we know the coordinate of the M heavy atoms of them e.g. from Patterson function. We can split the structure factor into a contribution of heavy atoms and the contribution of the rest of the atoms.

$$
\begin{aligned}
F(hkl) &= \sum_{n=1}^{N} f_n \exp[2\pi i (hx_n + ky_n + lz_n)] \\
&= \sum_{n \text{ heavy}} f_n \exp[2\pi i (hx_n + ky_n + lz_n)] + \sum_{n \text{ rest}} f_n \exp[2\pi i (hx_n + ky_n + lz_n)] \\
&= F_{\text{heavy}}(hkl) + F_{\text{rest}}(hkl).
\end{aligned}
\tag{4.24}
$$

From experiment, we can only measure $|F(hkl)|$, not $\Phi(hkl)$. But since we know the coordinate of the M heavy atoms, we can calculate

$$
F_{\text{heavy}} = \sum_{n=1}^{M} f_n \exp[2\pi i (hx_n + ky_n + lz_n)].
\tag{4.25}
$$

In particular, we know both the amplitude and the phase of the heavy atom contribution. Since the diffracting power of an atom f_n, is proportional to the number of electrons, which is in turn proportional to the atomic number, heavy atom contributions tend to dominate in $F(hkl)$.

$$|F_{\text{heavy}}(hkl)| \overset{?}{\gg} |F_{\text{rest}}(hkl)|.$$
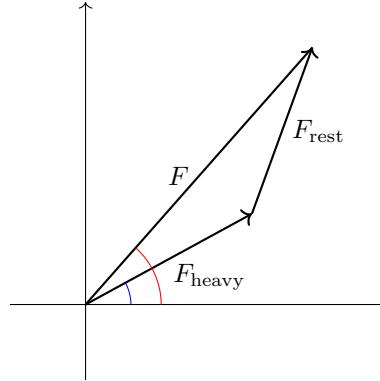
Figure 4.7: We can use Φ_{heavy} (blue) to approximate Φ (red).

This allows us to argue that hopefully $\Phi(hkl)$ is close to $\Phi_{\text{heavy}}(hkl)$, so we can use it as an estimate of $\Phi(hkl)$. This allows us to combine the measured $|F(hkl)|$ and calculated $F_{\text{heavy}}(hkl)$ to get an estimate of the structure factor

$$F(hkl) = |F(hkl)| \exp[\mathrm{i}\Phi(hkl)] \overset{?}{\approx} |F(hkl)| \exp[\mathrm{i}\Phi_{\text{heavy}}(hkl)]. \tag{4.26}$$

Having obtained approximate $F(hkl)$ values, we can perform the inverse Fourier transform to get an approximated electron density

$$
\begin{aligned}
\rho(x, y, z) &= \mathcal{F}^{-1}[|F(hkl)| \exp[\mathrm{i}\Phi(hkl)]] \\
&\approx \mathcal{F}^{-1}[|F(hkl)| \exp[\mathrm{i}\Phi_{\text{heavy}}(hkl)]] \\
&= \frac{1}{V_{\text{cell}}} \sum_{h,k,l} |F(hkl)| \exp[\mathrm{i}\Phi_{\text{heavy}}(hkl) - 2\pi\mathrm{i}(hx + ky + lz)].
\end{aligned}
\tag{4.27}
$$

This seems to be very crude so far, probably only slightly better than just guessing random Φ values. However, we will usually find that we are getting sufficiently close to the true ρ such that we can identify some peaks other than those due to the heavy atoms. If this is the case, we know some extra coordinates of some atoms. We can include them in our set of "heavy" atoms when calculating Φ_{heavy} — this will generally mean a better approximation of the true phase. A further inverse transform will give us a better approximation to the true electron density from which, hopefully, more atomic positions can be deduced. The iteration process can be repeated until all of the atoms are found. At this stage Φ_{heavy} will be Φ exactly, and we have completely solved the structure.
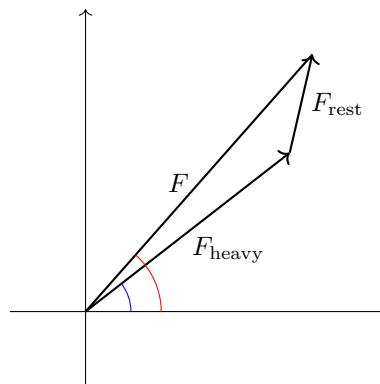

Figure 4.8: A refined approximation of Φ by including more atoms in the set of known 'heavy' atoms.

5 Phase Determining Methods

There are many occasions in which the heavy atom fails to work, for example because the heavy atoms are not 'heavy' enough to dominate the diffraction. Here, we will introduce two phase determining methods that can determine the phase of the diffraction beam more rigorously once the positions of the heavy atoms are known. Both methods achieve this by changing the way that the known heavy atom interacts with the X-ray, either by changing the identity of the atom or changing the way it scatters the X-ray.

5.1 Isomorphous Replacement

If we know the identity and positions of some type of atom (labelled a) in the lattice, then we can write the structure factor as

$$F_1(hkl) = F_{\text{rest}}(hkl) + \sum_{i=1}^{n} f_a \exp[2\pi\mathrm{i}(hx_i + ky_i + lz_i)]. \tag{5.1}$$

Now suppose we can replace these atoms with another type of atom (labelled b) at exactly the same positions, and suppose such replacement does not change any other part of the cell (the size and shape of the lattice is not changing as well as the coordinate of other atoms). This is known as an *isomorphous replacement*. Then the structure factor after such a replacement is

$$F_2(hkl) = F_{\text{rest}}(hkl) + \sum_{i=1}^{n} f_b \exp[2\pi\mathrm{i}(hx_i + ky_i + lz_i)]. \tag{5.2}$$

The difference of the two structure factors, before and after replacement, is

$$\delta F(hkl) = F_2(hkl) - F_1(hkl) = \sum_{i=1}^{n} (f_b - f_a) \exp[2\pi\mathrm{i}(hx_i + ky_i + lz_i)]. \tag{5.3}$$

This is something we can calculate, for both the magnitude and the phase angle, since we know the positions of these atoms. Then since we can measure the magnitude of the structure factor before and after replacement, $|F_1(hkl)|$ and $|F_2(hkl)|$, we can calculate the phase of F_1 and F_2.

However, as you can see in figure 5.1, a set of isomorphous replacement data give two possible values of phase angles. These two possible phase values are related by a reflection with respect to an axis of angle $\Phi(\delta F)$ passing through the origin. To eliminate this ambiguity, we can try to do a second replacement. However, if we replace the same atoms, say a to c, then the change in structure factor is given by

$$\delta F' = \sum_{n} (f_c - f_a) \sum_{i=1}^{n} \exp[2\pi\mathrm{i}(hx_i + ky_i + lz_i)]. \tag{5.4}$$

This is parallel to the first change in the structure factor δF — it will just be either shorter or longer, so it just gives the same set of two possible Φ values. To make progress, we have to perform isomorphous replacement for a different set of atoms, say d to e. Then the change in the structure factor is

$$\delta F' = \sum_{n} (f_e - f_d) \sum_{j=1}^{m} \exp[2\pi\mathrm{i}(hx_j + ky_j + lz_j)]. \tag{5.5}$$

This will give a different set of Φ values — the common one is the true phase of the beam.

This of course brings some experimental difficulties. We need to make three different crystals, and we need to make sure that our replacement does not change anything other than the identity of the atoms that we are replacing — this is not easy to do in practice.
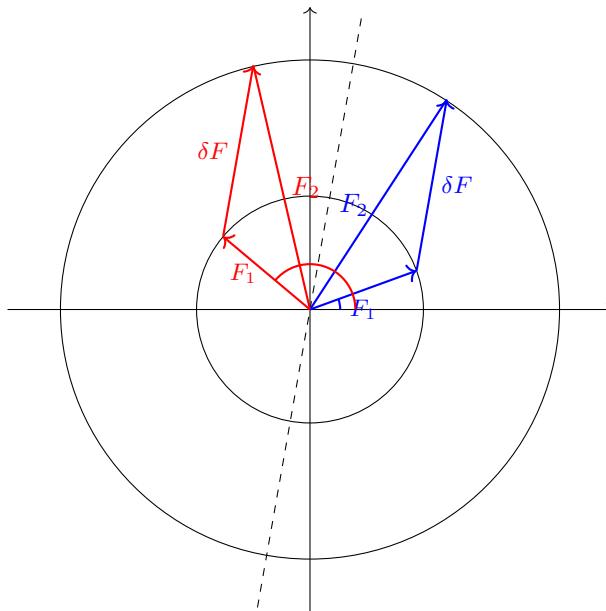
Figure 5.1: There are two possible set of phase angles (blue and red) for a set of isomorphous replacement.
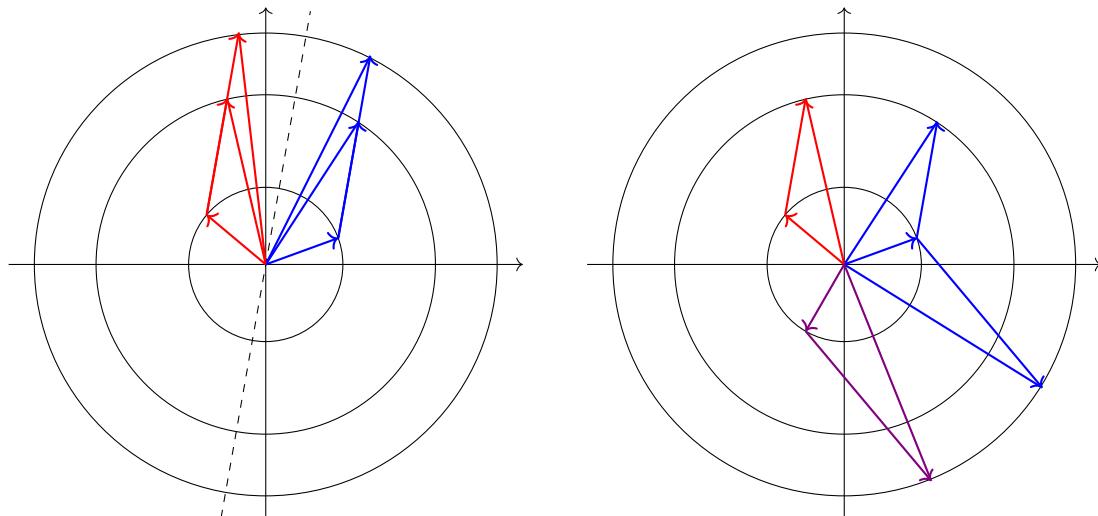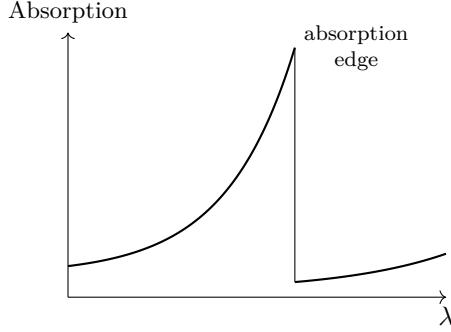
Figure 5.2: An isomorphous replacement of one set of atoms gives two possible set of phase angles (blue and red), and replacing another set of atoms will give a different set of phase angles (blue and violet). The common one is the true answer.

Luckily, such difficulty does not exist in centrosymmetric crystals since the phase will either be 0 or π, i.e. they are real numbers. Once we know δF, including its sign, we can directly determine the sign of F_1 and F_2 from their magnitudes.

5.2 Anomalous Scattering

A second method to determine the phase is to change the way that the electrons scatter the X-ray.

The discussion up to this point has been based on the assumptions of Thomson scattering, which assume that electrons in crystals behave like free electrons. But actually, the electrons are bound

Absorption

absorption
edge

λ

into atoms, which means that the oscillations set up by the incident X-rays are to some extent restrained. Under most circumstances, the deviation from Thomson behaviour is negligible and the description we have given so far is adequate. However, significant differences can arise when the wavelength of the incident X-ray comes close to an *absorption edge* for a particular atom type, where the energy corresponds to an electronic transition or ionisation process. Under this circumstance, the scattering behaviour of the atom will be different from the normal cases. This phenomenon is known as *anomalous scattering* or *resonant scattering*, where 'anomalous' here means deviation from normal (Thomson) behaviour.

To account for anomalous scattering in the structure factor equation, the atomic scattering factors are modified to

$$f_{\text{anom}} = f + (f' + \text{i}f''), \tag{5.6}$$

where f is the normal scattering factor, and $f' + \text{i}f''$ is a complex correction, where both f' and f'' are real. The values of f' and f'' depend on the X-ray wavelength but they are independent of scattering angle, and they are generally small compared to f. f' can be either positive or negative, but f'' is always positive.

Since f' and f'' are small compared to f, it is usually sufficient to approximate all of the effect solely by adding the imaginary correction, so

$$f_{\text{anom}} \approx f + \text{i}\Delta f. \tag{5.7}$$

We can then divide the atoms into the set that scatters normally, and the ones that scatter anomalously. Then the structure factor under anomalous conditions is

$$\begin{aligned}
F'(hkl) &= \sum_{n \text{ norm}} f_n \text{e}^{2\pi\text{i}(hx_n + ky_n + lz_n)} + \sum_{m \text{ anom}} (f_m + \text{i}\Delta f_m)\text{e}^{2\pi\text{i}(hx_m + ky_m + lz_m)} \\
&= \sum_n f_n \text{e}^{2\pi\text{i}(hx_n + ky_n + lz_n)} + \text{i}\sum_{\text{anom}} \Delta f_m \text{e}^{2\pi\text{i}(hx_m + ky_m + lz_m)} \\
&= F(hkl) + \delta F(hkl), \tag{5.8}
\end{aligned}$$

where $F(hkl)$ is the structure factor under normal scattering and change in the structure factor due to anomalous scattering is

$$\delta F(hkl) = \text{i}\sum_{\text{anom}} \Delta f_m \text{e}^{2\pi\text{i}(hx_m + ky_m + lz_m)}. \tag{5.9}$$

Graphically, if an atom originally contribute $f_n \text{e}^{\text{i}\phi}$ to the structure factor, then the anomalous change is

$$\text{i}\Delta f_n \text{e}^{\text{i}\phi} = \Delta f_n \text{e}^{\text{i}(\phi + \frac{\pi}{2})}, \tag{5.10}$$

i.e. it is 90° counterclockwise to the normal scattering contribution.

An important consequence of anomalous scattering is the breakdown of the Friedel's law: $F(hkl)$ and $F(\bar{h}\bar{k}\bar{l})$ are no longer complex conjugates of each other, and so $I(hkl)$ and $I(\bar{h}\bar{k}\bar{l})$ are no longer
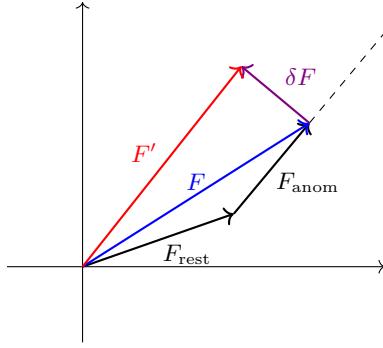
Figure 5.3: Anomalous Scattering.
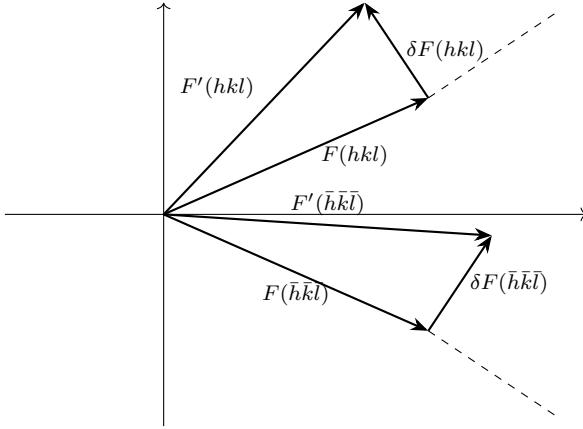
Figure 5.4: The Friedel's law is broken when anomalous scattering happens.

necessarily equal. This is because

$$F'(hkl) = \sum_n f_n e^{2\pi i(hx_n + ky_n + lz_n)} + i \sum_{\text{anom}} \Delta f_m e^{2\pi i(hx_m + ky_m + lz_m)} \qquad (5.11)$$

and so

$$F'(hkl)^* = \sum_n f_n e^{-2\pi i(hx_n + ky_n + lz_n)} - i \sum_{\text{anom}} \Delta f_m e^{-2\pi i(hx_m + ky_m + lz_m)}, \qquad (5.12)$$

while

$$F'(\bar{h}\bar{k}\bar{l}) = \sum_n f_n e^{-2\pi i(hx_n + ky_n + lz_n)} + i \sum_{\text{anom}} \Delta f_m e^{-2\pi i(hx_m + ky_m + lz_m)}. \qquad (5.13)$$

This is illustrated in the figure below.

Note however that we still have $I(hkl) = I(\bar{h}\bar{k}\bar{l})$ if the structure is centrosymmetric because then the two triangles will overlap.

The breakdown of Friedel's law under anomalous scattering is actually extremely useful because it is now possible to distinguish the *absolute structure* (handedness) of a non-centrosymmetric crystal.

If the coordinates of all atoms in the non-centrosymmetric crystal is inverted with respect to the origin $(x_n, y_n, z_n) \rightarrow (-x_n, -y_n, -z_n)$, then we see that we have obtained its enantiomer. Hence, if A and B are enantiomers, we must have

$$F_{\text{A}}(hkl) = F_{\text{B}}(\bar{h}\bar{k}\bar{l}), \qquad (5.14)$$

and so $I_{\text{A}}(hkl) = I_{\text{B}}(\bar{h}\bar{k}\bar{l})$. But now since $I_{\text{B}}(\bar{h}\bar{k}\bar{l}) \neq I_{\text{B}}(hkl)$ under anomalous scattering, $I_{\text{A}}(hkl) \neq I_{\text{B}}(hkl)$. This allows us to determine the handedness of the crystal — we just need to check which handedness of the model matches the measured intensities better than the other.

In practice, we would set our model to one of the handedness, and work out the value of x such that

$$I_{\text{measured}}(hkl) = (1 - x)I_{\text{model}}(hkl) + xI_{\text{model}}(\bar{h}\bar{k}\bar{l}), \tag{5.15}$$

and take the average over *hkl*. x is known as the *Flack parameter*, and if our current handedness is correct, x should be 0 within some statistical error, while if x is near 1, the actual handedness of the crystal should be flipped.

The assignment of absolute structure becomes more confident as the magnitude of the anomalous signal increases, which is achieved when heavy atoms are present. MoK_α radiation ($\lambda = 0.7107\,\text{Å}$) generally requires the crystal to contain Cl atoms or heavier, while CuK_α radiation ($\lambda = 1.5418\,\text{Å}$) can be successful with lighter atoms. More detailed statistical analysis of the intensities can often confidently determine the absolute structure of CHNO compounds with CuK_α radiation. This is incredibly useful for organic chemists.

However, due to the same reason as for isomorphous replacement, a calculated δF and measured $|F(hkl)|$ and $|F'(hkl)|$ will give two possible values of phase angle $\Phi(hkl)$. The solution is also the same — we use another X-ray wavelength under which another set of atoms would scatter anomalously, and this will give another set of two possible phase angles. The common answer is the true answer. This technique is known as *multiple anomalous scattering (MAD)*. It is particularly easy to apply if a source of variable wavelength is available, such as a synchrotron, so that several types of atoms can be made to scatter anomalously in turn.

6 Direct Methods

All the techniques introduced above rely on the fact that we know the positions of some atoms in the cell (the heavy atoms) using Patterson function. However, sometimes we can't even make a start using Patterson function and so none of the above methods work. We have to find an alternative route to the structure. Methods to do this are referred to as *direct methods* because we are attempting to deduce the phases directly from the observed intensities without going through the Patterson function.

The development of direct methods, coupled with the simultaneous growth of computing power, is principally responsible for establishing X-ray diffraction as a 'standard' analytical technique in chemistry. It eliminates the requirement for heavy atoms, and hence enables structure determination for any compound.

The theory is associated with probabilities, and can be fully described with the associated mathematical language rigorously. Here, we will only attempt to illustrate the basic concept using the understanding of Fourier synthesis.

6.1 The Basic Principle

For simplicity, we will consider a 1D crystal. The electron density is given by the Fourier synthesis (3.35)

$$\rho(x)V_{\text{cell}} = \sum_{h=-\infty}^{\infty} F(h)e^{-2\pi ihx}$$

$$= F(0) + 2\sum_{h=1}^{\infty} |F(h)|\cos[\Phi(h) - 2\pi hx]\,, \tag{6.1}$$

where we have used the fact that $\Phi(-h) = -\Phi(h)$. To simplify things further, we assume that the structure has a centre of symmetry, so $\Phi(h) \in \{0, \pi\}$. Then the summation becomes

$$\rho(x)V_{\text{cell}} = F(0) + 2\sum_{h=1}^{\infty} S(h)\,|F(h)|\cos(2\pi hx)\,, \tag{6.2}$$

where $S(h) = +1$ if $\Phi(h) = 0$, while $S(h) = -1$ if $\Phi(h) = \pi$. Consequently, the electron density is the sum of a series of cosine functions with period $1/h$, amplitude $|F(h)|$ and sign $S(h)$. In principle, we could determine $\rho(x)$ by simply trying every possible combinations of $S(h)$ values until we have found one which gives a reasonable electron density. But for N measured $|F(h)|$ values, there are 2^N combinations, so it is highly impractical to check them one by one.

The fundamental fact we will rely on is that electron density is a positive quantity. There may be a small negative region in our synthesis of $\rho(x)$ due to the inaccuracies in the measurement of $F(hkl)$ values (subject to experimental error) and series termination errors (due to the finite number of structure factors that can actually be measured), but there cannot be any significant negative regions — this requirement will allow us to rule out certain combination of phases.

6.2 Fixing the Origin

Before determining the phases *ab initio*, let's go back to the basics and consider again what does the "phase" mean. When defining the phase, we set an origin in our structure. This does not mean that we necessarily need to physically identify a point inside the crystal during a diffraction experiment.

Rather, we need to specify the values of a few phases, so that we can define all other phases relative to those.

For example, let's consider the $(h00)$ beam, where we write

$$F(h00) = \sum_n f_n \exp[2\pi \mathrm{i} h x_n] . \tag{6.3}$$

What we actually mean is

$$F(h00) = \sum_n f_n \exp[2\pi \mathrm{i} h (x_n - x_O)] , \tag{6.4}$$

where x_O is the x coordinate of the origin. When we drop $-x_O$, it is understood that x_n is a fractional coordinate defined relative to the origin. Obviously, the value of $x_n - x_O$, and therefore the phase of $F(h00)$, depends on the value of x_O. But oppositely, if we choose the phase of one of the $h00$ beams to be 0, we effectively define x_O. In a 3D structure, we could define the phase of three independent beams, say $h00$, $0k0$ and $00l$ for some h, k and l, then we have effectively fixed the origin in the structure.

Let's illustrate this using the common space group $P2_1/c$. We claimed before that if a space group is centrosymmetric, as for this one, then the structure factor would always be a real number, with phase 0 or π. This is actually true if and only if we set the origin to coincide with a centre of inversion. Of course it doesn't have to be, but we would always do so for simplicity. In $P2_1/c$, there are 8 centres of inversion within a unit cell, located at

$$(0,0,0), \ (0,0,\tfrac{1}{2}), \ (0,\tfrac{1}{2},0), \ (\tfrac{1}{2},0,0), \ (\tfrac{1}{2},\tfrac{1}{2},0), \ (\tfrac{1}{2},0,\tfrac{1}{2}), \ (0,\tfrac{1}{2},\tfrac{1}{2}) \ \text{and} \ (\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}), \tag{6.5}$$

where $(0,0,0)$ is set to be one of these centres of inversion. Then the structure factor referring to this origin is

$$F(hkl)_{0,0,0} = \sum_n f_n \cos[2\pi(hx_n + ky_n + lz_n)] . \tag{6.6}$$

Alternatively, if we move the origin to a different one, say the one we currently label as $(\tfrac{1}{2},\tfrac{1}{2},0)$, an atom that originally had fractional coordinate (x_n, y_n, z_n) would now have coordinate $(x_n - \tfrac{1}{2}, y_n - \tfrac{1}{2}, z_n)$. Hence, relative to this origin, the structure factor becomes

$$F(hkl)_{\frac{1}{2},\frac{1}{2},0} = \sum_n f_n \cos\left[2\pi\left((hx_n + ky_n + lz_n) - \frac{h+k}{2}\right)\right] = (-1)^{h+k} F(hkl)_{0,0,0} . \tag{6.7}$$

Consequently, the magnitude of each structure factor will have the same magnitude after this change, but its sign will change according to whether $h + k$ is odd or even.

Similar expressions can be obtained for all choices of the origin.

$$\begin{aligned}
F(hkl)_{0,0,\frac{1}{2}} &= (-1)^l F(hkl)_{0,0,0} , \\
F(hkl)_{0,\frac{1}{2},0} &= (-1)^k F(hkl)_{0,0,0} , \\
F(hkl)_{\frac{1}{2},0,0} &= (-1)^h F(hkl)_{0,0,0} , \\
F(hkl)_{0,\frac{1}{2},\frac{1}{2}} &= (-1)^{k+l} F(hkl)_{0,0,0} , \\
F(hkl)_{\frac{1}{2},0,\frac{1}{2}} &= (-1)^{h+l} F(hkl)_{0,0,0} , \\
F(hkl)_{\frac{1}{2},\frac{1}{2},0} &= (-1)^{h+k} F(hkl)_{0,0,0} , \\
F(hkl)_{\frac{1}{2},\frac{1}{2},\frac{1}{2}} &= (-1)^{h+k+l} F(hkl)_{0,0,0} .
\end{aligned} \tag{6.8}$$

Hence, if we specify the phases of three carefully chosen beams, say 312, 322 and 221 all to have phase 0, then there will be only one of the eight possible origins to agree with all of these three phase values (try it!), and hence we have successfully specified the origin.

Moreover, we can expand the set of known phases by symmetry. By inspection on the general equivalent positions of the $P2_1/c$ space group, we get

$$F(hkl) = F(h\bar{k}l) = F(\bar{h}k\bar{l}) = F(\bar{h}\bar{k}\bar{l}) \,. \tag{6.9}$$

Hence, by solely defining the origin, we get 12 phase values already.

6.3   Projected Structures and Unitary Structure Factors

The basic principle of the direct method is the non-negativity of electron density. For a centrosymmetric crystal,

$$\rho(x,y,z) = \sum_{h,k,l=-\infty}^{\infty} S(hkl)\,|F(hkl)|\cos[2\pi(hx+ky+lz)]\,, \tag{6.10}$$

where $S(hkl) = \pm1$. We see that a beam with strong $|F(hkl)|$ has a large influence on the resulting Fourier summation, so we would expect to be able to establish their phases most confidently. However, we would expect a very large number of terms in this triple summation, and hence even strong $|F(hkl)|$ beams will have a small contribution, and we would be less confident with any phase we propose.

This problem will be reduced if there are fewer terms in the series, which can be achieved if we project the structure onto a line. For example,

$$\rho(x) = F(000) + 2\sum_{h=1}^{\infty} S(h00)\,|F(h00)|\cos(2\pi hx) \tag{6.11}$$

would describe the electron density projected onto the x axis, and

$$\rho(d) = F(000) + 2\sum_{h=1}^{\infty} S(hhh)\,|F(hhh)|\cos(6\pi hd) \tag{6.12}$$

would describe the electron density projected onto the body diagonal of the cell, where d is the coordinate projected to the diagonal. In both cases, the non-negativity principle must still apply, and the number of terms in the sum has drastically reduced as we hoped to.

However, the projection brings another issue. In the original 3D structure, the electron density consists of electron-rich atoms separated by regions of very low electron density. However, this property is likely to be lost if the electron density is projected onto a line. The consequence is that, in general, it will be harder to generate regions of negative electron density even if some $S(hkl)$ are incorrect. Therefore the predictive power of the non-negativity principle will decrease.
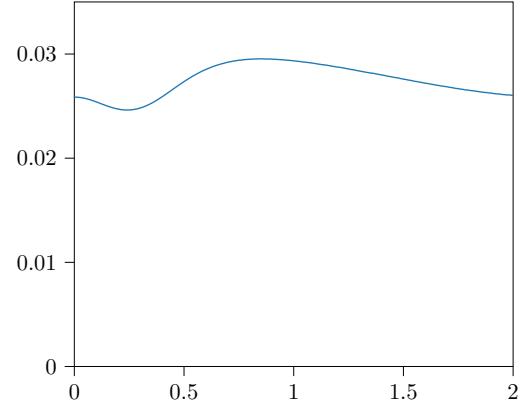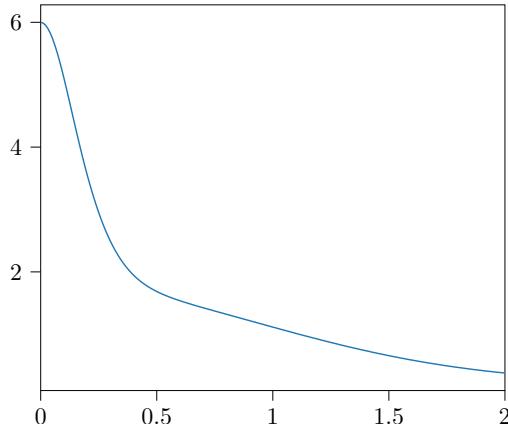
There is an elegant way to get around this. Since atomic scattering factor looks roughly the same shape for all atoms, differing mainly only their scale, we can define a *unitary scattering factor*

$$n_n = \frac{f_n}{\sum_m f_m}\,. \tag{6.13}$$

This represents the fraction of the scattering power of the atom relative to the overall scattering power of the cell, and since f_n has similar shapes, n_n will be roughly constant without any decay. Hence, the unitary scattering factor represents the scattering of a point atom, as the inverse Fourier transform of a constant function is a delta function.

From the unitary scattering factor, we can define the *unitary structure factor*

$$U(hkl) = \sum_n n_n \exp[2\pi\mathrm{i}(hx_n+ky_n+lz_n)] = \frac{F(hkl)}{\sum_n f_n}\,. \tag{6.14}$$

(a) Scattering factor of carbon.

(b) Unitary scattering factor of a carbon atom in a cell with content $C_{24}H_{24}O_8$.

An extra benefit of doing so is that because of the fall-off in the scattering power with diffraction angle, some beams which are inherently strong (in that all atoms scatter in phase) will nevertheless be weak. Unitary structure factors nicely remove this dependence on scattering angle. Finally, an extra helpful thing to do is to normalise it by the mean-squared unitary structure factor

$$E(hkl) = \frac{U(hkl)}{\sqrt{\langle U(hkl)^2 \rangle}} \,. \tag{6.15}$$

6.4   Pair and Triple Relationships

6.4.1   Pair Relationship

Suppose we have a normalised electron density projected onto the x axis

$$\rho(x) = 1 + 2 \sum_{h>0} S(h) \, |E(h)| \cos(2\pi h x) \,. \tag{6.16}$$

Now consider the case when there are two very strong beams, with indices h and $2h$, of similar magnitudes $|E(h)| \approx |E(2h)|$, we write

$$\rho(x) \approx 1 + 2 \, |E| \, [S(h) \cos(2\pi h x) + S(2h) \cos(4\pi h x)] \,. \tag{6.17}$$
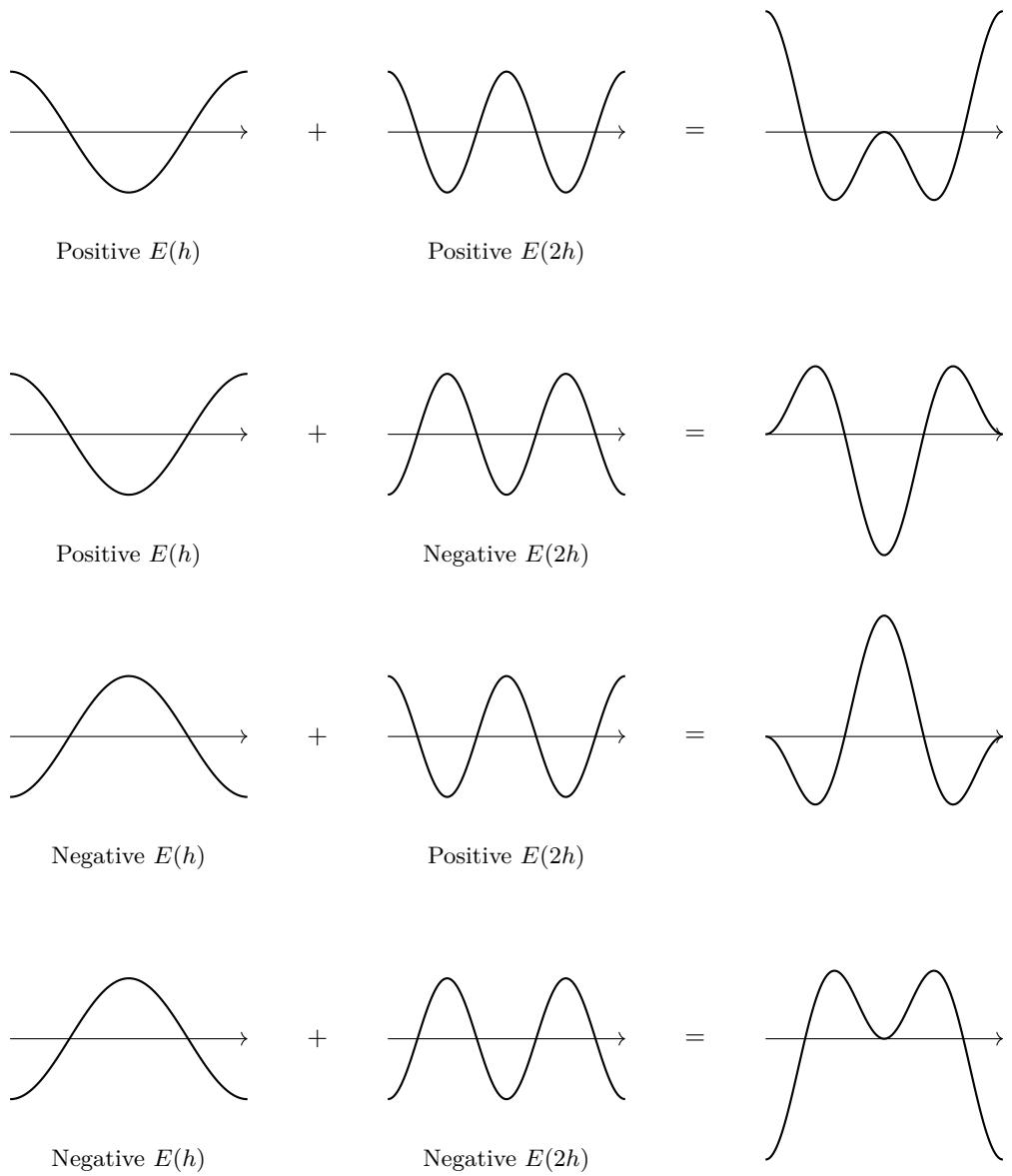
If the sum of these two contributions have a significantly negative region, it is unlikely that other terms in the Fourier synthesis will compensate for that, so we wish their minima will not overlap. Let's check different combinations of $S(h)$ and $S(2h)$, as shown in figure 6.2.

We see that regardless of $S(h)$ and $S(2h)$, we will have some negative regions. However, if $S(2h)$ is negative, whatever $S(h)$ is, the minima of the two cosine waves will overlap, resulting in a strongly negative region. This is unlikely. Hence, we get the pair relationship: if $E(h)$ and $E(2h)$ are both large, then $S(2h)$ is likely to be $+1$.

As we can take projection along any axis, this result can be generalised: if $E(\mathbf{h})$ and $E(2\mathbf{h})$ are large, where $\mathbf{h} = hkl$ is the Miller index, then $S(2\mathbf{h}) = +1$.

6.4.2   Triple Relationship

A similar, but more involved analysis can be done for $E(\mathbf{h})$, $E(\mathbf{h}')$ and $E(\mathbf{h} + \mathbf{h}')$, assuming all large and of similar magnitudes.

Positive $E(h)$ +   Positive $E(2h)$ =

Positive $E(h)$ +   Negative $E(2h)$ =

Negative $E(h)$ +   Positive $E(2h)$ =

Negative $E(h)$ +   Negative $E(2h)$ =

Figure 6.2: The minima of the two terms will overlap if $E(2h)$ is negative.
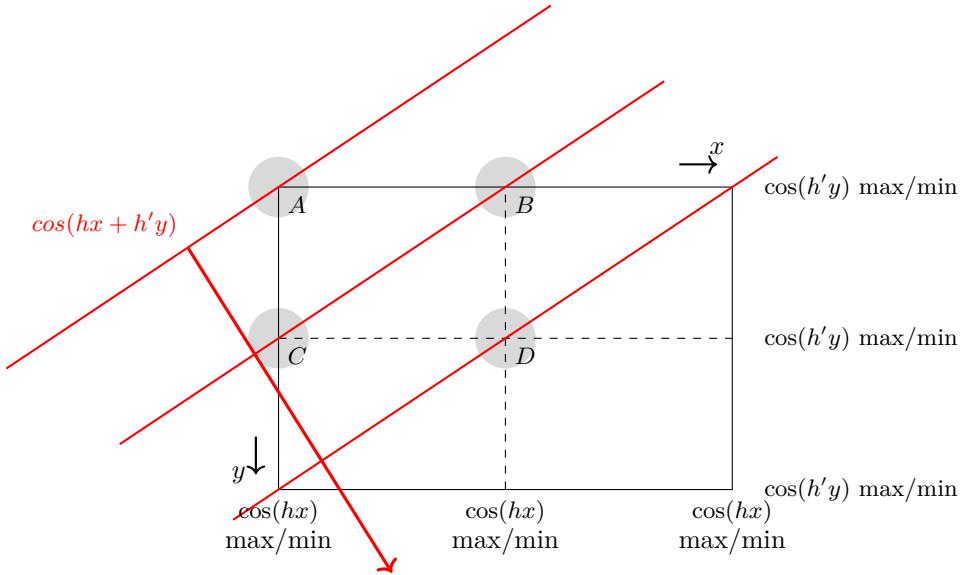
Figure 6.3: Triple relationship.

Let's imagine forming a two dimensional space, plotting $\pm \cos(hx)$ and $\pm \cos(h'y)$ along the two axis, where $h = \|\mathbf{h}\|$ and $h' = \|\mathbf{h}'\|$. This is equivalent to reorient the two Fourier waves $\pm \cos(\mathbf{S} \cdot \mathbf{r})$ and $\pm \cos(\mathbf{S}' \cdot \mathbf{r})$ along x and y axes, respectively. As shown in figure 6.3, the minimum of the sum of these two waves will be at one of the grid points A, B, C or D. The third wave $\pm \cos(hx + h'y)$, corresponding to $\pm \cos((\mathbf{S} + \mathbf{S}') \cdot \mathbf{r})$ with miller index $\mathbf{h} + \mathbf{h}'$, will have its wavefront at the diagonal, and we hope it to have its maximum aligned with the minimum of the previous two waves combined in order to compensate for it.

$S(h)$	$S(h')$	$E(h + h')$ must diminish minimum at position	$S(h + h')$	$S(h)S(h')S(h + h')$
$+1$	$+1$	D	$+1$	$+1$
$+1$	-1	C	-1	$+1$
-1	$+1$	A	-1	$+1$
-1	-1	B	$+1$	$+1$

This can be summarised as the *triple relationship*

$$S(\mathbf{h})S(\mathbf{h}')S(\mathbf{h} + \mathbf{h}') = +1 \,. \tag{6.18}$$

One can see that this reduces to the pair relationship if we just set $\mathbf{h} = \mathbf{h}'$.

The advantage of this method is that it can be automated with trials and errors on a computer, requiring virtually no user input.

6.4.3 Limitation of Direct Methods

Recall that the way we figure out $\rho(x, y, z)$ is to use a Fourier series, and if we include an infinite sum of diffracted beams, we should be able to reproduce it exactly. However, we can only measure diffraction up to a finite maximum scattering angle, and hence we have a finite number of terms in the Fourier synthesis. This results in the *series termination error*. This can be particularly problematic in direct methods.

Let's consider the example of caesium picrate $Cs[C_6H_2(NO_2)_3O]$, projected onto the x axis. Figure 6.4 shows progressively larger number of $E(h)$ values included in the Fourier synthesis of
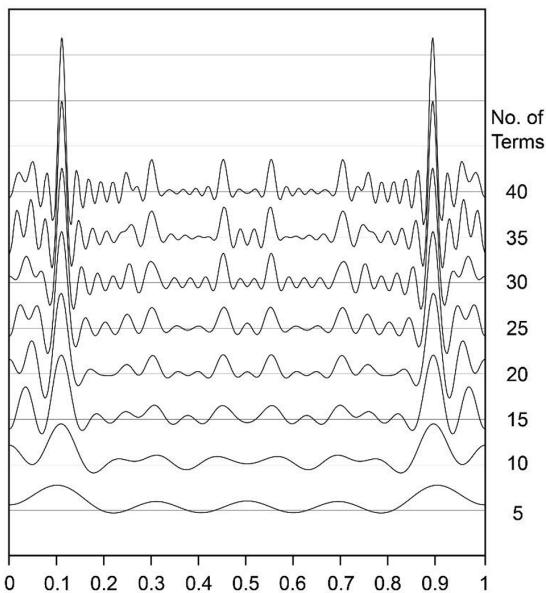
Figure 6.4: Fourier synthesis of electron density in caesium picrate $Cs[C_6H_2(NO_2)_3O]$ as number of terms increases.

electron density. We see with only five terms included in the sum, only Cs atoms at $x \approx 0.1$ and $x \approx 0.9$ can be observed, and the C, N and O atoms are not resolved. As more terms are included, the lighter atoms become visible, and the large Cs peaks also becomes sharper. However, the Cs peaks are also surrounded by large negative ripples, which arises due to the series termination errors. These remain strong even when 40 terms are included in the series, only disappearing when the number of terms exceeds 100. Therefore, we are likely to observe negative electron density regions around heavy atoms even when our phases are correct. Consequently, any method relying on the non-negativity of electron density may falsely rule out the correct phases combinations.

The solution is also simple. If heavy atoms are present, then Patterson function plus phase determining methods based on heavy atoms would be ideal. Direct method and heavy atom based methods are perfectly complementary.

Moreover, for molecules like proteins when there are no particularly heavy atoms and the unit cells are large, there might be no particularly strong peaks, and the intensities are measured at low resolutions. For them, multiple isomorphous replacement, multiple anomalous scattering or (most commonly) single anomalous scattering would be used.

6.5   Dual-Space Methods

In fact, most chemical structures solved today use neither the Patterson function nor direct methods. Instead, modern computing power enables application of the following algorithm called the *dual-space method*.

1. Produce a random phase $\Phi(hkl)$ for each measured $F(hkl)$.

2. Generate $\rho(x, y, z)$ at points on a defined grid within the unit cell by the inverse Fourier transform using measured $|F(hkl)|$ and current $\Phi(hkl)$.

3. For any grid point with negative $\rho(x, y, z)$, multiply the value by -1.

4. Calculate new $F(hkl)$ by the Fourier transform of $\rho(x, y, z)$.

5. Replace the current phase set with the $\Phi(hkl)$ values obtained in step 4, then return to step 2.

The above operation will converge to a $\rho(x, y, z)$ which is necessarily non-negative. We would repeat the procedure many times from different initial sets and to assess which of the converged solutions is the most likely to be correct.

The fundamental basis of the dual space method is clearly the non-negativity of electron density, and the process of flipping negative $\rho(x, y, z)$ is an example of *density modification*, meaning modifying the calculated electron density to conform to our expectations. The established principle of atomicity can also be applied to the structure at some defined frequency during the iteration. This is achieved by identifying a number of maxima in ρ and replacing them by isolated spherical electron densities at these positions.

An additional benefit of dual-space method is that it does not require the symmetry of the crystal structure to be firmly established. They are usually applied to the whole unit cell content.