lattice sites, and the lateral and axial lattice disorder weights are given by

$$w_{\rm lat}(R, r) = \exp\left(-4\pi^2 R^2 \sigma_{\rm lat}^2 [1 - \rho_{\rm lat}(r)]\right) \qquad (4.5.2.42)$$

and

$$w_{\rm axial}(Z, r) = \exp\left(-4\pi^2 Z^2 \sigma_{\rm axial}^2 [1 - \rho_{\rm axial}(r)]\right). \qquad (4.5.2.43)$$

Equation (4.5.2.41) is an expression for the continuous intensity distribution along the layer lines and does not separate into Bragg and continuous components as in the case of uncorrelated disorder. However, calculations using these expressions show that the continuous intensity is sharply peaked around the projected reciprocal-lattice points at low resolution, the peaks broadening with increasing resolution until they have the character of continuous diffraction at high resolution (Stroud & Millane, 1996*a*). This is consistent with the character of diffraction patterns from some disordered polycrystalline fibres. A detailed study of the effects of correlated disorder on fibre diffraction patterns, and analysis of such disorder, can be found in Stroud & Millane (1996*a*) and Stroud & Millane (1996*b*).

4.5.2.5. Processing diffraction data

Since the diffraction pattern from a fibre is two-dimensional, it can be collected with a single exposure of a stationary specimen. Diffraction data are collected either on film, which is subsequently scanned by a two-dimensional microdensitometer to obtain a digitized representation of the diffracted intensity, or using an electronic area detector (imaging plate, CCD camera, wire detector *etc.*) (Fraser *et al.*, 1976; Namba, Yamashita & Vonderviszt, 1989; Lorenz & Holmes, 1993). We assume here that the diffraction pattern is recorded on a flat film (or detector) that is normal to the incident X-ray beam, although other film geometries are easily accommodated (Fraser *et al.*, 1976). The fibre specimen is usually oriented with its axis normal to the incident X-ray beam, although, as is described below, it is sometimes tilted by a small angle to the normal in order to better access reciprocal space close to the meridian. The diffraction and camera geometry are shown in Fig. 4.5.2.1. Referring to this figure, P and S denote the intersections of the diffracted beam with the sphere of reflection and the film, respectively. The fibre, and therefore reciprocal space, is tilted by an angle β to the normal to the incident beam. The angles μ and χ define the direction of the diffracted beam and θ is the Bragg angle. Cartesian and polar coordinates on the film are denoted by (u, v) and (r, φ), respectively, and D denotes the film-to-specimen distance.

Inspection of Fig. 4.5.2.1 shows that the cylindrical (R, ψ, Z) and spherical (ρ, ψ, σ) polar coordinates in reciprocal space are related to μ and χ by

$$\rho = (1/\lambda)[2(1 - \cos\mu - \cos\chi)]^{1/2}, \qquad (4.5.2.44)$$

$$Z = (1/\lambda)[\sin\beta(1 - \cos\mu\cos\chi) + \cos\beta\sin\chi]^{1/2}, \qquad (4.5.2.45)$$

$$R = (\rho^2 - Z^2)^{1/2}, \qquad (4.5.2.46)$$

$$\sin\psi = \frac{\sin\mu\cos\chi}{R\lambda} \qquad (4.5.2.47)$$

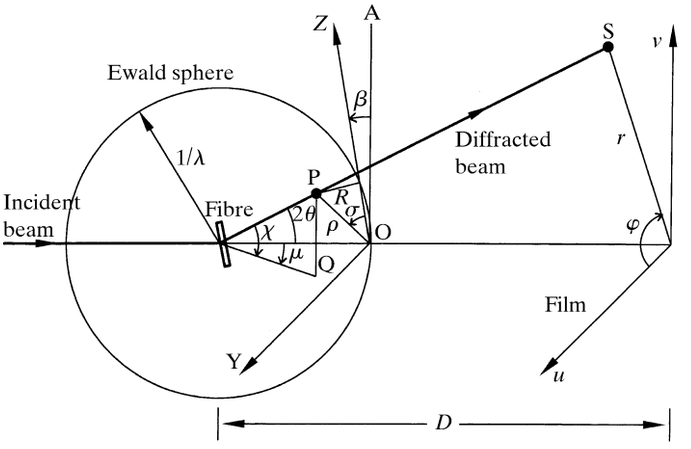

Fig. 4.5.2.1. Fibre diffraction geometry (see text). O is the origin of reciprocal space and **OA** is normal to the incident X-ray beam. Reciprocal space is rotated about **OY** so that the Z axis is inclined at an angle β to **OA**. Q is the projection of P onto the plane containing the incident beam and **OY**.

and

$$\tan\sigma = R/Z. \qquad (4.5.2.48)$$

The coordinates on the film are related to μ and χ by

$$u = D\tan\mu \qquad (4.5.2.49)$$

and

$$v = D\cos\mu\tan\chi, \qquad (4.5.2.50)$$

and we also have that

$$r = D\tan 2\theta. \qquad (4.5.2.51)$$

Use of the above equations allows the reciprocal-space coordinates to be calculated from film-space coordinates, and *vice versa.* The film coordinates (u, v) represent a relatively undistorted map of reciprocal space (R, Z), except near the v (vertical) axis of the diffraction pattern. The meridian of reciprocal space does not map onto the film. Inspection of Fig. 4.5.2.1 shows that the only point on the meridian that does appear on the film is at $Z = \lambda^{-1}\sin\beta$. The region *close* to the meridian that appears on the film can therefore be manipulated by adjusting the fibre tilt.

The film-to-specimen distance can be determined by including with the specimen a crystalline power that gives a diffraction ring of known spacing and adjusting the film-to-specimen distance so that the calculated and observed rings coincide. A nonzero fibre tilt leads to differences between the upper and lower halves of the diffraction pattern, and these differences can be used to determine the tilt. This can be done by either calculating the ρ and χ values for several sets of the same reflection above and below the equator and using the relationship

$$\tan\beta = \frac{\sin\chi_{\rm U} + \sin\chi_{\rm L}}{\lambda^2\rho^2}, \qquad (4.5.2.52)$$

where $\chi_{\rm U}$ and $\chi_{\rm L}$ refer to the upper and lower ($\chi < 0$) reflections (Millane & Arnott, 1986; Lorenz & Holmes, 1993), or by finding the tilt that minimizes the differences between optical densities at the same reciprocal-space coordinates above and below the

equator (Fraser *et al.*, 1976). The optical densities may also be corrected for the effects of film (or detector) nonlinearity (Fraser *et al.*, 1976) and the effects of variable absorption owing to the oblique passage of the beam through the film using expressions given by Fraser *et al.* (1976) and Lorenz & Holmes (1993).

Accurate subtraction of background diffraction is important in order to obtain accurate intensity measurements. One approach to estimating background diffraction is to fit a global background function, usually expanded as a polynomial (Lorenz & Holmes, 1993) or a Fourier–Bessel (Millane & Arnott, 1985) series, to optical densities at a set of points on the diffraction pattern that represent background alone. The background function may or may not be circularly symmetric. The background function is subtracted from the whole diffraction pattern. Another approach, suitable only for Bragg diffraction patterns, is to fit a plane under each reflection, either to the peripheral regions of the reflection or as part of a profile-fitting procedure (Fraser *et al.*, 1976). A different plane is required for each reflection. A third approach, more suitable for continuous diffraction patterns, is to fit a one-dimensional polynomial in angle, for each value of r on the film, possibly as part of a deconvolution procedure (Makowski, 1978). Recently, Ivanova & Makowski (1998) have described an iterative low-pass filtering technique for estimating the background on diffraction patterns from poorly oriented specimens in which there is little space between the layer lines for sampling the background.

A polarization correction is applied to the diffraction pattern, where for unpolarized X-rays (laboratory sources) the polarization factor p is given by (Fraser *et al.*, 1976)

$$p = (1 + \cos^2 2\theta)/2. \tag{4.5.2.53}$$

The diffraction pattern is usually mapped into reciprocal space (R, Z) for subsequent analysis. The mapping is performed by assigning to the intensity $I(R, Z)$ at position (R, Z) in reciprocal space the value given by (Fraser *et al.*, 1976)

$$I(R, Z) = \frac{I(u(R, Z); v(R, Z))}{\cos^3 \mu \cos^3 \chi}, \tag{4.5.2.54}$$

where $I(u; v)$ denotes the intensity on the film. The functions $u(R, Z)$ and $v(R, Z)$ can be derived from the equations given above. Note that equation (4.5.2.54) includes, implicitly, the Lorentz factor.

Subsequent processing depends on whether the diffraction pattern is continuous (*i.e.* from a noncrystalline specimen) or Bragg (*i.e.* from a polycrystalline specimen). Diffraction patterns from partially crystalline specimens that contain both components have been analysed using a combination of both approaches (Arnott *et al.*, 1986; Park *et al.*, 1987).

For a diffraction pattern containing continuous diffraction on layer lines, one usually extracts the cylindrically averaged transform $I_l(R)$ from the intensity $I(R, Z)$ on the diffraction pattern mapped into reciprocal space. This involves correcting for the effects of coherence length and disorientation expressed by equation (4.5.2.19), and for the overlap of the smeared layer lines that results from their increasing width with increasing R. The diffracted intensity $I(\rho, \sigma)$ in polar coordinates in reciprocal space is equal to the sum of the diffraction $I_l(\rho, \sigma)$ due to each (overlapping) layer line so that

$$I(\rho, \sigma) = \sum_l I_l(\rho, \sigma). \tag{4.5.2.55}$$

Referring to equations (4.5.2.55), (4.5.2.19) and (4.5.2.20) shows that if the smearing due to disorientation dominates over that due to coherence length, then for fixed ρ, equation (4.5.2.55) represents a convolution along σ of the layer-line intensities $I_l(\rho \sin \sigma_l)$ with the Gaussian angular profile in equation (4.5.2.19). By mapping the intensity $I(R, Z)$ into polar coordinates as $I(\rho, \sigma)$, or by simply sampling $I(R, Z)$ for fixed ρ and equally spaced samples of σ, $I_l(R)$ can be calculated from $I(\rho, \sigma)$ by deconvolution, usually by some appropriate solution of the resulting system of linear equations (Makowski, 1978). If the effects of coherence length are significant, as they often are, then equation (4.5.2.55) does not represent a convolution since the width of the Gaussian smearing function depends on σ through equation (4.5.2.20). However, the problem can still be posed as the solution of a system of linear equations and becomes one of profile fitting rather than deconvolution (Millane & Arnott, 1986). This allows the layer-line intensities to be extracted from the data beyond the resolution where they overlap, although there is a limiting resolution, owing to excessive overlap, beyond which reliable data cannot be obtained (Makowski, 1978; Millane & Arnott, 1986). This procedure requires that α_0 and l_c be known; these parameters can be estimated from the angular profiles at low resolution where there is no overlap, or they can be determined as part of the profile-fitting procedure.

For a diffraction pattern from a polycrystalline specimen containing Bragg reflections, the intensities $I_l(R_{hk})$ given by equation (4.5.2.24) need to be extracted from the intensity $I(R, Z)$ on the diffraction pattern mapped into reciprocal space. Each composite reflection $I_l(R_{hk})$ is smeared into a spot whose intensity profile is given by equation (4.5.2.27), and adjacent reflections may overlap. The intensity $I_l(R_{hk})$ is equal to the intensity $I(R, Z)$ integrated over the region of the spot, and the intensity at the centre of a spot is reduced, relative to $I_l(R_{hk})$, by a factor that increases with the degree of smearing.

The c repeat can be obtained immediately from the layer-line spacing. Initial estimates of the remaining cell constants can be made from inspection of the (R, Z) coordinates of low-order reflections. These values are refined by minimizing the difference between the calculated and measured (R, Z) coordinates of all the sharp reflections on the pattern.

One approach to measuring the intensities of Bragg reflections is to estimate the boundary of each spot (or a fixed proportion of the region occupied by each spot) and integrate the intensity over that region (Millane & Arnott, 1986; Hall *et al.*, 1987). For spots that overlap, an integration region that is the union of the region occupied by each contributing spot can be used, allowing the intensities for composite spots to be calculated (Millane & Arnott, 1986). This is more accurate than methods based on the measurement of the peak intensity followed by a correction for smearing. Integration methods suffer from problems associated with determining accurate spot boundaries and they are not capable of separating weakly overlapping spots. A more effective approach is one based on profile fitting. The intensity distribution on the diffraction pattern can be written as

$$I(R, Z) = \sum_l \sum_{h,k} I_l(R_{hk}, R, Z), \tag{4.5.2.56}$$

where $I_l(R_{hk}, R, Z)$ denotes the intensity distribution of the spot $I_l(R_{hk})$, and the sums are over all spots on the diffraction pattern. Using equation (4.5.2.27) shows that equation (4.5.2.56) can be written as

$$I(R, Z) = \sum_l \sum_{h,k} I_l(R_{hk}) S(R_{hk}; l/c; R; Z), \tag{4.5.2.57}$$

where $S(R_{hk}; l/c; R; Z)$ denotes the profile of the spot centred at $(R_{hk}, l/c)$ [which can be derived from equation (4.5.2.27)]. Given estimates of the parameters $l_{\rm lat}$, $l_{\rm axial}$ and α_0, equation (4.5.2.57) can be written as a system of linear equations that can be solved

for the intensities $I_l(R_{hk})$ from the data $I(R, Z)$ on the diffraction pattern. The parameters $l_{\rm lat}$, $l_{\rm axial}$ and α_0, as well as the cell constants and possibly other parameters, can also be refined as part of the profile-fitting procedure using nonlinear optimization.

A suite of programs for processing fibre diffraction data is distributed (and often developed) by the Collaborative Computational Project for Fibre and Polymer Diffraction (CCP13) in the UK (http://www.ccp13.ac.uk/) (Shotton *et al.*, 1998).

4.5.2.6. Structure determination

4.5.2.6.1. Overview

Structure determination in fibre diffraction is concerned with determining atomic coordinates or some other structural parameters, from the measured cylindrically averaged diffraction data. Fibre diffraction analysis suffers from the phase problem and low resolution (diffraction data rarely extend beyond 3 Å resolution), but this is no worse than in protein crystallography where phases derived from, say, isomorphous replacement or molecular replacement, coupled with the considerable stereochemical information usually available on the molecule under study, together contribute enough information to lead to precise structures. What makes structure determination by fibre diffraction more difficult is the loss of information owing to the cylindrical averaging of the diffraction data. However, in spite of these difficulties, fibre diffraction has been used to determine, with high precision, the structures of a wide variety of biological and synthetic polymers, and other macromolecular assemblies. Because of the size of the repeating unit and the resolution of the diffraction data, methods for structure determination in fibre diffraction tend to mimic those of macromolecular (protein) crystallography, rather than small-molecule crystallography (direct methods).

For a noncrystalline fibre one can determine only the molecular structure from the continuous diffraction data, whereas for a polycrystalline fibre one can determine crystal structures from the Bragg diffraction data. However, there is little fundamental difference between methods used for structure determination with noncrystalline and polycrystalline fibres. For partially crystalline fibres, little has so far been attempted with regard to rigorous structure determination.

As is the case with protein crystallography, the precise methods used for structure determination by fibre diffraction depend on the particular problem at hand. A variety of tools are available and one selects from these those that are appropriate given the data available in a particular case. For example, the structure of a polycrystalline polynucleotide might be determined by using Patterson functions to determine possible packing arrangements, molecular model building to define, refine and arbitrate between structures, difference Fourier synthesis to locate ions or solvent molecules, and finally assessment of the reliability of the structure. As a second example, to determine the structure of a helical virus, one might use isomorphous replacement to obtain phase estimates, calculate an electron-density map, fit a preliminary model and refine it using simulated annealing alternating with difference Fourier analysis, and assess the results. The various tools available, together with indications of where and how they are used, are described in the following sections.

Although a variety of techniques are used to solve structures using fibre diffraction, most of the methods do fall broadly into one of three classes that depend primarily on the size of the helical repeat unit. The first class applies to molecules whose repeating units are small, *i.e.* are represented by a relatively small number of independent parameters or degrees of freedom (after all stereochemical constraints have been incorporated). The structure can then be determined by an exhaustive exploration of the parameter space using molecular model building. The first example above would belong to this class. The second class of methods is appropriate when the size of the helical repeating unit is such that its structure is described by too many variable parameters for the parameter space to be explored *a priori*. It is then necessary to phase the fibre diffraction data and construct an electron-density map into which the molecular structure can be fitted and then refined. The second example above would belong to this class. The second class of methods therefore mimics conventional protein crystallography quite closely. The third class of problems applies when the structure is large, but there are too few diffraction data to attempt phasing and the usual determination of atomic coordinates. The solution to such problems varies from case to case and usually involves modelling and optimization of some kind.

An important parameter in structure determination by fibre diffraction is the degree of overlap (that results from the cylindrical averaging) in the data. This parameter is equal to the number of significant terms in equation (4.5.2.17) or the number of independent terms in equation (4.5.2.24), and depends on the position in reciprocal space and, for a polycrystalline fibre, the space-group symmetry. The number of degrees of freedom in a particular datum is equal to twice this number (since each structure factor generally has real and imaginary parts), and is denoted in this section by m. Determination of the $G_{nl}(R)$ from the cylindrically averaged data $I_l(R)$ therefore involves separating the $m/2$ amplitudes $|G_{nl}(R)|$ and assigning phases to each. The electron density can be calculated from the $G_{nl}(R)$ using equations (4.5.2.7) and (4.5.2.11).

4.5.2.6.2. Helix symmetry, cell constants and space-group symmetry

The first step in analysis of any fibre diffraction pattern is determination of the molecular helix symmetry u_v. Only the zero-order Bessel term contributes diffracted intensity on the meridian, and referring to equation (4.5.2.6) shows that the zero-order term occurs only on layer lines for which l is a multiple of u. Therefore, inspection of the distribution of diffraction along the meridian allows the value of u to be inferred. This procedure is usually effective, but can be difficult if u is large, because the first meridional maximum may be on a layer line that is difficult to measure. This difficulty was overcome in one case by Franklin & Holmes (1958) by noting that the second Bessel term on the equator is $n = u$, estimating $G_{00}(R)$ using data from a heavy-atom derivative (see Section 4.5.2.6.6), subtracting this from $I_0(R)$, and using the behaviour of the remaining intensity for small R to infer the order of the next Bessel term [using equation (4.5.2.14)] and thence u.

Referring to equations (4.5.2.6) and (4.5.2.14) shows that the distribution of $R_{\rm min}$ for $0 < l < u$ depends on the value of v. Therefore, inspection of the intensity distribution close to the meridian often allows v to be inferred. Note, however, that the distribution of $R_{\rm min}$ does not distinguish between the helix symmetries u_v and u_{u-v}. Any remaining ambiguities in the helix symmetry need to be resolved by steric considerations, or by detailed testing of models with the different symmetries against the available data.

For a polycrystalline system, the cell constants are determined from the (R, Z) coordinates of the spots on the diffraction pattern as described in Section 4.5.2.6.4. Space-group assignment is based on analysis of systematic absences, as in conventional crystallography. However, in some cases, because of possible overlap of systematic absences with other reflections, there may be some ambiguity in space-group assignment. However, the space group can always be limited to one of a few possibilities, and ambiguities can usually be resolved during structure determination (Section 4.5.2.6.4).