Processing of X-ray Diffraction Data from Partially Oriented Specimens

BY LEE MAKOWSKI

Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02154, USA

(Received 30 January 1978; accepted 10 April 1978)

A numerical deconvolution procedure has been developed to measure X-ray diffraction data from partially oriented specimens to the highest possible resolution. The resolution to which X-ray data from non-crystalline specimens can be measured is limited by the disorientation of the particles in the specimen. The distribution of particle orientations causes the X-ray reflections to be smeared into arcs about the center of the diffraction pattern. At sufficiently high diffraction angles, this arcing results in the overlap of adjacent reflections. Conventional methods of densitometry used to measure intensities from X-ray films cannot be used in this region. By measuring all the optical densities on an X-ray film and placing them on a polar grid, an angular deconvolution can be performed to correct for the disorientation and separate the intensities from overlapping reflections. This calculation results in a determination of the X-ray intensities and background, and a quantitative measurement of the highest possible resolution to which the data can be reliably measured. X-ray data from partially oriented specimens of several macromolecular aggregates have been collected with this method.

Introduction

There is a wide variety of macromolecules and macromolecular assemblies that do not crystallize under any known conditions. These include many one-dimensionally periodic structures (e.g. rod-shaped macromolecular assemblies such as microtubules and the helical viruses) and two-dimensionally periodic arrays (e.g. some differentiated membrane structures). It is often possible to obtain partially oriented fibers or gels of these one- and two-dimensionally periodic assemblies. The measurement of X-ray diffraction data from partially oriented specimens presents several problems not encountered in diffraction from crystals. This paper addresses the problem of processing X-ray diffraction data from partially oriented specimens to obtain accurate intensity measurements to the highest possible resolution.

When particles in the specimen exhibit periodicities in one or two dimensions, the diffraction is limited to discrete layer planes or lattice lines. The geometry of diffraction from one- and two-dimensionally periodic structures is shown in Fig. 1. Diffraction from one-dimensionally periodic structures is confined to layer planes in reciprocal space. The intersection of the sphere of reflection with these layer planes gives rise to the observed layer lines which lie along curves of constant height, Z, above and below the equator in reciprocal space. Diffraction from two-dimensionally periodic assemblies is confined to lattice lines in reciprocal space. Intersection of the sphere of reflection with the reciprocal-lattice lines in a rotationally disordered specimen gives rise to diffraction which lies along curves of constant distance, R, from the meridian in reciprocal space.

The correlation in the positions of the particles may be partial, weak or absent. Thus, some or all of the diffraction on layer planes or lattice lines may exhibit crystalline sampling or weaker interference effects, or there may be completely continuous diffraction. The presence of interference will not affect the measurement of intensities as discussed here, but will affect the interpretation of these measurements.

Often there is no correlation in the rotational orientation about the long axes of parallel rod-shaped particles or about an axis normal to the plane of parallel planar assemblies. If no interference effects are present, the recorded diffraction pattern may be thought of as the cylindrically averaged intensity distribution from a single rod or planar assembly. The cylindrical averaging of diffracted intensity severely limits the amount of data that can be collected from partially oriented specimens.

Imperfect orientation implies that there is a distribution in the orientations of the particle axes within the specimen. This has the effect of spreading out the diffracted intensity from a reflection along Debye-Scherrer arcs. That is, the intensity at a given position in the particle transform is arced about the center of the pattern at a constant diffraction angle. The distribution of intensity in the X-ray diffraction pattern as a function of angle about the center of the pattern is related to the distribution of particle orientations within the specimen. At small angles of diffraction, a small degree of disorientation will not significantly affect data collection. However, at higher angles any disorientation will ultimately cause the diffraction on neighboring layer planes or lattice lines to overlap and this will limit the resolution to which reliable data can be collected.

The procedure generally used for measuring the intensity distribution along the lines of data in diffrac-
tion patterns from helical assemblies has been to measure the optical density on a film with a linear scan of a microdensitometer along the center of a layer line and to measure the background by measuring the region between layer lines (Holmes & Barrington Leigh, 1974; Stubbs, 1974; Fraser, MacRae, Miller & Rowlands, 1976). Where disorientation causes adjacent layer lines to overlap, it is extremely difficult to make quantitative measurements of background by conventional procedures. This problem can be dealt with to some extent by the use of an interactive display system to determine approximate base lines for each reflection (Mandelkow, 1973). Fraser, MacRae, Miller & Rowlands (1976) have discussed the advantages of making use of all the diffraction data on an X-ray film, and briefly described a possible method for dealing with overlapping layer lines. Lovell & Windle (1977) discuss an azimuthal sharpening procedure useful in very poorly oriented specimens. Holmes & Barrington Leigh (1974) pointed out that the general problem of correcting for disorientation can only be treated by a numerical deconvolution procedure.

In this paper the implementation of a numerical deconvolution procedure making use of all the data on an X-ray film is described, and its application to several systems of biological interest is demonstrated. All the X-ray intensities measured at a given radius in the diffraction pattern are used to determine the background and intensity of each reflection falling at that radius. The data processing is very sensitive to the quality of the experimental measurements and can only be used reliably on carefully collected diffraction data. Although data from one- and two-dimensionally periodic assemblies can be treated using these methods, for simplicity the discussion throughout the remainder of the paper will assume that the object of interest is a one-dimensional array. The practical problems and limitations of the method are discussed. Its application to the measurement of diffraction from filamentous bacteriophage P1, from tobacco mosaic virus, and from microtubules is described. The limiting resolution to which diffraction data can be accurately collected is determined from the degree of disorientation and repeat period of the specimen.

**Theory**

Several papers have been published on the form and measurement of diffraction from partially oriented specimens (Deas, 1952; Holmes & Barrington Leigh, 1974; Stubbs, 1974; Fraser, MacRae, Miller & Rowlands, 1976). The purpose of this study is not to extend the previously published theory, but to apply the theory to develop a usable method of data collection. In the sections that follow, it will be assumed that the X-ray diffraction data were collected on flat films with a point-focus camera and that the order in the specimen is such that the layer lines can be considered infinitely thin. Diffraction data collected under other conditions can often be corrected such that these restrictions apply. For instance, data can be deconvoluted for the effect of beam height (e.g. Lake, 1967) and positions on films collected with other geometries can be readily transformed into positions that they would have had on a flat film (e.g. Vainstein, 1966; Fraser, MacRae, Miller & Rowlands, 1976).

In general, the form of the optical density \( D(r, \phi) \) measured on a flat film on which an X-ray diffraction pattern from a partially oriented specimen has been recorded can be expressed in film coordinates as the sum of contributions from all reflections \( f_i(r, \phi_i) \) plus a background \( B(r, \phi) \).

\[
D(r, \phi) = \sum_i f_i(r, \phi_i) + B(r, \phi),
\]  
(1)

where \( r \) is the distance from the center of the diffraction pattern and \( \phi \) is the angle about the center of the
diffraction pattern. The intensity distribution function \( f(\phi, \varphi) \) describes the spreading out of the intensity as a function of angle \( \phi \) from a reflection falling at \((r, \varphi)\), which causes it to contribute to the optical density measured at \((r, \varphi)\). For instance, in a completely disoriented specimen, the intensity distribution function is a constant function of \((\varphi - \varphi_0)\);

\[
D(r, \varphi) = \sum I_i(r, \varphi_i) + B(r, \varphi) \quad \text{(complete disorientation)}
\]

and in a perfectly oriented specimen it is a delta function, \( \delta(\varphi - \varphi_0) \);

\[
D(r, \varphi) = \sum I_i(r, \varphi_i) \delta(\varphi - \varphi_i) + B(r, \varphi) \quad \text{(perfect orientation)}.
\]

However, in general, the intensity distribution function is not related in a simple manner to the distribution of particle orientations in the specimen.

**Relation between particle orientation and intensity distribution**

The distribution of particle orientations about some common axis can be defined in terms of a disorientation function \( N(\varphi) \), where \( N(\varphi) \) d\( \varphi \) is the probability of the axis of a particle being at an angle \( \varphi \) to the common axis in an element of solid angle d\( \varphi \). The intensity distribution on the film from a single reflection, \([ I_i(r, \varphi_i) f(\varphi, \varphi_0) ] \), can be calculated by an integral over all possible particle orientations which will contribute to the intensity at \((r, \varphi)\). For a Gaussian distribution of particle orientations,

\[
N(\varphi) = \frac{2}{\sigma^2} \exp \left( -\frac{\varphi^2}{2\sigma^2} \right). \tag{2}
\]

Holmes & Barrington Leigh (1974) showed that, except in the immediate vicinity of the meridian, the intensity distribution function was approximately equal to a Gaussian;

\[
f(\varphi, \varphi_0) = \frac{1}{R_i} \frac{1}{2\pi \sigma^2} \exp \left[ -\frac{(\varphi - \varphi_0)^2}{2\sigma^2} \right]. \tag{3}
\]

where \( R_i \) is the distance from the meridian to the reflection. For this case, the intensity distribution in the pattern is simply a function of \((\varphi - \varphi_0)\) and (1) can be written as a convolution;

\[
D(r, \varphi) = \sum I_i(r, \varphi_i) f(\varphi - \varphi_i) + B(r, \varphi). \tag{4}
\]

This equation will apply whenever the shape of a reflection as a function of \( \varphi \) is independent of \( \varphi \); that is, its angular profile will be the same whether the reflection falls on the equator, near the meridian or somewhere in between.

In order to test the approximation that \( f(\varphi, \varphi_0) \) can be treated simply as \( f(\varphi - \varphi_0) \) and determine the limits of its validity, the intensity distribution function was calculated exactly for one simple case. For a distribution of particle orientations

\[
N(\varphi) = \begin{cases} 
\frac{1}{2\pi(1 - \cos \sigma)} & \sigma \leq \sigma \\
0 & \sigma > \sigma
\end{cases} \tag{5}
\]

the intensity distribution function can be derived in closed form. This form for the disorientation function was used by Fraser, MacRae, Parry & Suzuki (1971) to derive the disorientation correction factor for fiber diffraction patterns. Although this disorientation function is not physically plausible, variations of the intensity distribution resulting from this form of disorientation in different regions of the film are easily recognized and provide us with a measure of the variation in the angular profiles of reflections as a function of position on the film. The intensity at any point on the film will depend on the number of particles contributing to that point. As shown in Fig. 2, this is proportional to the length of the arc \( M_1M_2 \) on the circle with center at point \( O' \) radius \( q \) sin \( \varphi_i \), where \( q \) is

![Fig. 2. Diagram illustrating the derivation of the intensity distribution function in (6). Consider a point \( P \) on a layer line at a radius \( q \) in reciprocal space, at an angle \( \varphi_i \) from the meridian. For the disorientation function \( N(\varphi) \) given in (5), which is a constant over an angular range \( \sigma \) and zero outside that range, the distribution of particle orientations in the specimen will smear this reflection out onto points \( P' \) within an angular \( \sigma \) of the point \( P \). All particles that contribute to the point \( P' \) must be oriented such that their axes fall on a cone at an angle \( \varphi_i \) to the point \( P' \) with vertex at the origin, \( O \). The width of the layer line, \( \Delta \), will add contributions from neighboring cones over an angular range \( \Delta \varphi = \Delta / q \sin \varphi_i \). The proportion of particles contributing to the point \( P' \) can be calculated by a line integral on a circle about the point \( O' \) and falling on the surface of the sphere of radius \( q \) in reciprocal space. For the disorientation function chosen, this integral is particularly simple since the density of particle orientations is a constant over the shaded disk in the figure and zero elsewhere. The proportion of particles contributing to \( P' \) is equal to twice the length of the arc \( M_1M_2 \) times the width due to the layer line width, \( \Delta / q \sin \varphi_i \), divided by the area of the disk, \( 2\pi q^2 (1 - \cos \sigma) \). The length of the arc \( M_1M_2 \) about \( O' \) is \( q' \sin \varphi_i \), where \( q' \) is the angle subtended by the arc. \( \gamma \) is found from the spherical-triangle relation \( \cos \gamma = (\cos\varphi - \cos\varphi_0) sin\varphi_0 / sin\varphi_i \). Thus, the intensity is proportional to \( \cos^{-1}[(\cos\varphi - \cos\varphi_0) sin\varphi_0 / sin\varphi_i] / \pi q'^2 (1 - \cos \sigma) \). When \( \varphi \leq \varphi - \varphi_0 \), \( \gamma \) is always equal to \( \pi \), and the intensity is proportional to \( 1/4q'^2(1 - \cos \sigma) \).]
the distance from the origin of reciprocal space to the reflection at point $P$. The calculation as outlined in Fig. 2 gives:

$$f(\varphi, \varphi_i) = \begin{cases} 
\cos^{-1} \left( \frac{\cos \sigma - \cos \varphi \cos \varphi_i}{\sin \varphi \sin \varphi_i} \right) & \varphi > \sigma - \varphi_i \\
\frac{1}{\varrho A^{-1}(1 - \cos \sigma)} & \varphi \leq \sigma - \varphi_i,
\end{cases}$$

(6)

where $\varrho$ is the layer-line width. In Fig. 3, $f(\varphi, \varphi_i)$ is plotted for $\sigma = 5^\circ$ and $3^\circ$. The curves have been normalized to constant heights to demonstrate the shape changes. Only on the meridian is the intensity distribution function equal to the disorientation function, and once the curvature of the Ewald sphere is taken into account they are not equal anywhere on the film. In order to demonstrate the origins of the different effects, for $\sigma = 5^\circ$ the curves have been plotted with and without taking into account the curvature of the sphere of reflection. The curvature of the Ewald sphere has the effect of broadening the reflections near the meridian, and results in the very near meridional part of the Fourier transform being unobserved for well oriented fibers with axes perpendicular to the X-ray beam. Generally, the shape of the intensity distribution function is approximately constant for reflections at angles greater than $2\sigma$ from the meridian. The changes in shape of reflections near the meridian will be considerably less dramatic for physically reasonable disorientation functions which are smoothly varying. Except for near-meridional diffraction the approximation that the intensity distribution function is only a function of $(\varphi - \varphi_i)$ is valid, and (4) will apply to the data. This agrees with the results of Holmes & Barrington Leigh (1974) and Stubbs (1974). Experimental measurements made in this laboratory on fiber diffraction patterns from several types of specimens have been analyzed in a self-consistent way assuming that the intensity distribution function depends only on $(\varphi - \varphi_i)$.

Equation (4) represents a set of linear equations, one for each optical density measurement $D(r, \varphi)$. As long as the number of measurements at a given radius, $r$, exceeds the number of reflections at that radius and the forms of the intensity distribution function $f(\varphi - \varphi_i)$ and background $B(r, \varphi)$ are known, this equation set can be used to solve for the unknown intensities $I(r, \varphi_i)$. The intensity distribution function $f(\varphi - \varphi_i)$ can usually be measured in a region on an X-ray diffraction pattern where there is no overlapping of reflections. Holmes & Barrington Leigh (1974) showed that for highly oriented specimens where disorientation is largely due to thermal displacements from mean orientation, the intensity distribution function is Gaussian. Although there is no a priori reason for the disorientation to be Gaussian, the angular intensity distribution in many patterns that we have analyzed can be accurately approximated by a Gaussian distribution. Equation (4) then becomes

$$D(r, \varphi) = \sum I(r, \varphi_i) \exp \left[ -\left( \frac{\varphi - \varphi_i}{2\sigma^2} \right)^2 \right] + B(r, \varphi),$$

(7)

where $\sigma$ is the standard deviation of the angular disorientation. Other forms of the intensity distribution function can be used where appropriate.

The solution of (7) results in a set of measured intensities, $I(r, \varphi_i)$, which are equal to the background-subtracted optical densities at the centers of the layer lines; $I(r, \varphi_i)$ is the height of the Gaussian distribution
centered on the layer line at $\varphi_i$ and, where layer lines are not overlapping, this should correspond closely to the height of a layer line as measured by a linear densitometer trace after background subtraction. This can be seen, for instance, in the close correspondence between the calculated intensities and the densitometer traces near the meridian in Fig. 8. Thus, the calculated intensities must be corrected for geometric factors in the same way as intensities measured by conventional methods. The disorientation correction has been discussed by several authors (Franklin & Gosling, 1953; Langridge, Wilson, Hooper, Wilkins & Hamilton, 1960; Fraser, MacRae, Parry & Suzuki, 1971; Mandelkow, 1973; Holmes & Barrington Leigh, 1974; Stubbs, 1974; Fraser, MacRae, Miller & Rowlands, 1976). The problem involves the fact that because of disorientation a layer plane on which the intensity is a constant (as might occur in diffraction from a very thin, periodic structure) will not give rise to a layer line of constant observed intensity. For instance, for the particle disorientation function in (5), the exact correction factor has been calculated (see also Fraser, MacRae, Parry & Suzuki, 1971) and is plotted in Fig. 4. Outside a region within an angle of $2\sigma$ from the meridian, the correction factor is very nearly equal to $R$, the distance from the meridian. This is consistent with the calculations done by Stubbs (1974) for a Gaussian disorientation function. The disorientation correction has the same general form for any disorientation function, differing only in detail near the meridian. For non-crystalline fibers the correction factor approaches $R$ at distances far from the meridian and is approximately equal to $R$ at angles greater than $2\sigma$ from the meridian. Crystalline fibers must be treated somewhat differently (Franklin & Gosling, 1953; Stubbs, 1974).

![Fig. 4. Plot of the reciprocal of the intensity, $1/I$, that would be observed along a layer line from a specimen of very long, infinitely thin scattering objects (intensity a constant on any given layer plane) with orientations distributed as described by the disorientation function in (5). The curve was calculated from (6) with $\varphi = \varphi_i$ for a disorientation of $\sigma = 5^\circ$ and a layer line at $q = 0.1$ Å$^{-1}$. The broken line is equal to $R$ where $R$ is the distance from the meridian. When $\varphi_i > 2\sigma$, the disorientation correction is, to an excellent approximation, equal to $R$.]

Background

The background presents additional problems, for without knowledge of the form of the background, the problem is indeterminate. Two possible strategies are available for dealing with the background. It can be measured with a blank replacing the specimen, in which case it need not appear as an unknown in (7), or it can be expanded as a sum of orthogonal functions increasing the number of unknowns in (7). Background is due to film fog, X-ray scattering from air, sample holder, solvent and components of the camera. It can be approximately measured by taking an X-ray diffraction photograph under exactly the same conditions as used in obtaining the diffraction pattern, but without the sample. This is not always possible to obtain, and even under the best conditions the background measured in this way is not likely to completely account for the diffraction pattern background. However, subtraction of a blank from a diffraction pattern will improve the accuracy of the calculated intensities by minimizing any discrepancies between the theoretical form of the background and the actual background. When a blank cannot be used to account completely for the background, an analytical form of the background can be assumed; for instance, the background may be expanded as a sum of orthogonal functions such as a polynomial or Fourier series. Unless the exact form of the background is known, this expansion greatly increases the number of unknowns in the equations and lowers the reliability of the calculated intensities.

Except for regions close to the center of the diffraction pattern, it is often possible to assume that the background is a constant at a given distance from the center of the diffraction pattern; that is, $B(r, \varphi) = B(r)$. This is the assumption that has been used in most of the film processing described below. With this assumption, (7) becomes simply

$$D(r, \varphi) = \sum_{i=1}^{n} I_i(r, \varphi_i) \exp \left[ -((\varphi - \varphi_i)^2/2\sigma^2) \right] + B(r). \quad (8)$$

There are $(n + 1)$ unknowns in this equation where $n$ is the number of reflections contributing at radius $r$ on the film.

Calculation of intensities

There are several ways of solving a set of linear equations such as those represented by (8). Fourier methods (Stokes, 1948), iterative methods (Lovell & Windle, 1977) or matrix inversion may be used to solve (8). Each of these methods has certain advantages and all three have been implemented in this laboratory. The Fourier method is very useful for partial deconvolution which can be informative for extremely disoriented specimens. However, because a quantitative evaluation of the amount of information available at a given radius on the film can be calculated using
matrix inversion, it has been used in all the examples described here. In matrix notation, (8) becomes

$$D = Fl, \tag{9}$$

where $D$ is the vector of observed optical densities at a given radius, $r$, $F$, is the disorientation matrix and $I$ is the vector of unknowns at radius $r$ that includes the parameters describing the background. Formally, the least-squares solution to this equation can be written as

$$I = (FF)^{-1}F^TD \tag{10}$$

where $F'$ is the transpose of the matrix $F$, and $(FF)^{-1}$ is the inverse of the symmetric, square matrix $FF$. The reliability of this solution can be tested by a calculation of the eigenvalues $\lambda$ of the symmetric matrix $FF$. When errors in the measurement of the optical densities $D(r, \phi)$ are independent and normally distributed with standard deviation $\sigma^2$, the standard deviation of errors in the calculated intensities, $\delta^2$, is given by (Crowther, DeRosier & Klug, 1970)

$$\delta^2 = \frac{1}{\lambda} \sigma^2 \tag{11}$$

where the angle brackets indicate average value. Thus if $\langle 1/\lambda \rangle < 1$, the uncertainties in the calculated intensities will be less than those of the measured optical densities. The effect of increasing $\langle 1/\lambda \rangle$ can be observed in the examples presented below. At small diffraction angles, $\langle 1/\lambda \rangle$ is small and the noise in the intensities calculated from (8) is lower than in the intensities measured by a densitometer trace. At higher diffraction angles, the layer lines are closer together (as a function of angle $\phi$) and the value of $\langle 1/\lambda \rangle$ is larger. This has the effect of greatly increasing the noise level in the intensities calculated from (8). Thus, at high diffraction angles the noise level in a linear densitometer trace is much less accurate.

Methods

Optical densities on films to be processed are measured on an Optronics rotating-drum microdensitometer. The optical densities are measured on a square raster over the entire region of interest on the film and stored on magnetic tape. The magnetic-tape file is transferred to a computer disk for further processing on a Digital PDP-10. The optical densities must be placed in a polar coordinate system in order to allow solution of (8). To do this, each optical density is assigned a position ($x$, $y$) relative to the center of the diffraction pattern with the equator and meridian chosen as the principal axes. The optical densities in small regions ($\Delta r$, $\Delta \phi$) are then averaged

$$D(r, \phi) = \frac{1}{n} \sum_{j,k} D(x_j, y_k), \tag{12}$$

for all $(x_j, y_k)$ within $(r \pm \Delta r/2, \phi \pm \Delta \phi/2)$ where $n$ is the number of data points in the sum. For specimens with fiber axis perpendicular to the X-ray beam, data from all four quadrants can be averaged. Note that at higher diffraction angles, the region of the film averaged to obtain $D(r, \phi)$ will be larger. This averaging improves the signal-to-noise ratio at high diffraction angles where the signal will be weaker.

In order to assign a coordinate pair $(x, y)$ to each data point, the center of the diffraction pattern must be found accurately. This is usually done by locating symmetrically equivalent regions on the film (see also Fraser et al., 1976). The orientation of the film relative to the raster of data points can be determined by plotting the optical density at a given radius as a function of the angle $\phi$ about the center of the diffraction pattern, and comparing positions of equivalent reflections at that radius.

The intensities are calculated from the optical density data at a given radius using (8). At each radius, the positions, $\phi_i$, of all contributing layer lines are calculated. The matrix $F$ is then calculated from the form of the intensity distribution function and background parameters. Once the intensities are determined, the optical densities, $D$, can be recalculated.

$$D_{\text{calc}} = F_{\text{calc}},$$

and compared with the observed data $D$. If the recalculated data do not agree with the observed data, the assumptions made in the process, for instance the form of the background or the angular width, $\sigma$, of the intensity distribution function or the calculated layer-line positions, $\phi_i$, must be questioned. The residual from this equation can be used to improve the estimates of $\sigma$ and $\phi_i$ where necessary.

In order to process a film successfully using the angular deconvolution procedures described here, all the optical densities on the portion of the film to be processed must be due to sources described by (4); that is, to reflections found at positions $(r, \phi_i)$ smeared out by an intensity distribution function $f(\phi - \phi_i)$ plus a background that can be described as a constant or a small number of analytical terms. Some preliminary processing of a data set may be necessary to put the data in a form suitable for this calculation. For instance, correction for beam width, or subtraction of a blank to remove background intensity. Films with small defects such as scratches or peaks due to crystalline salt in the specimen or a non-monochromatic beam are usually not suitable for processing since any defect at a radius, $r$, will affect the calculated intensities of all the reflections at that radius.

Results

X-ray diffraction patterns from oriented specimens of tobacco mosaic virus, filamentous bacteriophage P1, and microtubules are shown in Fig. 5. The three photographs are reproduced on the same scale in reciprocal space.
The pattern from tobacco mosaic virus in Fig. 5(c) is from an exceptionally well oriented specimen. The layer lines correspond to an axial repeat of 69 Å. The standard deviation of the disorientation in this specimen is about 1.4°, and the layer lines are well separated to at least 5 Å spacing. At higher angles the layer lines are overlapping.

Fig. 5. Fiber diffraction patterns from three partially oriented specimens. (a) is a diffraction pattern of an oriented specimen of microtubules obtained by J. Thomas of Brandeis University. The standard deviation of the orientation in this specimen is about 9°. The principal layer lines are widely separated, the repeat period being about 40 Å compared with 75 Å in the filamentous bacteriophage and 69 Å in tobacco mosaic virus. The larger separation of the layer lines allows the data from this specimen to be collected to at least 10 Å. (b) is a diffraction pattern from a fiber of filamentous bacteriophage P22 taken by Dr. W. C. Phillips at Brandeis University using a fiber provided by Dr. D. Marvin of the European Molecular Biology Laboratory in Heidelberg, Germany. This pattern was taken using one of the best oriented fibers of P22. Its orientation has a standard deviation of 4.2°. The first four layer lines in this pattern are not well separated in the region of strong diffraction at about 10 Å spacing. (c) is a diffraction pattern from an oriented gel of tobacco mosaic virus taken by Dr. S. Warren and Dr. G. Stubbs at the Max-Plank-Institut in Heidelberg, Germany. Tobacco mosaic virus solutions at concentrations of about 25% produce some of the best oriented non-crystalline specimens ever used in diffraction studies. The standard deviation of the orientation in this gel is about 1.4°. The layer lines are distinct and clearly separated out to at least 5 Å spacing. However, even in this pattern, at higher angles of diffraction the layer lines are overlapping.
lines begin to overlap and there is significant contribution to the background from water in the specimen and the quartz capillary containing the specimen. This pattern was taken on a Guinier camera, and the arcs which the reflections are smeared into by the disorientation are not circularly symmetric.

Specimens of Pfl are more difficult to orient. The standard deviation of disorientation in the specimen shown in Fig. 5(b) is about 4.2°. The closely spaced layer lines (1/4 Å⁻¹) overlap even in the region of strong diffraction at about 10 Å spacing. Diffraction from the weak second layer line is completely obscured by

---

Fig. 6. Calculated and observed diffraction from X-ray patterns from microtubules, filamentous bacteriophage and tobacco mosaic virus plotted as a function of angle about the center of the diffraction pattern at a constant radius. The diagrams at the top represent the circular paths along which the data are plotted and the positions of the layer lines intersecting these paths. The data for microtubules are plotted for spacings of 15 and 10 Å, and the data for filamentous bacteriophage and tobacco mosaic virus are plotted for spacings of 15, 10, and 7 Å. The tobacco mosaic virus data are plotted as discrete circles at 1° intervals, the others at 3° intervals. The solid lines are the recalculated optical-density data from the set of intensities for the layer lines convoluted with a Gaussian intensity distribution function and added on to a constant background. For the tobacco mosaic virus data, an additional background parameter was calculated to account for the diffuse scatter from the quartz capillary, which is not circularly symmetric in the Guinier camera used. For each plot the average of the reciprocal of the eigenvalues, \(1/\lambda\), is given. The high value for tobacco mosaic virus at 7 Å is due to the presence of two background parameters in the equations.
overlap from the strong first and third layer lines in this region. Sharp reflections on the equator are due to the hexagonal packing of virions in the fiber. Weak inter-particle interference is also evident on the third layer line.

The principal layer lines in fiber diffraction patterns from microtubules are separated by ~1 Å. Although in most specimens the orientation is poorer than that of Pfl, the layer lines are reasonably well separated out to about 10 Å spacing as seen in the diffraction pattern in Fig. 5(a). In this pattern, the standard deviation of disorientation is about 9°.

Fig. 6 shows plots of the optical densities on these three diffraction patterns as a function of angle, φ, about the center of the patterns at different diffraction angles. The diagrams at the top show the paths followed by the plots on the diffraction patterns and the layer lines which these paths intersect. The measured optical densities are marked as open circles and the solid lines are the optical densities calculated by convoluting the calculated intensities with the intensity distribution function. Plots are shown for microtubules at 15 and 10 Å spacings and for Pfl and TMV at 15, 10, and 7 Å spacings. On each plot, the average of 1/2 of the matrix which was inverted to perform the deconvolution is given. The angular profiles of the reflections remain the same at different spacings but, as a function of angle, the reflections get closer together at higher diffraction angles.

Fig. 7 contains the results of the angular deconvolution of the filamentous bacteriophage data for the equator and first five layer lines of the diffraction pattern shown in Fig. 5(b), and the average (1/λ) as a function of distance from the center of the pattern.

Fig. 8. A comparison of the intensities calculated by angular deconvolution with densitometer traces not employing the deconvolution procedure for the first three layer lines of the diffraction pattern from Pfl shown in Fig. 5(b). In each case the lower plot is that calculated by angular deconvolution. The baselines of the densitometer traces have been adjusted to facilitate comparison of the pairs. The baselines of the calculated intensities are determined by the deconvolution. From these plots, it is clear that the second layer line could not be measured by any conventional procedure because of the large contribution from the first and third layer lines in that region. The baselines of the first and third layer lines would be difficult to determine by conventional densitometry because of the increase in background between 15 and 7 Å spacing. The variation in noise level as a function of diffraction angle is apparent in these plots. The noise level on the traces calculated by angular deconvolution continually increases with diffraction angle because of the increase in (1/λ). The noise level in the densitometer traces decreases with increasing diffraction angle because they were calculated as polar scans. That is, each point was determined by averaging over a specified range (Δr, Δφ) on the film where Δφ was centered on the layer line. For these plots, AΔr is 100 μm and AΔφ is 3°. At higher diffraction angles, a larger portion of the film is averaged over. This form of densitometry is useful where layer lines are not overlapping because near the meridian conventional scans often smear adjacent portions of the layer line data into one another. At higher diffraction angles, the signal-to-noise ratio using this type of scan is more favorable than using linear scans.
For diffraction data at spacings between 10 and 7 Å, the $\langle 1/\lambda \rangle$ rises steeply. In this same region, the noise in the plots of the deconvoluted intensities along the layer lines also increases. Since the intensity along a layer line is a smoothly varying function, the increase in noise of these plots reflects the lowered reliability of the deconvolution with increased $\langle 1/\lambda \rangle$. Since intensities at each radius are calculated independently of those at other radii, correlation of intensities calculated on adjacent radii can be used to extend the resolution of the calculated data.

Fig. 8 shows a comparison of the intensities calculated by angular deconvolution with data collected by a densitometer trace for the first three layer lines of Pfl. The data from the densitometer traces are plotted just above the calculated intensities with an arbitrary choice of baseline for ease of comparison. The angular deconvolution allows calculation of the second layer line, which in linear traces is completely dominated by the contribution from the first and third layer lines. It also allows an accurate background to be determined for all layer lines.

In processing of data at high resolutions, a somewhat more elaborate scheme than that described here can be used. Tobacco mosaic virus is a helical virus with approximately 49.01 subunits in three turns of a helix. This slight deviation from an integral number of units in three turns causes the layer lines to split at high diffraction angles (Franklin & Klug, 1955). This splitting causes a shift in the apparent angular position and angular half-width of the layer lines. In order to process this data, the constraints on the layer line width, $\sigma$, and position, $q$, were relaxed slightly, allowing them to vary by up to 0.5°. The set of non-linear equations resulting from this formulation were solved by alternating a simple search procedure with three iterations of the matrix inversion, assuming new values for layer-line positions and widths on each iteration. This resulted in considerable improvement in the agreement with the data, as would be expected because of the increase in the number of variable parameters, and an improvement in the accuracy of the calculated intensities. Also, the apparent shift in position and width of a layer line provides a measure of the distribution of intensity among the two or more azimuthal (Bessel function) terms contributing to each layer line.

From the calculations which have been made on a variety of macromolecular systems, an estimate can be made of the resolution to which data can be reliably collected from a non-crystalline specimen. The highest spacing (in Å) to which data can be measured on a diffraction pattern from a specimen with a (one-dimensional) repeat distance, $c$, and standard deviation of disorientation, $\sigma$, is approximately $c/[\sin \{1.5\sigma\}]$, where the limit of reliable measurement has been taken to be the spacing where $\langle 1/\lambda \rangle = 1$. Correlation of intensities calculated independently at neighboring radii may allow some extension of this resolution limit.

### Discussion

The numerical deconvolution procedure described here is designed to produce estimates of diffracted intensity from partially oriented specimens to the highest possible resolution. The method requires that the form of the background be known and that the angular intensity distribution for a reflection be approximately independent of the position of the reflection on the film. It has been shown that in many cases both of these requirements are met by diffraction data from partially oriented specimens. The background from the X-ray cameras used in this laboratory is very nearly circularly symmetric except near the center of the diffraction pattern. Calculations of the angular intensity distribution for reflections in different parts of the film show that the intensity distribution function is approximately the same for all reflections outside the immediate neighborhood of the meridian. Angular deconvolution of data from several different types of specimen have shown that within the limitations of the approximations made, a set of layer-line intensities can be calculated which, when convoluted with the intensity distribution function, reproduces the data to within experimental error.

The object of an X-ray diffraction study is to construct a model for the structure of a structural unit. This model must be able to account for all the observed diffraction. In order to accomplish this it may be necessary to specify the average structure, variations in the structure, packing relations between the structural units and variability in the orientation of the units. The data may then be recalculated, taking into account instrumental smearing effects and background due to scattering from elements of the camera and specimen holder. The correspondence of this recalculated data to the observed pattern represents the ultimate test of a physically reasonable structural model.

It is possible, in principle, to make a rigorous test of a structural model taking into account structural variations, disorder and disorientation. In a rigorous test, the data would be recalculated without using approximations. For instance, rather than using an approximate intensity distribution function, a disorientation function would be found which could be used to recalculate the data, not by a convolution, but using the exact relations which depend on both $\phi$ and $\psi$. Such laborious exact calculations have not been necessary to process accurately the diffraction patterns described here.

Generally, it is sufficient to use the raw data, the optical densities on a film, to calculate a set of crystal reflection or layer line intensities. In processing of X-ray diffraction data from partially oriented specimens this step in data reduction is very important. Once a set of layer-line intensities has been collected, there may be no further reference to the raw data during a structural study. The angular deconvolution at-
tempts to extract as much information as possible from a diffraction pattern and the raw data is utilized to the limits imposed by the signal-to-noise ratio. This warrants the detailed comparison of data and re-calculated data which has been undertaken here.

Some extension of the limits on the resolution to which data can be collected may be possible. Because of the finite size of a particle in the specimen, the continuous intensity distribution must be smoothly varying along a layer line which does not exhibit interference effects due to correlation in particle positions. For a particle of diameter, \( d \), the convolution of the intensity distribution along a layer line with the function \( \sin(2\pi d R)/2\pi d R \) is equal to the intensity distribution itself. Thus, intensities calculated at positions along the layer line much closer together than \( 1/2d \) are highly correlated. If errors in the data at adjacent radii are independent then the noise in the intensity distributions such as that seen at high angles in Fig. 8 can be removed by convoluting the intensities with the function \( \sin(2\pi d R)/2\pi d R \). At sufficiently high diffraction angles the noise will be large compared to the intensity and this procedure must also fail. At higher diffraction angles only qualitative separation of intensities on adjacent layer lines may be possible. These may be obtained by using a partial deconvolution applying Stokes method or an iterative procedure.

The major problem in structural studies of oriented, non-crystalline specimens is the paucity of data due to cylindrical averaging of intensities. The angular deconvolution of data allows collection of the maximum amount of information from a given diffraction pattern. Further, it defines the limiting resolution to which accurate intensity information can be obtained. Its application has proven feasible for a number of macromolecular systems with widely varying degrees of disorientation. The calculations significantly improve the accuracy of data measured compared with conventional methods and extends the resolution to which data from these specimens can be reliably collected.

I would like to thank Dr. K. C. Holmes, Dr. G. Stubbs and Dr. S. Warren for the diffraction patterns of tobacco mosaic virus, Dr. C. Cohen and J. Thomas for the diffraction patterns of microtubules, Dr. D. Marvin for providing a fiber of filamentous bacteriophage P1 and Dr. W. Phillips for taking many of the diffraction photographs used in this study. Much of the computer processing was done with the help of M. Hagghooie. The work was aided by very helpful discussions with Dr. D. J. DeRosier. I would particularly like to thank Dr. D. L. D. Caspar; his advice, encouragement and friendship have been essential throughout the course of this work.

The work was supported by a National Institutes of Health postdoctoral fellowship to the author and by National Institutes of Health grant No. CA15468-04 and National Science Foundation grant No. PCM77-16271 to Dr. D. L. D. Caspar.

References