Acta Cryst. (1987). A43, 45-49

X-ray Scattering from Helical Structures Possessing Random Variable Twist

By Richard Barakat

Division of Applied Sciences, Harvard University, Cambridge, MA 02138, USA

(*Received* 18 *July* 1985; *accepted* 27 *May* 1986)

Abstract

The Egelman-DeRosier model of X-ray scattering from a helical structure possessing cumulative random twist is studied. The present analysis assumes that the random rotations between subunits are zero-mean uncorrelated Gaussian random variables whose variances are small. The average scattered intensity is obtained in closed form for an arbitrary number of scattering subunits, and also when the number of scattering subunits is itself a random variable governed by a Poisson distribution. When the number of scattering subunits is large, the probability density function of the scattered intensity at a given layer line is obtained as the product of a negative exponential probability density and an infinite series of modified Bessel functions.

1. Introduction

Egelman, Francis & DeRosier (1982, 1983) have shown that the F-actin helix can be described by a constant rise per subunit but with a random twist. Egelman & DeRosier (1982) developed the formal aspects of the model with respect to X-ray scattering and carried out calculations of the average scattered X-ray intensity using analogies with polymer statistics, backed up by Monte Carlo simulations. They point out that their model may be applicable to many helical structures.

The purpose of the present communication is an *ab initio* study of the X-ray scattering model in view of its importance. We obtain an expression for the probability density function of the scattered intensity at the layer lines when the number of scattering subunits N is large. In addition, the average scattered intensity is obtained in closed form for *arbitrary* N, and when N itself is a random variable governed by a Poisson distribution.

According to Egelman & DeRosier (1982), the complex amplitude due to scattering from N subunits which gives rise to layer-line intensities is

$$C(n,Z)=\sum_{j=1}^{N}\exp\left\{i[j\Delta\psi(pZ-n)]-in\sum_{i=1}^{i}\delta_i\right\} \quad (1.1)$$

where $Z=l/c$ (l = layer-line number, c = helical repeat) and $j\Delta\psi=\psi_{j-1}$, with ψ_j the polar angle at the jth subunit. n is the order associated with the azimuthal symmetry of the corresponding helix ($n=0,\pm1,\pm2,\ldots$) and p is the pitch of a helix corresponding to $n=1$. The random variable δ_t describes the rotation between subunits, which random variation is cumulative.

The δ_t are taken to be zero-mean uncorrelated Gaussian random variables

$$\begin{aligned}\langle\delta_t\rangle&=0 && \forall t\\ \langle\delta_s\delta_t\rangle&=\langle\delta^2\rangle, && s=t\\ &=0, && s\neq t\end{aligned} \quad (1.2)$$

where $\langle\delta^2\rangle$ *is generally small.* Note that Egelman & DeRosier (1982) do not make the Gaussian assumption.

At the layer lines, *i.e.*

$$\Delta\psi(pZ-n)=2\pi m, \qquad m=0,\pm1,\pm2,\ldots, \quad (1.3)$$

the sample realizations of $C(n,Z)$ are now given by

$$C=\sum_{j=1}^{N}\exp\left[-in\sum_{t=1}^{j}\delta_t\right]. \quad (1.4)$$

This is equation (4) of Egelman & DeRosier (1982); it is the basic equation for the subsequent analysis. For convenience we set

$$\theta_j\equiv n\sum_{t=1}^{j}\delta_t \quad (1.5)$$

so that $C=U-iV$ where

$$U=\sum_{j=1}^{N}\cos\theta_j,\qquad V=\sum_{j=1}^{N}\sin\theta_j. \quad (1.6)$$

Note that C itself is not physically observable; its absolute square (scattered intensity)

$$I=U^2+V^2 \quad (1.7)$$

is the basic observable.

The ensemble averaging over the δ_t in (1.5) can be carried out explicitly *via* the formula (Thomas, 1981)

$$\left\langle\exp\left(\pm in\sum_{t=m_1}^{m_2}a_t\delta_t\right)\right\rangle=\exp\left(-\frac{n^2\langle\delta^2\rangle}{2}\sum_{t=m_1}^{m_2}a_t^2\right) \quad (1.8)$$

where the a_t are deterministic constants. This expression is simply the multivariate characteristic

0108-7673/87/010045-05$01.50

function for zero-mean uncorrelated Gaussian random variables. In addition, since the probability density functions of the sums of δ_t are also zero-mean Gaussian, it follows that they are symmetric so that

$$\left\langle \cos\left(n \sum_{t=m_1}^{m_2} a_t\delta_t\right)\right\rangle = \exp\left(-\frac{n^2\langle\delta^2\rangle}{2}\sum_{t=m_1}^{m_2} a_t^2\right)$$

$$\left\langle \sin\left(n \sum_{t=m_1}^{m_2} a_t\delta_t\right)\right\rangle = 0. \quad (1.9)$$

2. Preliminaries

The random variables U and V, themselves being sums of random variables, tend toward normality as N increases by virtue of the central limit theorem [see Thomas (1981)]. Consequently the joint probability density function of U and V is given by the two-dimensional correlated Gaussian

$$f_{U,V}(U, V) = [2\pi\sigma_U\sigma_V(1-\rho^2)]^{-1} \times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{U-\langle U\rangle}{\sigma_U}\right)^2 - 2\rho\left(\frac{U-\langle U\rangle}{\sigma_U}\right)\left(\frac{V-\langle V\rangle}{\sigma_U}\right) + \left(\frac{V-\langle V\rangle}{\sigma_V}\right)^2\right]\right\} \quad (2.1)$$

where

$$\rho \equiv \frac{(\langle UV\rangle - \langle U\rangle\langle V\rangle)}{\langle U\rangle\langle V\rangle} \quad (2.2)$$

$$\sigma_U^2 = \langle U^2\rangle - \langle U\rangle^2, \qquad \sigma_V^2 = \langle V^2\rangle - \langle V\rangle^2. \quad (2.3)$$

We now proceed to determine the various terms in (2.1), beginning with $\langle U\rangle$ and $\langle V\rangle$. Thus

$$\langle U\rangle = \sum_{j=1}^{N}\langle\cos\theta_j\rangle = \sum_{j=1}^{N}\beta^j \quad (2.4)$$

where

$$\beta \equiv \exp(-n^2\langle\delta^2\rangle/2) < 1. \quad (2.5)$$

In deriving (2.4) and (2.5) we used (1.9). The series in (2.4) is a geometric series

$$1+\beta+\beta^2+\ldots+\beta^{N-1} = (1-\beta^N)/(1-\beta) \quad (2.6)$$

and we easily obtain

$$\langle U\rangle = [\beta/(1-\beta)](1-\beta^N). \quad (2.7)$$

In addition

$$\langle V\rangle = \sum_{j=1}^{N}\langle\sin\theta_j\rangle \equiv 0 \quad (2.8)$$

from (1.9).

The term $\langle UV\rangle$ can be expressed as

$$\langle UV\rangle = \tfrac{1}{2}\sum_{j=1}^{N}\langle\sin 2\theta_j\rangle + \sum_{j=1}^{N-1}\sum_{k=j+1}^{N}[\langle\sin(\theta_j-\theta_k)\rangle + \langle\sin(\theta_j+\theta_k)\rangle] = 0. \quad (2.9)$$

All terms vanish because the sin terms average to zero by virtue of (1.9). Consequently $\rho \equiv 0$.

The second moments of U and V are

$$\langle U^2\rangle = \sum_{j=1}^{N}\langle\cos^2\theta_j\rangle + \sum_{j=1}^{N-1}\sum_{k=j+1}^{N}[\langle\cos(\theta_j-\theta_k)\rangle + \langle\cos(\theta_j+\theta_k)\rangle] \quad (2.10)$$

$$\langle V^2\rangle = \sum_{j=1}^{N}\langle\sin^2\theta_j\rangle + \sum_{j=1}^{N-1}\sum_{k=j+1}^{N}[\langle\cos(\theta_j-\theta_k)\rangle - \langle\cos(\theta_j+\theta_k)\rangle]. \quad (2.11)$$

The single sums are

$$\sum_{j=1}^{N}\langle\cos^2\theta_j\rangle = \tfrac{1}{2}\sum_{j=1}^{N} + \tfrac{1}{2}\sum_{j=1}^{N}\langle\cos 2\theta_j\rangle = \tfrac{1}{2}N + \tfrac{1}{2}\sum_{j=1}^{N}\beta^{2j} = \tfrac{1}{2}N + \tfrac{1}{2}\left(\frac{\beta^2}{1-\beta^2}\right)(1-\beta^{2N-2}), \quad (2.12)$$

$$\sum_{j=1}^{N}\langle\sin^2\theta_j\rangle = \tfrac{1}{2}\sum_{j=1}^{N} - \tfrac{1}{2}\sum_{j=1}^{N}\langle\cos 2\theta_j\rangle = \tfrac{1}{2}N - \tfrac{1}{2}\left(\frac{\beta^2}{1-\beta^2}\right)(1-\beta^{2N-2}). \quad (2.13)$$

The double sums are

$$\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}\langle\cos(\theta_j-\theta_k)\rangle = \sum_{j=1}^{N-1}\beta^{-j}\sum_{k=j+1}^{N}\beta^k = \left(\frac{\beta}{1-\beta}\right)(N-1) - \left(\frac{\beta}{1-\beta}\right)^2(1-\beta^{N-1}), \quad (2.14)$$

$$\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}\langle\cos(\theta_j+\theta_k)\rangle = \sum_{j=1}^{N-1}\beta^{3j}\sum_{k=j+1}^{N}\beta^k = \left(\frac{\beta}{1-\beta}\right)\left(\frac{\beta^4}{1-\beta^4}\right)(1-\beta^{4N-4}) - \left(\frac{\beta}{1-\beta}\right)\left(\frac{\beta^2}{1-\beta^2}\right)\beta^{N-1}(1-\beta^{2N-2}). \quad (2.15)$$

The double series over the β's were evaluated by repeated use of the finite geometric series, (2.6).

Note that the independent variables n and $\langle\delta^2\rangle$ enter into all the above expressions only through the product $n^2\langle\delta^2\rangle$.

We are now in possession of all the terms in (2.1).

3. Average scattered intensity

Before continuing let us determine the average scattered intensity. *We will (for the present) make no assumptions concerning the magnitude of N.* Now from (1.7)

$$\langle I\rangle=\langle U^2\rangle+\langle V^2\rangle. \tag{3.1}$$

Upon collecting the various terms in the previous section, we have

$$\langle I\rangle=\left(\frac{1+\beta}{1-\beta}\right)N-2\left(\frac{\beta}{1-\beta}\right)-2\left(\frac{\beta}{1-\beta}\right)^2(1-\beta^{N-1}). \tag{3.2}$$

When N is very large, this reduces to

$$\langle I\rangle\simeq[(1+\beta)/(1-\beta)]N \tag{3.3}$$

showing an N dependence in the presence of the random angular disorder.

The scattered intensity for an ideal helix ($\langle\delta^2\rangle\equiv 0$) behaves very differently. We have from (1.4)

$$I=N^2 \tag{3.4}$$

showing an N^2 dependence irrespective of the magnitude of N. At the other extreme where $\beta\to 0$, it follows from (3.2) that

$$\langle I\rangle=N \tag{3.5}$$

irrespective of the magnitude of N.

In Fig. 1, we show the behavior of $\langle I\rangle$ as a function of the number of scattering subunits N for some representative values of β. As N gets larger, the cumulative effect of the scattering relative to the ideal helix becomes evident. At $N=10^3$, the average intensity for $\beta=0{\cdot}9$ is almost two orders of magnitude smaller than that caused by the ideal helix. Even at $\beta=0{\cdot}99$, the difference is almost an order of magnitude.

In many situations, it is virtually impossible to know the exact number of δ_t's contributing to the scattered intensity; and it is of some interest to determine $\langle I\rangle$ when N itself is a discrete random variable governed by a probability distribution $P(N)$. It is a simple exercise in probability theory to prove that if N is a discrete random variable, then the first moment of the scattered intensity is

$$\langle I\rangle=\sum_{N=0}^{\infty}\langle I\,|\,N\rangle P(N) \tag{3.6}$$

where $\langle I\,|\,N\rangle$ is the first moment of I, given that N is deterministic [see (3.2)]. We assume that N is governed by a Poisson distribution

$$P(N)=(\langle N\rangle^N/N!)\exp(-\langle N\rangle) \tag{3.7}$$

where $\langle N\rangle$ is the first moment of N with respect to the Poisson distribution. Upon carrying out the necessary manipulations, we have

$$\begin{aligned}\langle I\,|\,\langle N\rangle\rangle &=\left(\frac{1+\beta}{1-\beta}\right)\langle N\rangle-2\left(\frac{\beta}{1-\beta}\right)\\ &\quad-2\left(\frac{\beta}{1-\beta}\right)^2\left[1-\frac{1}{\beta}\sum_{N=0}^{\infty}\frac{(\beta\langle N\rangle)^N}{N!}\exp(-\langle N\rangle)\right]\\ &=\left(\frac{1+\beta}{1-\beta}\right)\langle N\rangle-2\left(\frac{\beta}{1-\beta}\right)\\ &\quad-2\left(\frac{\beta}{1-\beta}\right)^2\{1-(1/\beta)\exp[-\langle N\rangle(1-\beta)]\}\end{aligned} \tag{3.8}$$

where we have made explicit the dependence on $\langle N\rangle$. When $\langle N\rangle$ is large

$$\langle I\,|\,\langle N\rangle\rangle\simeq[(1+\beta)/(1-\beta)]\langle N\rangle. \tag{3.9}$$

It is only for very small values of $\langle N\rangle$ (*i.e.* $\langle N\rangle<10$) that the mean intensities given by (3.3) and (3.9) vary significantly. Thus for all *practical* purposes when N or $\langle N\rangle$ is large, we can employ (3.3).

In some cases, it is useful to note that since

$$\frac{\exp(n^2\langle\delta^2\rangle/2)+1}{\exp(n^2\langle\delta^2\rangle/2)-1}=\coth(n^2\langle\delta^2\rangle/4) \tag{3.10}$$

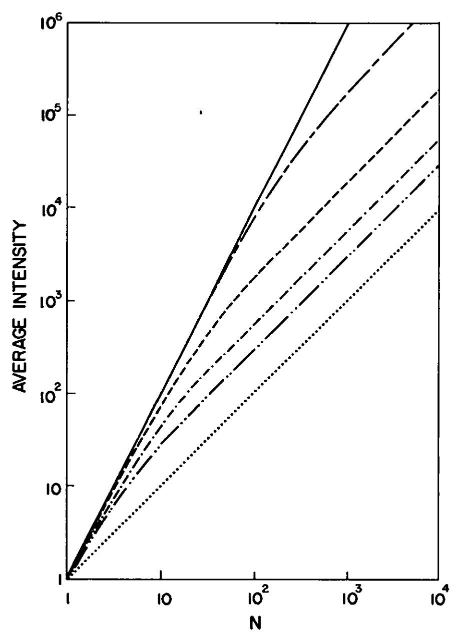

Fig. 1. Average scattered intensity as a function of N for various values of $\beta\equiv\exp(n^2\langle\delta^2\rangle/2)$. Solid line: $\beta=1$; broken curves: $\beta=0{\cdot}99$, $0{\cdot}90$, $0{\cdot}75$, $0{\cdot}50$ respectively. Dotted line: $\beta\simeq 0$.

we can rewrite (3.3) and (3.9) as

$$\langle I\,|\,N\rangle \simeq [\coth(n^2\langle\delta^2\rangle/4)]N \qquad (3.3a)$$

$$\langle I\,|\langle N\rangle\rangle \simeq [\coth(n^2\langle\delta^2\rangle/4)]\langle N\rangle. \qquad (3.9a)$$

When $n^2\langle\delta^2\rangle \ll 1$, $\langle I\,|\,N\rangle$ reduces to

$$\langle I\,|\,N\rangle \simeq (4/n^2\langle\delta^2\rangle)N \qquad (3.3b)$$

and when $n^2\langle\delta^2\rangle \gg 1$, $\langle I\,|\,N\rangle$ reduces to

$$\langle I\,|\,N\rangle \simeq [1-2\exp(-n^2\langle\delta^2\rangle)]N. \qquad (3.3c)$$

We note that (3.3b) was first obtained by Egelman & DeRosier (1982) using an analogy with polymer statistics.

4. Probability density of scattered intensity

We now resume determination of the probability density function of the scattered intensity when N is large. Upon returning to (2.1) with $\rho \equiv 0$, we transform to the joint density function of r and ψ, the envelope and phase

$$U = r\cos\psi, \qquad V = r\sin\psi. \qquad (4.1)$$

The final result is

$$f_{r,\psi}(r,\psi) = (r/2\pi\sigma_U\sigma_V)\exp[-R(r,\psi)] \qquad (4.2)$$

where

$$R(r,\psi) = \frac{(r\cos\psi - \langle U\rangle)^2}{2\sigma_U^2} + \frac{r^2\sin^2\psi}{2\sigma_V^2}. \qquad (4.3)$$

The probability density function of the envelope r follows by integrating out the phase; consequently

$$f_r(r) = \frac{r}{2\pi\sigma_U\sigma_V}\exp\left(-\frac{\langle U\rangle^2}{2\sigma_U^2}\right)\exp\left(-\frac{r^2}{2\sigma_V^2}\right) \times \int_0^{2\pi}\exp(-Sr^2\cos^2\psi + Tr\cos\psi)\,d\psi \qquad (4.4)$$

with

$$S \equiv (\sigma_V^2 - \sigma_U^2)/4\sigma_U^2\sigma_V^2, \qquad T \equiv \langle U\rangle/\sigma_U^2. \qquad (4.5)$$

To evaluate the integral, rewrite the exponent as

$$-Sr^2\cos^2\psi + Tr\cos\psi = -Sr^2/2 - (Sr^2/2)\cos 2\psi + Tr\cos\psi \qquad (4.6)$$

and use the expansion (Watson, 1947)

$$\exp[-(S^2r^2/2)\cos 2\psi] = \sum_{m=-\infty}^{\infty}(-1)^m I_m(S^2r^2/2)\exp(i2m\psi) \qquad (4.7)$$

where I_m is the modified Bessel function of the first kind. Hence

$$\int_0^{2\pi}(\,\cdot\,)\,d\psi = \exp(-Sr^2/2)\sum_{m=-\infty}^{\infty}(-1)^m I_m(Sr^2/2) \times \int_0^{2\pi}\exp(Tr\cos\psi + i2m\psi)\,d\psi. \qquad (4.8)$$

However (Watson, 1947),

$$\int_0^{2\pi}\exp(Tr\cos\psi + i2m\psi)\,d\psi = 2\pi I_{2m}(Tr) \qquad (4.9)$$

so that

$$\int_0^{2\pi}(\,\cdot\,)\,d\psi = 2\pi\exp(-Sr^2/2) \times \sum_{m=-\infty}^{\infty}(-1)^m I_m(Sr^2/2)I_{2m}(Tr). \qquad (4.10)$$

Upon collecting the various terms, we finally obtain the probability density of the envelope

$$f_r(r) = (r/\sigma_U\sigma_V)\exp(-\langle U\rangle^2/2\sigma_U^2) \times \exp[-(\sigma_U^2+\sigma_V^2)r^2/4\sigma_U^2\sigma_V^2] \times \sum_{m=0}^{\infty}\varepsilon_m(-1)^m I_m\left[\left(\frac{\sigma_V^2-\sigma_U^2}{4\sigma_U^2\sigma_V^2}\right)r^2\right] \times I_{2m}(\langle U\rangle r/\sigma_U^2) \qquad (4.11)$$

where $\varepsilon_m = 1$ for $m = 0$, $\varepsilon_m = 2$ for $m > 0$.

The probability density of the scattered intensity $I = r^2$ follows by the usual rules for transformation of variables in probability theory:

$$f_I(I) = (2\sigma_U\sigma_V)^{-1}\exp(-\langle U\rangle^2/2\sigma_U^2) \times \exp[-(\sigma_U^2+\sigma_V^2)I/4\sigma_U^2\sigma_V^2] \times \sum_{m=0}^{\infty}\varepsilon_m(-1)^m I_m\left[\left(\frac{\sigma_V^2-\sigma_U^2}{4\sigma_U^2\sigma_V^2}\right)I\right] \times I_{2m}[(\langle U\rangle/\sigma_U^2)\,I^{1/2}]. \qquad (4.12)$$

Returning to § 2, we may examine $\langle U\rangle$, $\langle U^2\rangle$ and $\langle V^2\rangle$ when N is large. We have from (2.7) that

$$\langle U\rangle \simeq [\beta/(1-\beta)] \qquad (4.13)$$

and from (2.10)-(2.15)

$$\langle U^2\rangle = \langle V^2\rangle \simeq \tfrac{1}{2}[(1+\beta)/(1-\beta)]N. \qquad (4.14)$$

It follows that

$$\begin{aligned}\sigma_V^2+\sigma_U^2 &\simeq [(1+\beta)/(1-\beta)]N = \langle I\rangle\\ \sigma_V^2-\sigma_U^2 &\simeq [\beta/(1-\beta)]^2 = \langle U\rangle^2.\end{aligned} \qquad (4.15)$$

Upon defining the parameter α,

$$\alpha = \langle U\rangle/\langle I\rangle^{1/2} \ll 1, \qquad (4.16)$$

we can rewrite (4.12) as

$$f_I(I) = \langle I\rangle^{-1}\exp(-I/\langle I\rangle)\exp(-\alpha^2) \times \sum_{m=0}^{\infty}\varepsilon_m(-1)^m I_m\left(\frac{\alpha^2 I}{\langle I\rangle}\right)I_{2m}\left[2\alpha\left(\frac{I}{\langle I\rangle}\right)^{1/2}\right]. \qquad (4.17)$$

The probability density function (PDF) of the corresponding normalized scattered intensity

$$h \equiv I/\langle I\rangle \qquad (4.18)$$

is

$$f_h(h) = \exp(-h)\exp(-\alpha^2) \times \sum_{m=0}^{\infty} \varepsilon_m(-1)^m I_m(\alpha^2 h) I_{2m}(2\alpha h^{1/2}). \quad (4.19)$$

These series converge rapidly because α has such a small magnitude.

The results of some numerical computations are shown in Fig. 2. The solid curve corresponds to the

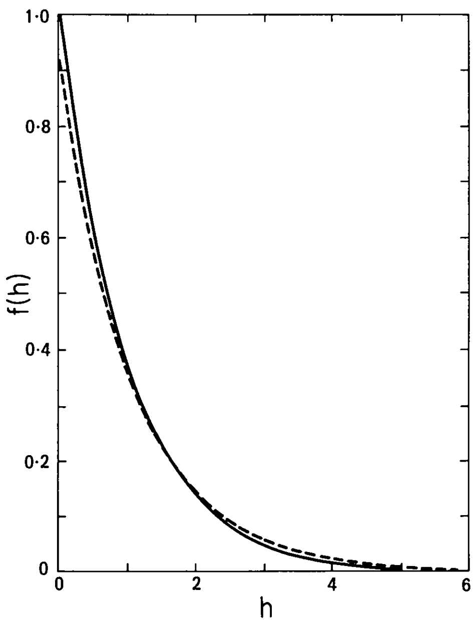

Fig. 2. Probability density function of the normalized scattered intensity, equation (4.19), for $\alpha = 0$ (solid curve), $\alpha = 0{\cdot}3$ (broken curve).

negative exponential PDF

$$f_h(h) = \exp(-h) \quad (4.20)$$

which would arise when $\alpha = 0$. Even for the unrealistically large value of $\alpha = 0{\cdot}3$ (dotted curve), the PDF is only a minor perturbation of the negative exponential.

This suggests that we seek a simpler version of $f_h(h)$ consistent with the fact that α is small. To this end we employ the usual power-series expansion of the modified Bessel functions. We can easily show that

$$f_h(h) \simeq (1+\alpha^2)^{-1}\exp(-h)(1+\alpha^2 h) \quad (4.21)$$

is an excellent approximation to (4.19) when α is very small. In fact, if $\alpha = 0{\cdot}2$, then (4.21) differs from (4.19) by less than 0·1%. This result is not surprising; after all, when β is small, then $\langle U \rangle \simeq 0$ in relation to $\langle U^2 \rangle$. The probability density function of the sum of the squares of two Gaussian distributed random variables, where first moments vanish and second moments are equal, is known to have a negative exponential probability density function.

References

EGELMAN, E. H. & DEROSIER, D. J. (1982). *Acta Cryst.* **A38**, 796-799.

EGELMAN, E. H., FRANCIS, N. & DEROSIER, D. J. (1982). *Nature (London)*, **298**, 131-135.

EGELMAN, E. H., FRANCIS, N. & DEROSIER, D. J. (1983). *J. Mol. Biol.*, **166**, 605-629.

THOMAS, J. B. (1981). *Applied Probability and Random Processes*, pp. 124, 139, 199. New York: Krieger.

WATSON, G. N. (1947). *Treatise on the Theory of Bessel Functions*, pp. 22, 79. Cambridge Univ. Press.

Acta Cryst. (1987). **A43**, 49-54

Diffraction Profile from Small Crystallites with Anisotropic Temperature Factor

BY K. MIURA AND T. INO

Department of Physics, Osaka City University, Osaka, Japan

AND N. MINAMI

Department of Physics, Kinki University, Osaka, Japan

(*Received* 20 *January* 1986; *accepted* 27 *May* 1986)

Abstract

The diffraction intensity from small crystallites with lattice vibrations is expressed by a sum over direct-lattice points as previously described, using atomic scattering factors modified by the anisotropic vibration tensor β specified by the lattice vibration, the thermal diffuse scattering not being taken into account. Since the temperature factor for the atomic pair of the αth and βth atoms is $\beta_\alpha + \beta_\beta$, the factor is proved to have the same rotation symmetry as the Laue symmetry corresponding to the atomic distance vector of the pair, $\mathbf{r}_{\alpha\beta} = \mathbf{r}_\alpha - \mathbf{r}_\beta$. Consequently the intensity profile for the crystallites with lattice

0108-7673/87/010049-06$01.50