PARTHASARATHY, S. & PARTHASARATHI, V. (1973). *Acta Cryst.* A**29**, 428–432.
PONNUSWAMY, M. N. (1979). PhD Thesis. Univ. of Madras.
PONNUSWAMY, M. N. & PARTHASARATHY, S. (1977). *Acta Cryst.* A**33**, 838–844.
RAMACHANDRAN, G. N. & PARTHASARATHY, S. (1965). *Science*, **150**, 212–214.
RAMACHANDRAN, G. N. & SRINIVASAN, R. (1970). *Fourier Methods in X-ray Crystallography*. New York: Wiley and Sons.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1981). A**37**, 162–169

Estimation of Experiment Time in Neutron Diffractometry of Large Structures. I. Anomalous Scattering

BY W. JAUCH AND H. DACHS

Hahn-Meitner-Institut für Kernforschung, Glienicker Strasse 100, *D*-1000 *Berlin* 39, *Federal Republic of Germany*

(*Received* 20 *March* 1980; *accepted* 18 *September* 1980)

Dedicated to Professor H. Jagodzinski on the occasion of his 65th birthday

Abstract

An *a priori* approach to the prediction of required neutron beam time for single-crystal analysis of biological structures is presented. Time economy is determined by several main features: (i) tolerable inaccuracy of the Fourier map, (ii) method of extracting phase information, (iii) data-collection technique. Phasing by anomalous scattering at two wavelengths is considered. An expression is derived for the error in scattering density arising from experimental intensity errors. Application of the theoretical probability distributions for the intensities leads to an equation for the expected total counting time. Conditions are established for which the time expenditure is a minimum. Tables which aid easy application of the results are given as well as a numerical example.

1. Introduction

The work of Schoenborn and his colleagues (Schoenborn, 1969; Norvell, Nunes & Schoenborn, 1975) has shown that neutron diffraction can be applied successfully to protein crystals. Protein neutron diffraction aims at the elucidation of structural features, particularly hydrogen atoms, which are not accessible to X-rays. The most serious problem of such work is the time and expense involved in data collection. In general, the practicability of a neutron study depends on the request for beam time.

Neutron diffraction offers the possibility of tackling the phase problem by means of anomalous scattering from nuclei such as ^{113}Cd, ^{149}Sm or ^{157}Gd. This method has been used to solve several small crystal structures (*e.g.* Koetzle & Hamilton, 1975; Sikka & Rajagopal, 1975). Results of an application to a protein structure have been reported by Schoenborn (1975).

The present paper is concerned with the prediction of experiment time when neutron anomalous scattering is used for phase determination. Its purpose is to provide a basis for experiment planning. The problem is approached by an analysis of the expected errors in the density map.

2. Tolerable density error

2.1. *Error model*

If series termination effects are not considered the true scattering density is defined by the truncated Fourier series

$$\rho(\mathbf{r}) = \bar{\rho} + V^{-1} \sum F_{\mathbf{H}} \cos(2\pi \mathbf{H}.\mathbf{r} - \varphi_{\mathbf{H}}), \qquad (1)$$

where V = unit-cell volume, $\bar{\rho} = F_0/V$ = average scattering density, $\varphi_{\mathbf{H}}$ = phase angle of structure factor $\mathbf{F}_{\mathbf{H}}$, $\mathbf{H}$ = reciprocal-lattice vector, and the summation is carried over a sphere in reciprocal space, up to a radius H_0.

The expected accuracy of the density can be predicted if a model for the errors in the Fourier components and a specific error criterion is assumed.

0567-7394/81/020162-08$01.00 © 1981 International Union of Crystallography

An integrated squared-error criterion is commonly accepted (its weakness consists in weighting image distortions just as heavily as lack of resolution). Blow & Crick (1959) introduced the concept of the 'best Fourier' with weighted coefficients ($w_{\mathbf{H}}\,\mathbf{F}_{\mathbf{H}}$) which has a lower noise level than a normal ($\mathbf{F}_{\mathbf{H}}$) synthesis. Their treatment, which is based upon phase errors only, may easily be generalized by the inclusion of errors in the structure amplitudes.

Consider the experimental density

$$\rho_e(\mathbf{r}) = \bar{\rho} + V^{-1} \sum w_{\mathbf{H}}(F_{\mathbf{H}} + \Delta F_{\mathbf{H}}) \times \cos(2\pi \mathbf{H}.\mathbf{r} - \varphi_{\mathbf{H}} + \Delta\varphi_{\mathbf{H}}), \quad (2)$$

$\Delta F_{\mathbf{H}}$ and $\Delta\varphi_{\mathbf{H}}$ being independent random errors with zero mean. If we define $\Delta\rho(\mathbf{r}) = \rho_e(\mathbf{r}) - \rho(\mathbf{r})$, the mean square error averaged over the unit cell is given by

$$\langle(\Delta\rho)^2\rangle = V^{-2} \sum \{F_{\mathbf{H}}^2 + w_{\mathbf{H}}^2[F_{\mathbf{H}}^2 + \sigma^2(F_{\mathbf{H}}^2)] - 2w_{\mathbf{H}} F_{\mathbf{H}}^2 \langle\cos\Delta\varphi_{\mathbf{H}}\rangle\}, \quad (3)$$

where $\sigma^2(F) \equiv \langle(\Delta F)^2\rangle$. With the weighting function

$$w_{\mathbf{H}} = \langle\cos\Delta\varphi_{\mathbf{H}}\rangle[1 + \sigma^2(F_{\mathbf{H}})/F_{\mathbf{H}}^2]^{-1}, \quad (4)$$

the following minimum value is obtained

$$\langle(\Delta\rho)^2\rangle = V^{-2} \sum F_{\mathbf{H}}^2(1 - w_{\mathbf{H}}\langle\cos\Delta\varphi_{\mathbf{H}}\rangle). \quad (5)$$

Owing to the phase errors and the weights, the mean value $\langle\rho_e(\mathbf{r})\rangle$ lies always below the true density level. Therefore, the mean square error (5) is larger than the variance $\sigma^2(\rho_e)$.

2.2. *Accuracy and correlation*

A convenient measure of the quality of density maps at any resolution is provided by

$$G^2 = \langle(\Delta\rho)^2\rangle / \langle(\rho - \bar{\rho})^2\rangle, \quad (6)$$

where the denominator is defined as

$$V^{-1} \int_V [\rho(\mathbf{r}) - \bar{\rho}]^2 \, \mathrm{d}\mathbf{r}.$$

Since

$$\langle(\Delta\rho)^2\rangle = \langle(\rho - \bar{\rho})^2\rangle + \langle(\rho_e - \bar{\rho})^2\rangle - 2\langle(\rho_e - \bar{\rho})(\rho - \bar{\rho})\rangle, \quad (7)$$

G^2 is related to the correlation coefficient defined by

$$C = \frac{\langle(\rho_e - \bar{\rho})(\rho - \bar{\rho})\rangle}{\{\langle(\rho - \bar{\rho})^2\rangle\langle(\rho_e - \bar{\rho})^2\rangle\}^{1/2}}. \quad (8)$$

For the Fourier series (1), (2) the following mean values are obtained

$$\langle(\rho - \bar{\rho})^2\rangle = V^{-2} \sum F_{\mathbf{H}}^2 \quad (9)$$

$$\langle(\rho_e - \bar{\rho})^2\rangle = V^{-2} \sum w_{\mathbf{H}}^2[F_{\mathbf{H}}^2 + \sigma^2(F_{\mathbf{H}})] \quad (10)$$

$$\langle(\rho - \bar{\rho})(\rho_e - \bar{\rho})\rangle = V^{-2} \sum w_{\mathbf{H}} F_{\mathbf{H}}^2 \langle\cos\Delta\varphi_{\mathbf{H}}\rangle, \quad (11)$$

where it is understood that the summations involve only the terms $\mathbf{H} \le \mathbf{H}_0$, $\mathbf{H} \neq (000)$. Hence

$$C = \frac{\sum w_{\mathbf{H}} F_{\mathbf{H}}^2 \langle\cos\Delta\varphi_{\mathbf{H}}\rangle}{\{\sum F_{\mathbf{H}}^2\}^{1/2}\{\sum w_{\mathbf{H}}^2[F_{\mathbf{H}}^2 + \sigma^2(F_{\mathbf{H}})]\}^{1/2}} \quad (12)$$

and

$$G^2 = 1 + k - 2k^{1/2} C \quad (13)$$

with

$$k = \sum w_{\mathbf{H}}^2[F_{\mathbf{H}}^2 + \sigma^2(F_{\mathbf{H}})] / \sum F_{\mathbf{H}}^2. \quad (14)$$

If the optimum weights (4) are applied, it follows from (13) and (14):

$$G^2 = 1 - C^2. \quad (15)$$

For an unweighted Fourier ($w_{\mathbf{H}} = 1$) which is affected by phase errors only one finds

$$G^2 = 2(1 - C), \quad (16)$$

whereas if it is subjected to amplitude errors only a nonlinear relationship occurs:

$$G^2 = C^{-2} - 1. \quad (17)$$

It is well known that the phase angles are of greater importance than the amplitudes (Ramachandran & Srinivasan, 1961). This fundamental property of the Fourier synthesis is reflected by the different behaviour of (16) and (17). From the two measures of agreement it is the correlation coefficient which is more apt for describing the recognizability of a structure.

2.3. *Low-resolution maps*

The interpretability of a protein density map at low resolution (6 Å $\ge d_{\min} \ge$ 4·5 Å) depends on structural features of the molecule and the crystal (helix content, aggregation of subunits, packing density) as well as on the accuracy of the structure factors.

The influence of errors was studied with a hypothetical two-dimensional chain molecule (548 atoms) and with myoglobin (Watson, 1968) both at 6 and 4·5 Å resolution. Calculated neutron structure factors defined the 'true' density. A random-number routine was then used to generate normally distributed phase errors which were independent of the amplitudes. For $C \lesssim 0{\cdot}85$ it was possible to trace the chain in accordance with the ideal maps.

2.4. *High-resolution maps*

An indication of the tolerable density error at high resolution ($d_{\min} \le$ 2 Å) is provided by the

following argument. In protein work the quality of the phases is usually assessed by the figure-of-merit m_H [= estimator of $\langle \cos \Delta\varphi_H \rangle$, Blow & Crick (1959)] and in general its mean value, taken over all reflections up to high resolution, is about 0·7. If we assume m_H to be the same for all reflections, (12) yields, for a weighted Fourier synthesis with only phase errors, $C = \langle \cos \Delta\varphi_H \rangle$. It is therefore suggested that $C \simeq 0{\cdot}7$ should be sufficient for the initial Fourier map. This estimate has been checked with an organic molecule where we could identify 13 from 18 H atoms ($d_{min} = 1{\cdot}7$ Å, $B = 10$ Å^2). In the corresponding error-free map 17 negative hydrogen peaks were clearly visible whereas the positive scatterers remained nonresolved.

3. Anomalous scattering

3.1. *Phase determination*

If the unit cell contains n identical nuclei of complex scattering length ($b = b^0 + b' + ib''$) such as ^{113}Cd, ^{149}Sm or ^{157}Gd (Fig. 1) and N normal scatterers the structure factor at wavelength λ_1 may be written as

$$\mathbf{F}_1(\mathbf{H}) = \mathbf{F}_N(\mathbf{H}) + \mathbf{F}_n^0(\mathbf{H}) + \mathbf{F}'_{n1}(\mathbf{H}) + i\mathbf{F}''_{n1}(\mathbf{H}), \quad (18)$$

where $\mathbf{F}_N$ and $\mathbf{F}_n^0$ denote the wavelength-independent parts and $\mathbf{F}'_{n1}$, $\mathbf{F}''_{n1}$ are the contributions arising from b'_1, b''_1.

From measurements of inverse reflections, $F_i^2(+)$ and $F_i^2(-)$, at two suitable wavelengths ($b'_1 \neq b'_2$), the phases can be determined uniquely if the resonant atom positions are known:

$$\sin \theta_1 = \frac{F_1^2(+) - F_1^2(-)}{4F'_1 F''_{n1}} \quad (19)$$

(Ramachandran & Raman, 1956);

$$\cos \theta_1 = \frac{F'_{n1}}{F'_1} + \frac{(F_1'^2 - F_{n1}'^2) - (F_2'^2 - F_{n2}'^2)}{2F'_1 |\mathbf{F}'_{n1} - \mathbf{F}'_{n2}|} \quad (20)$$

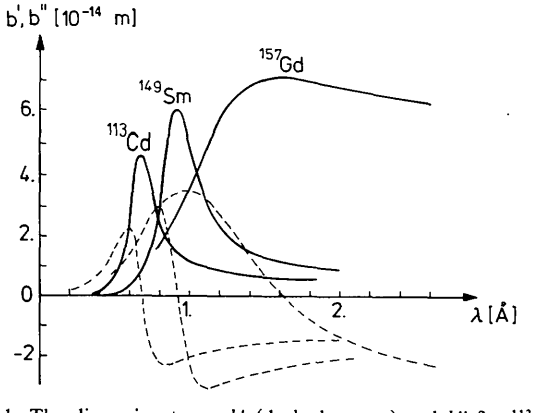

Fig. 1. The dispersion terms b' (dashed curves) and b'' for ^{113}Cd, ^{149}Sm and ^{157}Gd (after *Neutron Cross Sections*, 1973).

(Singh & Ramaseshan, 1968); where $\theta_i = \varphi_i - \varphi_n$,

$$F_i'^2 = \frac{F_i^2(+) + F_i^2(-)}{2} - F_{ni}''^2 \quad (21)$$

and the subscripts $i = 1, 2$ denote the two wavelengths; φ_i, φ_n are the phase angles of $\mathbf{F}'_i$ and $\mathbf{F}_n^0$. [Equation (20) differs from equation (23) of Singh & Ramaseshan (1968), but can be easily derived from it.]

3.2. *Errors in the structure factors*

Denote by $\mathbf{C}_x$ the covariance matrix of the vector of Bijvoet pairs at the two wavelengths, $\mathbf{x} = [F_1^2(+), \ldots, F_2^2(-)]$. We assume $\mathbf{C}_x$ to be diagonal, and since $F_i^2(+)$ and $F_i^2(-)$ are of nearly the same magnitude, $\sigma^2[F_i^2(+)] = \sigma^2[F_i^2(-)] \simeq 4F_N^2\,\sigma^2(F_i)$.

Let $\mathbf{y}_i$ be the column vector (F'_i, sin θ_i, cos θ_i) and $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$. The covariance matrix of $\mathbf{y}$ may then be approximated according to the propagation of error formula:

$$\mathbf{C}_y = \mathbf{T}\mathbf{C}_x\mathbf{T}',$$

where the prime denotes transposition and $T_{kl} = \partial y_k / \partial x_l$ is evaluated from (19), (20), (21). Since for large structures in most cases $|\mathbf{F}'_{n1} - \mathbf{F}'_{n2}|$, $F''_{ni} \ll F'_i$ certain elements of $\mathbf{T}$ can be simplified.

The constituents of the structure factors $\mathbf{F}'_i$ can be collected in a vector $\mathbf{z} = (F'_1, \sin \varphi_1, \ldots, \cos \varphi_2)$, which is related to $\mathbf{y}$ by a linear transformation $\mathbf{V}$. The covariance matrix of $\mathbf{z}$ is given by

$$\mathbf{C}_z = \mathbf{V}\mathbf{C}_y\mathbf{V}'. \quad (22)$$

3.3. *Errors in scattering density*

The scattering density based on the structure factors $\mathbf{F}'_1(H)$ may be written as

$$\rho_1(\mathbf{r}) = \bar{\rho} + V^{-1} \sum F'_1(\cos \varphi_1 \cos 2\pi\mathbf{H}.\mathbf{r} + \sin \varphi_1 \sin 2\pi\mathbf{H}.\mathbf{r}). \quad (23)$$

The experimental estimates of the structure amplitudes and the trigonometric functions differ from their true values by errors $\Delta F'_1$, $\Delta \cos \varphi_1$ and $\Delta \sin \varphi_1$. If a weighting function w is attached to the experimental Fourier coefficients the mean square error, averaged over the unit cell, is given by

$$\langle (\Delta\rho_1)^2 \rangle = V^{-1} \left\langle \int_V [\rho_1^{exp}(\mathbf{r}) - \rho_1(\mathbf{r})]^2 \, d\mathbf{r} \right\rangle$$

$$= V^{-2} \sum \{F_1'^2(1-w)^2 + w^2 F_1'^2[\sigma^2(\cos \varphi_1) + \sigma^2(\sin \varphi_1)] + w^2 \sigma^2(F'_1)\}. \quad (24)$$

With

$$w = [1 + \sigma^2(\cos\varphi_1) + \sigma^2(\sin\varphi_1) + \sigma^2(F_1')/F_1'^2]^{-1}, \quad (25)$$

the minimum value

$$\langle(\Delta\rho_1)^2\rangle = V^{-2}\sum F_1'^2(1-w) \quad (26)$$

is achieved.

Though useful with experimental data, (26) is not suitable for an *a priori* evaluation. We will therefore assume in the following that no weights are applied. With $w = 1, F_1' \simeq F_N$ and substitution of (22) into (24):

$$\langle(\Delta\rho_1)^2\rangle = \frac{1}{2V^2}\sum (F_N^2\,\sigma^2(F_1) \times \{1 + b_{11}[1 + \sigma^2(F_2)/\sigma^2(F_1)]\}/F_{n1}''^2), \quad (27)$$

where

$$b_{li} = (b_i'')^2/(b_1' - b_2')^2, \quad i = 1, 2.$$

The error of $\rho_2(\mathbf{r})$ follows from (27) by interchanging the indices.

Since only the densities of the anomalous scatterers are wavelength dependent the redundancy contained in ρ_1 and ρ_2 can be used to calculate an averaged density $\rho_m(\mathbf{r}) = [\rho_1(\mathbf{r}) + \rho_2(\mathbf{r})]/2$ of increased accuracy:

$$\langle(\Delta\rho_m)^2\rangle = [\langle(\Delta\rho_1)^2\rangle + \langle(\Delta\rho_2)^2\rangle + 2\,\mathrm{cov}\,(\rho_1,\rho_2)]/4. \quad (28)$$

For small anomalous effects

$$\mathrm{cov}\,(\rho_1,\rho_2) = V^{-2}\sum F_N^2\{\mathrm{cov}\,(\cos\varphi_1, \cos\varphi_2) + \mathrm{cov}\,(\sin\varphi_1, \sin\varphi_2)\};$$

hence with (22) one obtains

$$\langle(\Delta\rho_m)^2\rangle = \frac{1}{8V^2}\sum_{i=1}^{2}\left\{\sum F_N^2\,\sigma^2(F_i)\,(1 + 4b_{li})/F_{ni}''^2\right\}. \quad (29)$$

4. Experimental factors

4.1. *Luminosity and resolution*

A conventional two-crystal diffractometer is considered. If the collimations and reflectivities are described by Gaussians of equal horizontal and vertical angular dispersion α^2 for the in-pile and the monochromated beam collimator and of misorientation dispersion η_M^2 in both directions for a monochromator with peak reflectivity p_M, the following expression is obtained for the neutron flux at the sample:

$$\Phi = (\pi/2)^{1/2}\,p_M\,\eta_M\,\alpha^4(2\alpha^2 + 4\eta_M^2\sin^2\theta_M)^{-1/2} \times (2\alpha^2 + 4\eta_M^2)^{-1/2}\cot\theta_M\,\lambda\,\mathrm{d}\Phi/\mathrm{d}\lambda \quad (30)$$

with θ_M = Bragg angle of the monochromator. $\mathrm{d}\Phi/\mathrm{d}\lambda$ is the spectral flux density [$\mathrm{d}\Phi/\mathrm{d}\lambda = 2\Phi_0\,\lambda_T^4\,\lambda^{-5}\exp(-\lambda_T^2/\lambda^2)$ where Φ_0 = total thermal flux and $\lambda_T = h/(2mkT)^{1/2}$ for moderator temperature T].

The intensity distribution in the crystal rocking curve, $I(\Delta\omega)$, and the Bragg-scattered beam, $I(\Delta 2\theta)$, may be described by Gaussians of variance $\sigma^2(\Delta\omega)$ and $\sigma^2(\Delta 2\theta)$ respectively, which depend on the instrument parameters, the mosaic spread of the sample (η_s), the diffraction angle and the scanning mode (*e.g.* Dachs, 1961; Sequeira, 1974). $\sigma(\Delta\omega)$ is generally of minimum width when the reflecting planes of the specimen and the monochromator are parallel, where for large sample mosaicity

$$\sigma(\Delta\omega) \simeq \eta_s \quad (31)$$

and for the angular resolution in a stationary detector (ω scan) accepting all diffracted neutrons

$$\sigma(\Delta 2\theta) \simeq \alpha. \quad (32)$$

For protein crystals (high mosaic spread, small Bragg angles) the ω scan is preferable to the $\theta/2\theta$ scan. The focusing region ($\theta \simeq \theta_M$) will comprise the bulk of reflections if θ_M is chosen to be equal to the maximum Bragg angle under investigation, θ_0, which implies

$$\cot\theta_M \simeq 2d_{\min}/\lambda. \quad (33)$$

We postulate that the integrated intensity does not extend beyond $\pm 2{\cdot}5\sigma$ of the Gaussian profile. Then for the reflections to be separated at least the following conditions must be satisfied

$$\sigma(\Delta\omega) \le \lambda V^{-1/3}/5, \quad \sigma(\Delta 2\theta) \le \lambda V^{-1/3}/5, \quad (34)$$

where the real unit cell has been replaced by the equivalent cubic cell with side $V^{1/3}$. The upper limits for the full width at half height of collimation and sample mosaicity are then restricted to

$$\alpha' = \eta_s' \simeq 0{\cdot}5\,\lambda V^{-1/3} \quad (35)$$

[$\alpha' = 2(2\ln 2)^{1/2}\alpha$]. Substitution of (33) and (35) into (30) gives an estimate of the attainable incident neutron flux.

4.2. *Background and absorption*

Incoherent scattering from hydrogen is the principal source of noise in biological structure work. Hence the expected background counting rate to be recorded may be written as

$$R_B = \Phi A\,\frac{V_s}{V}\,\frac{\sum\sigma_{\mathrm{inc}}^{\mathrm{H}}}{4\pi}\,\varepsilon\Delta\Omega, \quad (36)$$

where A = transmission factor, V_s = volume of the sample, $\sum\sigma_{\mathrm{inc}}^{\mathrm{H}}$ = incoherent scattering cross section of hydrogen per unit cell, ε = detector efficiency, $\Delta\Omega$ = solid angle of the counter. $\sigma_{\mathrm{inc}}^{\mathrm{H}}$ varies with wavelength and binding potential between 20 and 80 $\times$ 10^{-28} m^2.

The value $\sigma_{inc}^{H} \simeq 50 \times 10^{-28}$ m^2 ($\lambda = 1{\cdot}5$ Å) was reported for myoglobin (Nunes & Norvell, 1975).

An effective transmission factor

$$A = \exp(-\mu V_s^{1/3}) \tag{37}$$

is used in the following. The dominating processes which contribute to the attenuation coefficient μ are hydrogen incoherent scattering and true absorption of the resonant scattering nuclei ($\sigma_{abs} = 2\lambda b''$); hence

$$\mu \simeq V^{-1} \sum (\sigma_{inc}^{H} + \sigma_{abs}), \tag{38}$$

where the summation is over the unit cell.

4.3. *Errors in measurement*

If N_P and N_B are the gross peak and background counts, recorded during the times t_P and t_B, the net integrated counts are given by

$$I = N_P - \frac{t_P}{t_B} N_B \tag{39}$$

and the variance is estimated as

$$\sigma^2(I) = N_P + (t_P/t_B)^2 N_B + P^2 I^2, \tag{40}$$

where the last term is due to fluctuations aside from counting statistics.

With

$$I = cLF^2/\dot{\omega}, \quad c = \Phi\lambda^3 V_s A\varepsilon/V^2 \tag{41}$$

(L = Lorentz factor, $\dot{\omega} = \Delta\omega/t_P$ = angular scanning speed) and

$$\sigma^2(F) = \sigma^2(F^2)/4F^2, \tag{42}$$

the variance in F is obtained as

$$\sigma^2(F) = \frac{\Delta\omega}{4cLt_P}\left[1 + \frac{\Delta\omega R_B}{cLF^2}\left(1 + \frac{t_P}{t_B}\right)\right] + P^2\frac{F^2}{4}. \tag{43}$$

From (36), (41) and the conditions of peak separation (34), it follows that

$$\Delta\omega R_B/c = \Delta\omega\Delta\Omega V \lambda^{-3} \sum \sigma_{inc}^{H}/4\pi \leq \sum \sigma_{inc}^{H}/4\pi. \tag{44}$$

Further specification of $\sigma(F)$ will depend on the mode of registration. Sequential recording with a single counter is susceptible to various optimizations (*e.g.* adjustment of precision for individual reflections according to their importance for a particular problem), whereas parallel recording with a multidetector implies the same counting time for every reflection.

In the following we will assume that a linear multidetector, covering the total range of scattering angles ($2\theta_0$), is used and that data are collected in normal beam geometry with crystal rotation. Integration of the Bragg peaks is performed over the rotation angle ($\Delta\omega$) and the respective detector channels ($\Delta 2\theta$), *i.e.* over n_P points of the (ω,2θ) matrix. The background is estimated from n_B points in the vicinity of the peak. One obtains from (43) and (44):

$$\sigma^2(F) = \frac{\pi}{2cL\tau}\left(1 + \frac{\sigma}{LF^2}\right) + P^2\frac{F^2}{4} \tag{45}$$

$$\sigma = (1 + n_P/n_B) \sum \sigma_{inc}^{H}/4\pi.$$

τ denotes the time for a 2π rotation of the sample. The Lorentz factor for the equatorial plane ($L = 1/\sin 2\theta$) will be used throughout the statistical calculations.

5. Formula for the counting time

Substitution of (45) into (29) yields

$$\langle(\Delta\rho_m)^2\rangle = V^{-2}\left[\sum_{i=1}^{2}\sum_{m} F_{ni}''^{-2}\{\gamma_i\tau_i^{-1}(L_i^{-1}F_N^2 + L_i^{-2}\sigma) + \delta_i F_N^4\} + \sum_{M-m} F_N^2\right], \tag{46}$$

where

$$\gamma_i = \pi c^{-1}(1 + 4b_{ii})/16, \quad \delta_i = P^2(1 + 4b_{ii})/32$$

and $M = 4\pi H_0^3 V/3$, m = number of reflections actually used in the Fourier synthesis. The last term in (46) takes into account that $M - m$ structure factors with resonance contributions smaller than a certain threshold value F_t'' are omitted from the Fourier.

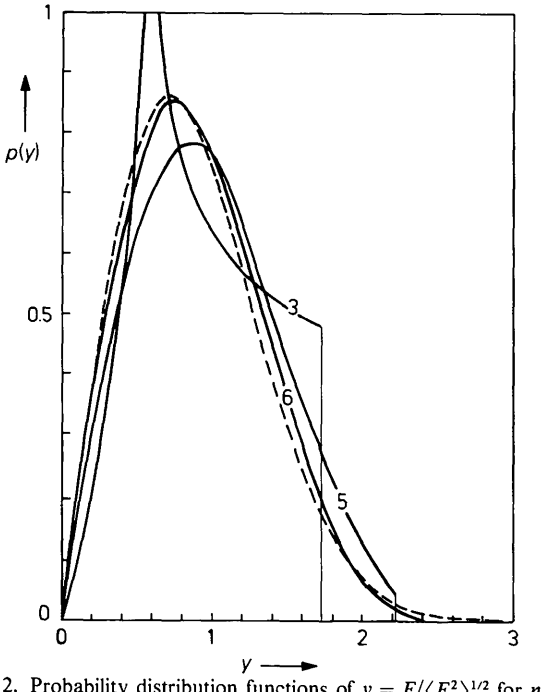

Fig. 2. Probability distribution functions of $y = F/\langle F^2\rangle^{1/2}$ for n = 3, 5 and 6 non-centrosymmetrically arranged equal atoms [calculated from Pearson's (1906) tables] and the acentric Wilson distribution (broken line).

Table 1. *Probability density functions and related integrals for $n = 1, 2$ and the asymptotic distributions*

n	$p(y)$	$N(y_t)$	$\int_{y_t}^{\sqrt{n}} y^{-2} p(y)\,dy$
1	$\delta(y-1)$	0	1
2	$2\pi^{-1}(2-y^2)^{-1/2}$	$2\pi^{-1}\arcsin(y_t/\sqrt{2})$	$(2-y_t^2)^{1/2}/(\pi y_t)$
C	$(2/\pi)^{1/2}\exp(-y^2/2)$	$\operatorname{erf}(y_t/\sqrt{2})$	$(2/\pi)^{1/2} y_t^{-1}\exp(-y_t^2/2) - \operatorname{erfc}(y_t/\sqrt{2})$
AC	$2y\exp(-y^2)$	$1-\exp(-y_t^2)$	$E_1(y_t^2)$

$$\operatorname{erf}(x) = 2\pi^{-1/2}\int_0^x \exp(-t^2)\,dt, \quad \operatorname{erfc}(x) = 1 - \operatorname{erf}(x), \quad E_1(x) = \int_x^\infty t^{-1} e^{-t}\,dt.$$

When the same temperature factor $T[= \exp(-BH^2/4)]$ is assigned to all atoms, the mean value of (46) takes the form

$$\langle\langle(\Delta\rho_m)^2\rangle\rangle = V^{-2}\Bigg[m\sum_{i=1}^{2}\langle F_{ni}''^{-2}\rangle\{\gamma_i\tau_i^{-1}(\langle L_i^{-1}\rangle\Sigma + \langle(L_i T)^{-2}\rangle\sigma) + 2\delta_i\Sigma^2\langle T^2\rangle\} + (M-m)\langle T^2\rangle\Sigma\Bigg] \tag{47}$$

with

$$\Sigma = \sum_{j=1}^{N} b_j^2 = \langle F_N^2\rangle \tag{48}$$

(Wilson, 1942),

$$\langle F_{ni}''^{-2}\rangle = (nb_i''^2)^{-1}\int_{y_t}^{\sqrt{n}} y^{-2}p(y)\,dy/[1-N(y_t)], \tag{49}$$

$$m = M[1-N(y_t)]. \tag{50}$$

$p(y)$ is the probability distribution and $N(y)$ the cumulative distribution function of the normalized anomalous structure amplitudes: $y = F_n''/n^{1/2}b''$ $(0 \le y \le n^{1/2})$. They depend on the number and symmetry of the resonant scatterers in the crystal. We will distinguish the cases: $n = 1$, 2 and the centric (C) and acentric (AC) distribution (Wilson, 1949). The integrals occurring in (49) are collected in Table 1. Fig. 2 shows that for small n the asymptotic Wilson function is approached very rapidly. $\langle F_N^4\rangle = 2\Sigma^2$ (AC distribution) has been used in (47). The average values of L^{-1} *etc.* over reciprocal space are given in the Appendix.

Table 2. *Optimum thresholds and numerical values of function* (56) *for different distribution functions*

G^2	Y_t(2)	Y_t(C)	Y_t(AC)	Γ(2)	Γ(C)	Γ(AC)
0·6	0·70	0·53	0·62	0·48	0·28	0·39
0·5	0·57	0·41	0·52	0·33	0·16	0·27
0·4	0·45	0·30	0·42	0·20	0·092	0·18
0·3	0·34	0·22	0·34	0·11	0·047	0·11
0·2	0·22	0·14	0·25	0·050	0·018	0·062
0·1	0·11	0·07	0·16	0·013	0·0043	0·024

Division of (47) by $V^{-2}\sum_M F_N^2$ $(= V^{-2} M\langle T^2\rangle\Sigma)$ leads to a relation between the accuracy factor G^2 (6) and the counting times τ_i:

$$G^2 = N(y_t) + \int_{y_t}^{\sqrt{n}} y^{-2}\,p(y)\,dy\left(\Gamma_P + \sum_{i=1}^{2}\tau_i^{-1}\Gamma_i\right) \tag{51}$$

with

$$\Gamma_i = \gamma_i(\langle L_i^{-1}\rangle + \sigma\Sigma^{-1}\langle(L_i T)^{-2}\rangle)/nb_i''^2\langle T^2\rangle, \tag{52}$$

$$\Gamma_P = 2\Sigma n^{-1}\sum_{i=1}^{2}\delta_i b_i''^{-2}. \tag{53}$$

From a simple calculation one finds that with

$$\tau_1/\tau_2 = (\Gamma_1/\Gamma_2)^{1/2} \tag{54}$$

a prescribed accuracy G^2 is achieved within a minimum amount of total counting time $\tau = \tau_1 + \tau_2$.

Similarly, a best threshold value Y_t may be defined. If $y_t > Y_t$, the intensities have to be measured more precisely in order to compensate for the loss in accuracy due to omitted reflections. For $y_t < Y_t$, on the other hand, time is wasted in phasing reflections with poorly defined label vectors. From (54) and (51), it follows from $\partial\tau_1/\partial y_t = 0$ that Y_t has to satisfy the equation

$$N(Y_t) + Y_t^2\int_{Y_t}^{\sqrt{n}} y^{-2}\,p(y)\,dy = G^2. \tag{55}$$

Numerical solutions of (55) are listed in Table 2. With the auxiliary function

$$\Gamma(n, G^2) = \Big[G^2 - N(Y_t)\Big]\Big/\int_{Y_t}^{\sqrt{n}} y^{-2}\,p(y)\,dy \tag{56}$$

(see Table 2), the total counting time per layer line takes the simple form

$$\tau = \frac{(\Gamma_1^{1/2} + \Gamma_2^{1/2})^2}{\Gamma(n, G^2) - \Gamma_P}. \tag{57}$$

We propose that (57) be used in estimating the required beam time for the neutron experiment.

6. Discussion

6.1. *Validity of Wilson statistics*

Equation (48) is based on uncorrelated atomic scattering elements. This assumption is not fulfilled with low-resolution data and deviations from Wilson's formula are observed (*e.g.* Blundell & Johnson, 1976). Even if (48) is valid there remains some uncertainty concerning N, the number of atoms per primitive cell which participate in crystalline order.

The influence of uncompensated correlations due to atomic overlap, however, should be different for X-ray and neutron data since scattering lengths of both signs can occur for neutrons. We have simulated this effect with myoglobin (Watson, 1968) where the negative scattering length of hydrogen has been attached to some atoms randomly distributed over the molecule. The average intensity, taken over all reflections within a sphere of radius H_0, was calculated from the atomic coordinates:

$$\langle F_H^2 \rangle = \sum_j b_j^2 + 3 \sum_{j \neq j'} b_j b_{j'} (\sin x_{jj'} - x_{jj'} \cos x_{jj'})/x_{jj'}^3 = \Sigma + \Delta\Sigma, \quad (58)$$

where $x_{jj'} = 2\pi H_0 |\mathbf{r}_j - \mathbf{r}_{j'}|$ and $\Delta\Sigma$ denotes the contribution from the double sum. Results are shown in Table 3.

Generally, for the validity of the form of Wilson's distribution functions it is sufficient that $\langle \mathbf{F}_H \rangle = 0$. This condition will be fulfilled if the structure can be divided into segments whose positions are not correlated. Such a division should be possible even with a low-resolution protein structure (at least in the absence of pseudo-symmetry). This conjecture is supported by observations of Nixon & North (1976).

6.2. *Optimum wavelength combinations*

The optimum wavelength combination is defined as the one for which $\tau(\lambda_1,\lambda_2) =$ minimum. Apart from the resonant isotope type it will depend mainly on the neutron flux distribution $\Phi(\lambda)$, the transmission factor $A(\lambda)$, the wavelength dependence of the crystal reflectivity ($\sim\lambda^3$) and the required instrumental resolution ($\sim\lambda^{-2}$). Taking these various features into account yields the following wavelength variation for Γ_i:

$$\Gamma_i \sim (1 + 4b_{ii}) b_i''^{-2} \exp(2\lambda_i b_i'' x + \lambda_T^2/\lambda_i^2) \quad (59)$$

with $x = nV_s^{1/3}/V$.

From (57) and (59) the optimum wavelength pairs can be obtained for specified values of x. Results for the most important isotopes are displayed in Fig. 3. The graphs correspond to a 10% deviation of $\tau(\lambda_1,\lambda_2)$ from the minimum. For ^{113}Cd a hot source ($T = 2000$ K) has been assumed in place of a thermal one ($T = 300$ K). In the case of strong absorption the optimum region of ^{157}Gd is completely shifted to the low-wavelength side of the resonance.

Table 3. *Calculated relative deviations $\Delta\Sigma/\Sigma$ at different resolutions as a function of the amount of hydrogen scattering $\sigma_H^2 = \sum b_H^2/\Sigma$ (real protein: $\sigma_H^2 \simeq 0{\cdot}2$–$0{\cdot}3$)*

d_{min}	6 Å	4 Å	2 Å
σ_H^2			
0	2·96	1·17	0·046
0·15	0·38	0·39	−0·070
0·2	−0·15	−0·12	0·034
0·3	−0·18	−0·30	−0·034

6.3. *Symmetry of the anomalous scatterers*

The required counting time depends on the symmetry arrangement of the anomalous scatterers. From (57) it follows that for the two Wilson distributions

$$\tau(\mathrm{C})/\tau(\mathrm{AC}) \geq \Gamma(\mathrm{AC})/\Gamma(\mathrm{C})$$

for otherwise identical conditions. With $G^2 = 0{\cdot}4$, $\tau(\mathrm{C}) \geq 2\tau(\mathrm{AC})$. This considerable influence of symmetry refers to the very different probabilities for the occurrence of weak structure factors in the two cases.

6.4. *Unit-cell size*

From (57) one obtains that τ varies approximately with $V^{8/3} n^{-1} A(n/V)$ where $A(n/V)$ = transmission factor. For small proteins the number of reference scatterers will be independent of the unit-cell size. As the molecular weight increases, however, one has to consider the fact that large biomolecules are aggregates of similar or identical subunits. Therefore n will become proportional to V and hence $\tau \sim V^{5/3}$. A resolution of d_{min} in the Fourier synthesis requires that $V^{1/3} d_{min}^{-1}$ layer lines are to be recorded.

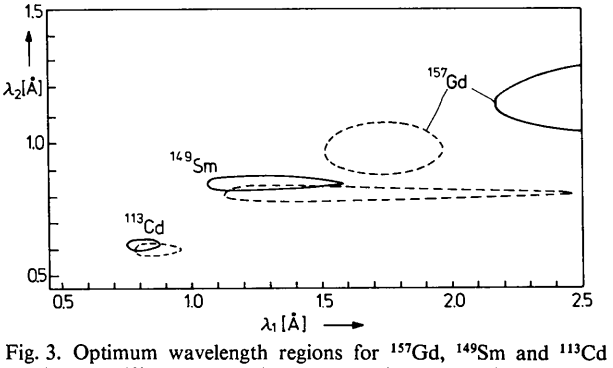

Fig. 3. Optimum wavelength regions for ^{157}Gd, ^{149}Sm and ^{113}Cd with two different absorption factors x (equation 59): $x = 0{\cdot}6 \times 10^{23}$ m^{-2} (full-line), $x = 3 \times 10^{23}$ m^{-2} (broken line).

7. Example

We take myoglobin as an example for the application of the above formulae. The values of the structural parameters are $V = 6{\cdot}64 \times 10^4$ Å^3, $\Sigma = 1800 \times 10^{-28}$ m^2, $\sum\sigma_{inc}^{H}/4\pi = 9400 \times 10^{-28}$ m^2, $B = 15$ Å^2 [for the atomic composition of the partially deuterated unit cell see Nunes & Norvell (1975)]. We suppose one anomalous scatterer to be bound per molecule ($n = 2$). Furthermore, let $V_s = 8$ mm^3, $P = 0{\cdot}02$, $n_B = 2n_P$. Table 4 summarizes the expected counting times for low- and high-resolution studies of various derivatives. Clearly, ^{113}Cd is by far the least suitable reference scatterer even if a hot source is available.

APPENDIX

For the calculation of average values the sums over reciprocal space can be replaced by integrals.

The average reciprocal Lorentz factor ($L^{-1} = \sin 2\theta$) depends on the maximum Bragg angle θ_0:

$$\langle L^{-1}\rangle = 3H_0^{-3}\int_0^{H_0} L^{-1}H^2\,dH$$

$$= 2(2 - 5\cos^3\theta_0 + 3\cos^5\theta_0)/5\sin^3\theta_0.$$

Introducing the function

$$f(x) = 2x^{-2}\left[\frac{\pi^{1/2}}{2x}\operatorname{erf}(x) - \exp(-x^2)\right],$$

Table 4. *Expected counting times for myoglobin derivatives*

	^{157}Gd		^{149}Sm		^{113}Cd	
λ (Å)	1·2	2·2	0·85	1·2	0·6	0·8
μ (mm^{-1})	0·66	1·11	0·49	0·41	0·34	0·37
Φ (10^{11} neutrons s^{-1} m^{-2})*	7	6·2	1·9	7	2·1	1·9
ε†	0·5	0·75	0·4	0·5	0·3	0·4
$\tau_i(h)$‡	46	40	146	69	444	355
$T(d)$‡	115¶		197		733	
$\tau_i(h)$§	6·1	5·4	20·6	9·4	76·6	58
$T(d)$§	4		10		45	

* $\Phi_0 = 10^{10}$ neutrons s^{-1} m^{-2} (HFR Grenoble), $\eta'_M = 25'$, $p_M = 0{\cdot}5$; a hot source is assumed for the ^{113}Cd derivative.

† $\varepsilon(\lambda)$ for a high-pressure ^{3}He detector, taken from Alberi (1975).

‡ $d_{min} = 1{\cdot}8$ Å, $G^2 = 0{\cdot}5$; $T = V^{1/3}\,d_{min}^{-1}\,(\tau_1 + \tau_2)$ = total experiment time.

§ $d_{min} = 5$ Å, $G^2 = 0{\cdot}2$.

¶ For $\lambda = 2{\cdot}2$ Å it is assumed that crystal rotation is around two different axes since $p = 12\%$ of the reflections lie in the blind region:

$$p = 1 + \frac{3\cos\theta_0}{8\sin^2\theta_0} - \tfrac{3}{4}\cos\theta_0 - \frac{3\theta_0}{8\sin^3\theta_0}.$$

[$\operatorname{erf}(x) = 2\pi^{-1/2}\int_0^x \exp(-t^2)\,dt$], the mean temperature factors may be written as

$$\langle T\rangle = 3H_0^{-3}\int_0^{H_0} TH^2\,dH = f[H_0(B/4)^{1/2}]$$

$$\langle T^2\rangle = f[H_0(B/2)^{1/2}]$$

$\langle (LT)^{-2}\rangle$ has to be evaluated by numerical integration:

$$\langle (LT)^{-2}\rangle = 3a^{-3/2}\,b(1 + \tfrac{5}{4}b)\int_0^a z^{3/2}\exp(z)\,dz$$

$$-\tfrac{3}{2}ab^2\exp(a)$$

with $a = BH_0^2/2$, $b = \lambda^2/B$.

References

ALBERI, J. L. (1975). *Brookhaven Symp. Biol.* **27**, VIII-26.

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*, p. 334. New York: Academic Press.

DACHS, H. (1961). *Z. Kristallogr.* **115**, 80–92.

KOETZLE, T. F. & HAMILTON, W. C. (1975). *Anomalous Scattering*, edited by S. RAMASESHAN & S. C. ABRAHAMS, pp. 489–502. Copenhagen: Munksgaard.

Neutron Cross Sections (1973). Report BNL 325. Brookhaven National Laboratory.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.

NORVELL, J. C., NUNES, A. C. & SCHOENBORN, B. P. (1975). *Science*, **190**, 568–570.

NUNES, A. C. & NORVELL, J. C. (1975). *Brookhaven Symp. Biol.* **27**, VII-57-65.

PEARSON, K. (1906). *Drapers' Co. Res. Mem. Biom. Ser.* No. 3.

RAMACHANDRAN, G. N. & RAMAN, S. (1956). *Curr. Sci.* **25**, 348–351.

RAMACHANDRAN, G. N. & SRINIVASAN, R. (1961). *Nature (London)*, **190**, 159–161.

SCHOENBORN, B. P. (1969). *Nature (London)*, **224**, 143–146.

SCHOENBORN, B. P. (1975). *Anomalous Scattering*, edited by S. RAMASESHAN & S. C. ABRAHAMS, pp. 407–416. Copenhagen: Munksgaard.

SEQUEIRA, A. (1974). *Acta Cryst.* A**30**, 839–843.

SIKKA, S. K. & RAJAGOPAL, H. (1975). *Anomalous Scattering*, edited by S. RAMASESHAN & S. C. ABRAHAMS, pp. 503–514. Copenhagen: Munksgaard.

SINGH, A. K. & RAMASESHAN, S. (1968). *Acta Cryst.* B**24**, 35–39.

WATSON, H. C. (1968). *Prog. Stereochem.* **4**, 299–333.

WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151–152.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.