Acta Cryst. (1965). **18**, 297

On the treatment of accidentally absent and composite reflexions in least-squares refinement of crystalline fibre structures.

By STRUTHER ARNOTT, *Medical Research Council Biophysics Research Unit, Department of Biophysics, King's College*, 26–29, *Drury Lane, London, W. C.* 2, *England*

(*Received* 8 *June* 1964)

This investigation has been prompted by the application of least-squares methods to diffraction data from crystalline fibres (*e.g.* Arnott & Coulter, 1963). For single-crystal data, the inclusion in least-squares analyses of contributions from reflexions with below-threshold intensities has been considered previously (Hamilton, 1955). Cruickshank (1961) has pointed out how markedly the accuracy of crystal-structure determinations is improved by such inclusions in the case where the number of accidental absences is relatively large. This is very often the case with data from crystalline fibres which show many accidental absences and where the number of reflexions recorded is a small fraction of the total possible. This is partly a consequence of the imperfect parallelism of the crystallites within the fibre which results in the diffracted intensity being dissipated over arcs, whose length increases with distance from the centre of the pattern, thereby giving higher threshold values for the observations.

The rotation-diagram characteristics of fibre diffraction patterns provide a further complication for reflexions, whatever their intensity, which have the same ξ-value and therefore are not resolvable. This problem of overlap has its most extreme form in powder diagrams for reflexions with similar ϱ-values.

Let us define F_i to be the structure amplitude and I_i the corrected intensity for a reflexion with structure factor $\mathbf{F}_i$, and the probability that F_i lies between F_i and F_i+dF_i to be $P(F_i)dF_i$. In a least-squares analysis which minimizes

$$R'=\Sigma w_i(F_i^{\text{obs}}-F_i^{\text{calc}})^2\,,$$

F_i should be assigned its expected or mean value $\mu_1(F_i)$, under the conditions provided by the observations, and a weight $w_i=1/\mu_2$ where μ_2 is the variance. When a reflexion has an intensity below the observable threshold, I_t, then

$$0\le F_i<F_t=\sqrt{I_t}\;.$$

It follows that

$$\mu_1(F_i)=\int_0^{F_t}F_iP(F_i)dF_i\Big/\int_0^{F_t}P(F_i)dF_i \qquad (1)$$

and

$$\mu_2(F_i)=\left\{\int_0^{F_t}F_i^2P(F_i)dF_i\Big/\int_0^{F_t}P(F_i)dF_i\right\}-\mu_1^2(F_i)\;. \qquad (2)$$

In the case where the intensity is below threshold, but composite, with N components then $0\le\sum_1^N F_i^2<I_t$ and

$$\mu_1(F_r)=\iint\ldots\int F_r\prod_1^N P_i(F_i)dF_i\Big/\iint\ldots\int\prod_1^N P_i(F_i)dF_i \qquad (3)$$

and

$$\mu_2(F_r)=\left\{\iint\ldots\int F_r^2\prod_1^N P_i(F_i)dF_i\Big/\iint\ldots\int\prod_1^N P_i(F_i)dF_i\right\}-\mu_1^2(F_r)\;. \qquad (4)$$

The limits of the N successive definite integrals are respectively 0 and F_t, 0 and $\sqrt{(I_t-F_1^2)}$, 0 and

$$\sqrt{(I_t-F_1^2-F_2^2)},\;\ldots,\;0\quad\text{and}\quad\sqrt{(I_t-\prod_1^{N=1}F_i^2)}\;,$$

expressing the fact that once some F's are chosen the values the remainder may take are constrained.

If $\langle F\rangle$ is the local average of all the intensities, we may assume that if $G_t=F_t/\langle F\rangle$ is small, which will usually be the case for below threshold intensities, the F's will be randomly distributed in the range 0 to F_t. The probability density $P(F)$ is then a constant for $\mathbf{F}$ real, and proportional to F, a radius in the complex plane, for $\mathbf{F}$ complex. Table 1 shows $\mu_1(F)$ and $\mu_2(F)$ calculated on this basis for the nine cases with $N\le 3$.

The case where a number of reflexions overlap and give an above threshold, and therefore measurable, intensity $I_s=F_s^2$ may be treated similarly. Here N components satisfy the condition

$$\sum_1^N F_i^2=I_s=F_s^2\;,$$

which is a more restricted case of the previous situation implying

$$0\le\Sigma' F_i^2<I_s$$

when any term F_r^2 is omitted from the summation. Therefore

Table 1. *Expected value* (μ_1) *and variance* (μ_2) *of below-threshold structure amplitudes*

No. of overlapping reflexions	Structure factor type		Real		Complex	
	Real	Complex	$\mu_1(F)/F_t$	$\mu_2(F)/F_t{}^2$	$\mu_1(F)/F_t$	$\mu_2(F)/F_t{}^2$
1	1	0	0·500	0·083	X	X
	0	1	X	X	0·667	0·056
2	2	0	0·424	0·070	X	X
	1	1	0·375	0·059	0·589	0·053
	0	2	X	X	0·533	0·049
3	3	0	0·375	0·059	X	X
	2	1	0·340	0·051	0·533	0·049
	1	2	0·313	0·045	0·491	0·045
	0	3	X	X	0·457	0·041

Table 2. *Means and variances for two overlapping, above-threshold, intensities*

	2 'real' reflexions				2 'complex' reflexions			
G_s	$\mu_1(G)$		$10^2 . \mu_2(G)$		$\mu_1(G)$		$10^2 . \mu_2(G)$	
0·2	0·16	(0·16)	0·20	(0·20)	0·13	(0·13)	0·19	(0·22)
0·4	0·32	(0·31)	0·76	(0·80)	0·27	(0·27)	0·89	(0·89)
0·6	0·48	(0·47)	1·65	(1·79)	0·42	(0·40)	1·70	(2·00)
0·8	0·65	(0·63)	2·14	(3·19)	0·57	(0·53)	3·10	(3·56)
1·0	0·82	(0·79)	3·95	(4·98)	0·73	(0·67)	5·00	(5·56)

$$\mu_1(F_r) = \iint \ldots \int F_r \Pi' P_i(F_i)\, dF_i \Big/ \iint \ldots \int \Pi' P_i(F_i)\, dF_i \quad etc., \qquad (5)$$

the prime indicating that no term $P_r(F_r)dF_r$ occurs in the product. The number of integrations will be $(N-1)$, betokening one degree of freedom less than in the below-threshold case. The limits of the pth definite integral being 0 and

$$\sqrt{(I_s - \sum_1^{(p-1)} F_i)} \, .$$

As is usual in X-ray intensity statistics it is convenient to consider $G = F/\langle F \rangle$ rather than F as the variable. Each $P_i(G_i)$ must be chosen accurately to represent the amplitude distribution and also to be mathematically tractable. When Wilson statistics apply, the probability densities for centric and acentric structure amplitude distributions are respectively proportional to $\exp(-\frac{1}{2}G^2)$ and $G \exp(-G^2)$. In the most common case when two reflexions overlap to give a total intensity $I_s = G_s^2 . \langle I \rangle$, it can be shown that if both reflexions are of the centric class

$$\mu_1(G_{\bar{1}}) = \sqrt{(2\pi)} G_s^2 \exp(-G_s^2/4) \{I_0(G_s^2/4) + I_1(G_s^2/4)\}/4 \operatorname{erf}(G_s/\sqrt{2}) \qquad (6)$$

and

$$\mu_2(G_{\bar{1}}) = (G_s^2 - 1) + \{2G_s \exp(-G_s^2/2)/\sqrt{(2\pi)} \operatorname{erf}(G_s/\sqrt{2})\} - \mu_1^2(G_{\bar{1}}) \qquad (7)$$

(I_n is the nth order modified Bessel function of the first kind); when both are from the acentric class

$$\mu_1(G_1) = G_s - \{\exp(-G_s^2) \int_0^{G_s} \exp(p^2)\, dp / (1 - \exp(-G_s^2))\} \qquad (8)$$

and

$$\mu_2(G_1) = \{G_s^2(1 - \exp(-G_s^2))\} - 1 - \mu_1^2(G_1) \, . \qquad (9)$$

In the third possible case where one belongs to each class, $\mu_1(G_{\bar{1}}) = \mu_1(G_1)$ above, and $\mu_1(G_1) = \mu_1(G_{\bar{1}})$ above, and similarly for the μ_2's.

Table 2 shows the values for a number of G_s. They are to be compared with the bracketed values obtained using the approximate probability densities, the use of which was suggested for the accidentally absent spectra. They can be seen to be reasonable approximations even when $G_s \sim 1$.

The inverses of the variances provide absolute weights for the observations. Generally some analytical function is used to provide relative weights for the single reflexion data and therefore a correlation for the two schemes has to be established. The treatment assumes that measurements of all I_t and I_s are subject to the same errors, it may sometimes be desirable also to subject these observations to a further weighting factor to take account of the variation of errors in I_t and I_s.

The author is obliged to Prof. Sir John Randall, F.R.S. for the provision of facilities and to Prof. M.H.F. Wilkins, F.R.S. for encouragement.

References

ARNOTT, S. & COULTER, C. L. (1963). *Acta Cryst.* **16**, A175.

CRUICKSHANK, D. W. J. & PILLING, D. E. (1961). *Computing methods and the phase problem in X-ray crystal analysis*, p. 46. Oxford: Pergamon Press.

HAMILTON, W. C. (1955). *Acta Cryst.* **7**, 185.

Acta Cryst. (1965). **18**, 298

A scattering curve for the ammonium ion.

By M. W. WEBB, *University of Bristol, England*

(*Received* 18 *June* 1964)

During two-dimensional least-squares calculations on nickel ammonium sulphate hexahydrate (Grimes, Kay & Webb, 1963) it was noticed that the recommended atomic shifts for the ammonium ion were unexpectedly much larger and more random than those for the oxygen atoms and there was no appreciable change in the R index from 0·13 in successive cycles. The ammonium ions and oxygen atoms were equally well defined on the electron density projection, so it was suspected that the strange behaviour of the ammonium ion was due to the use of the scattering curve of nitrogen as an approximation to that of the ammonium ion. It was therefore decided to calculate a scattering curve for the ammonium ion from the electron-density distribution given by Banyard & March (1961) for an assumed spherically symmetrical ammonium ion.