Acta Cryst. (1977). A33, 75–85

An Improved Method for Combining Isomorphous Replacement and Anomalous Scattering Diffraction Data for Macromolecular Crystals*

BY J. RALPH EINSTEIN

Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, *USA*

(*Received* 13 *November* 1975; *accepted* 1 *July* 1976)

The coefficients for the 'best' Fourier synthesis for electron density have been shown by Blow and Crick to be the centroids of probability distributions (over the complex plane) for the structure factors. In defining probabilities for the case in which Bijvoet differences are ignored or averaged, Blow and Crick utilized simplifying assumptions not in accord with probability theory. Raiz and Andreeva defined probabilities for the same case in a theoretically correct manner. The present treatment provides somewhat better approximations for that case, and an extension to the case in which Bijvoet differences are utilized. The variances arising from experimental errors are maintained separately from those due to uncertainties in the heavy-atom parameters. The treatment is restricted here to derivatives in any one of which the scattering factors for heavy-atom anomalous scatterers have the same ratio of real and imaginary components.

1. Glossary

$\mathbf{F}$	Structure factor
F,φ	Amplitude, phase of $\mathbf{F}$
$\mathbf{F}_C$	Centroid
F_o	'Parent', observed structure amplitude
δ_0^2	Variance of F_o
ϱ	$F-F_o$
N	Number of 'normal' derivatives
N'	Number of 'anomalous' derivatives
N_t	$N+N'$

For any derivative

δ_j^2	Variance of observed structure amplitude
$\mathbf{f}_j$	Structure-factor contribution from heavy-atom substitutions (see § 3)
f_j,α_j	Amplitude, phase of $\mathbf{f}_j$
$\mathbf{z}_j$	Hypothetical error of $\mathbf{f}_j$
ε_j^2	m.s. (mean square) amplitude of errors (for a given Bragg-angle range) of $\mathbf{f}_j$
x_j	Lack-of-closure error, $F_{jo}-\lvert\mathbf{F}+\mathbf{f}_j\rvert$
x_{jo}	Lack-of-closure error for $F=F_o$

For a 'normal' derivative

$\mathbf{F}_n$	Structure factor
F_{no}	Observed amplitude of $\mathbf{F}_n$
F_{nh}	Hypothetical value of amplitude of $\mathbf{F}_n$ (see § 3)
σ_n^2	$\delta_n^2+\varepsilon_n^2$

For an 'anomalous' derivative

$\mathbf{F}_m^+,\mathbf{F}_m^-$	Structure factors for $\mathbf{H}$, $-\mathbf{H}$
F_{mo}^+,F_{mo}^-	Observed amplitudes of $\mathbf{F}_m^+,\mathbf{F}_m^-$
F_{mh}^+,F_{mh}^-	Hypothetical values of amplitudes of $\mathbf{F}_m^+,\mathbf{F}_m^-$ (see § 3)
F_{mo}	$(F_{mo}^++F_{mo}^-)/2$
Δ_{mo}	$F_{mo}^+-F_{mo}^-$
k_m	For a heavy-atom anomalous scatterer, f''/f' (see § 3)
s_m,v_m	See Fig. 9
w_m	$\Delta_{mo}+2k_mv_m$
σ_m^2	$\frac{1}{2}\delta_m^2+\varepsilon_m^2$
τ_m^2	$2\delta_m^2+4k_m^2\varepsilon_m^2$

2. Introduction

The multiple isomorphous replacement method as applied with macromolecular crystals does not in practice yield unambiguous phase determinations because of several sources of error. These include the experimental errors in measuring structure amplitudes; errors in the determination of positional, thermal and occupancy parameters for the heavy-atom substitution sites; and deviations from exact isomorphism. Blow & Crick (1959) have treated each structure factor for the 'parent' crystal as a random complex variable with a probability distribution $p(F,\varphi)$ derivable from all the data for that reflection. They defined the 'best' Fourier synthesis as that having the minimum expected mean-square error in electron density, as averaged over the whole unit cell. They showed that, given certain assumptions regarding the randomness of errors, the 'best' Fourier is obtained by taking for each coefficient the centroid $\mathbf{F}_C$ of the probability distribution $p(F,\varphi)$:

$$\mathbf{F}_C=\frac{\int_0^{2\pi}\int_0^{\infty}(F\exp i\varphi)p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi}{\int_0^{2\pi}\int_0^{\infty}p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi}. \qquad (1)$$

* A part of this work, treating crystals assumed to contain only normal scatterers, was done in the Department of Biochemistry, College of Physicians and Surgeons, Columbia University, New York, N.Y., and was given in Abstract A9, American Crystallographic Association Meeting, Suffern, N.Y., 1965. Preliminary reports of the analysis for derivatives containing heavy-atom anomalous scatterers were given in Abstract 10·11, Seventh International Congress, International Union of Crystallography, Moscow, USSR, 1966, and in Abstract G1, American Crystallographic Association Meeting, Minneapolis, Minn., 1967. Research sponsored by the US Energy Research and Development Administration under contract with the Union Carbide Corporation.

In deriving an expression for the probability distribution, they assumed that errors in the measured structure amplitudes have Gaussian distributions, and that the (complex) errors in the calculated structure-factor contributions due to heavy-atom substitutions have amplitudes with Gaussian distributions and random phases. They treated only 'normal' derivatives – those for which Bijvoet differences are considered negligible or are averaged. Blow & Rossman (1961), North (1965) and Matthews (1966) have extended the Blow–Crick treatment to 'anomalous' derivatives – those containing heavy-atom anomalous scatterers – for which the Bijvoet differences are utilized.

The treatment described here differs from that of Blow & Crick (1959) principally in the definition of the probability distributions and the manner in which they are combined. Blow & Crick made several assumptions which simplified the derivation of an expression for the probability distribution and the calculation of the centroid. They first considered the probability distribution, here denoted p_n^{BC}, for the parent structure factor as determined from data for the pair, parent and nth derivative. They obtained p_n^{BC} as a function of φ alone on the circle $F=F_o$, where F_o is the observed parent structure amplitude,* by defining it in a manner equivalent to the following:

$$p_n^{BC}(\varphi)=\int_0^\infty p_0(F)p_n(F,\varphi)\mathrm{d}F, \qquad (2)$$

where $p_0(F)$ is the probability distribution related to errors in F_o, and $p_n(F,\varphi)$ is that related to errors in F_{no}, the observed structure amplitude, and in $\mathbf{f}_n$, the calculated heavy-atom structure-factor contribution for the derivative. The correctness of equation (2) will be shown in § 7.†

They obtained as an approximation

$$p_n^{BC}(\varphi)=\exp\{-[x_{no}(\varphi)]^2/2E_n^2\}, \qquad (3a)$$

where $x_{no}(\varphi)$ is the lack-of-closure error [Fig. 1(*a*)] and E_n^2 is the variance [see equation (14)]. (This probability and all those following are unnormalized.) The joint probability, here denoted p^{BC}, was defined,

$$p^{BC}(\varphi)=\prod_1^N p_n^{BC}(\varphi), \qquad (3b)$$

where N is the number of derivatives. The centroid could then be calculated by one-dimensional numerical integration:

$$\mathbf{F}_C^{BC}=F_o\frac{\int_0^{2\pi}(\exp i\varphi)p^{BC}(\varphi)\mathrm{d}\varphi}{\int_0^{2\pi}p^{BC}(\varphi)\mathrm{d}\varphi}. \qquad (4)$$

Since each p_n^{BC} is an integral which involves the term p_0, these probabilities are not independent and therefore should not be multiplied as in equation (4). To do so gives too high a weight to the observed F_o.

Raiz & Andreeva (1970) have considered the probability, for the case of normal derivatives only, as a function of two dimensions, F and φ. A separate distribution was defined for each crystal form – parent and derivatives – so that these distributions were independent and could properly be multiplied to obtain the joint probability. The probability distribution for derivative n was defined [by their equation (3), in present nomenclature and without the normalization constant] as $p_n^{RA}(F,\varphi)\sim\exp[-x_n^2(F,\varphi)/2(E_n^{RA})^2]$, where the lack-of-closure error x_n varies with both F and φ, and the superscript RA denotes Raiz–Andreeva. $(E_n^{RA})^2$ was said to be the sum of the variance of F_{no} and a term related to errors in $\mathbf{f}_n$, as indicated by Blow & Crick's (1959) analysis of the distribution $p_n^{BC}(\varphi)$ [see equations (3) and (14)]. By equation (8) below, with the assumption that $x_n \ll F_{no}$, it may be seen that the second term in $(E_n^{RA})^2$ is ε_n^2, the mean-square amplitude of errors in $\mathbf{f}_n$. Then

$$p_n^{RA}(F,\varphi)\sim\exp[-x_n^2(F,\varphi)/2(\delta_n^2+\varepsilon_n^2)]. \qquad (5)$$

The mathematical treatment of this paper is applied to both normal and anomalous derivatives. The Blow–

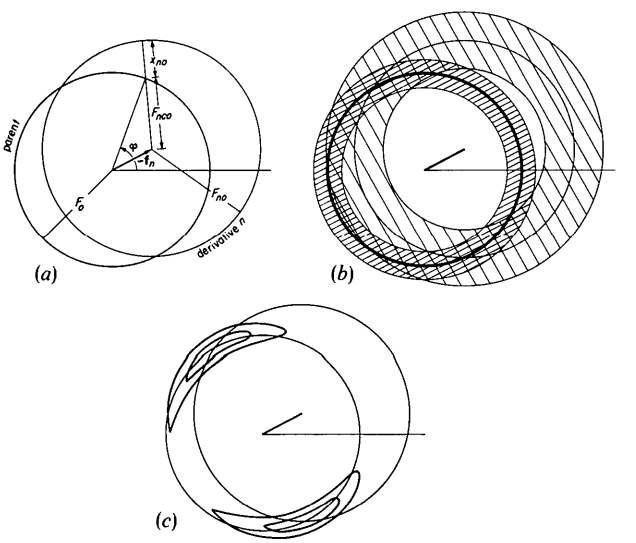

Fig. 1. (*a*) Construction showing the lack-of-closure error x_{no}, defined here as $F_{no}-F_{nco}$, the negative of the term as defined by Blow & Crick (1959). If the parent circle is replaced by one with arbitrary radius F, the lack-of-closure error is denoted x_n. (*b*) Shaded areas represent regions of high probability (p_0 and p_n). (*c*) Equiprobability contours for the joint probability p_0p_n.

* The term 'observed structure amplitude' will be used for simplicity, although, of course, the amplitudes are calculated from observed intensities.

† Equation (2) may not appear to correspond to Blow & Crick's Fig. 7(*b*), in which the integral was illustrated as if it related only to errors in $\mathbf{f}_n$ and F_{no}. However, it can be deduced from their discussion that the illustrated Gaussian probability distribution for errors in F_{no} (F_H in their figure) was actually a convolution of Gaussians for errors in F_o and F_{no}, a procedure which they showed to be valid if $f_n \ll F_{no}$. Moreover, the variance $\langle\delta^2\rangle$ of this illustrated Gaussian was defined as the mean-square error in F_{no}–F_o [their equation (3)].

Crick and Raiz–Andreeva assumption that $f_n \ll F_o$ is somewhat relaxed, so that better approximations to the 'true' probability distributions are obtained, for both normal and anomalous derivatives, than those of previous authors. The variances δ_j^2 of the observed structure amplitudes F_{jo}, and the variances ε_j^2 of the calculated heavy-atom structure factors $\mathbf{f}_j$, are carried separately throughout the analysis. In this way, the probability distributions (and calculated centroid) for a given reflection depend on the estimated precision of the observed structure amplitudes for that reflection. Estimation of the variances ε_j^2, following least-squares refinement of the heavy-atom parameters, is discussed below.

The treatment described here applies to acentric reflections. A related treatment of centric reflections is in preparation.

An elementary idea of the procedure followed here (and by Raiz & Andreeva, 1970) may be conveyed by Fig. 1(*b*,*c*). (The detailed treatment follows in § 3.) In Fig. 1(*b*) the probability distribution $p_0(F)$, related to errors in the observed parent amplitude, is indicated as a shaded annular region between circular equiprobability contours. The distribution $p_n(F,\varphi)$, related to errors in F_{no} and $\mathbf{f}_n$, is similarly indicated. This annulus is broader than that for p_0 because there is uncertainty in the position of the center of the derivative circle as well as in its radius. It is important to note that this distribution p_n represents the probability that (F,φ) is the structure factor of the *parent*, as related to errors in the amplitude and in the heavy-atom parameters for the *derivative*. Fig. 1(*c*) shows equiprobability contours of the joint distribution $p_0 p_n$. The joint distribution would be multiplied by another distribution of the same type as p_n for each additional normal derivative.

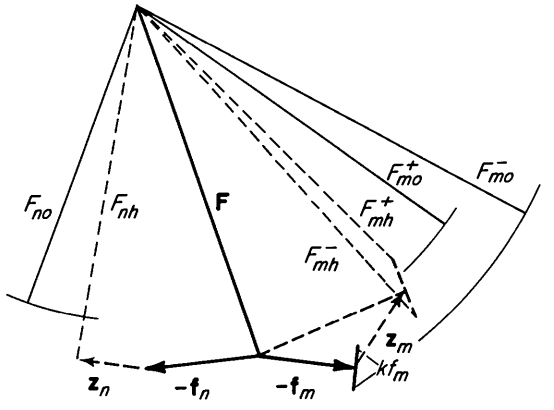

Fig. 2. Diagram of the right-hand sides of equation (6) for one normal (*n*) and one anomalous (*m*) derivative, where F_{no}, $\mathbf{f}_n$ and F^+_{mo}, F^-_{mo}, $\mathbf{f}_m$ are considered to be subject to error. F_{nh} is the hypothetical structure amplitude for the normal derivative if $\mathbf{z}_n$ is the error in $\mathbf{f}_n$; and F^+_{mh}, F^-_{mh} are the pair of hypothetical amplitudes for the anomalous derivative if $\mathbf{z}_m$ is the error in $\mathbf{f}_m$.

3. Probability distributions

The given quantities for a single reflection include (*a*) for the parent: the observed amplitude F_o; (*b*) for the *n*th normal derivative: the observed amplitude F_{no}, and the calculated structure-factor contribution due to the heavy-atom substitutions $\mathbf{f}_n$; (*c*) for the *m*th anomalous derivative: the observed amplitudes $F_{mo}(\mathbf{H})$ and $F_{mo}(-\mathbf{H})$ (to be denoted F^+_{mo} and F^-_{mo}), and the calculated structure-factor contributions $\mathbf{f}_m(\mathbf{H})$ and $\mathbf{f}_m(-\mathbf{H})$ (which are related to each other as discussed below). It is assumed in this paper that for all types of heavy-atom anomalous scatterers in any one derivative the same ratio k obtains, where

$$k \equiv f''/f',$$

and the scattering factor is $f' + if''$. The symbol $\mathbf{f}_m$ will hereafter be defined as follows:

$$\mathbf{f}_m \equiv \sum a_j f'_j \exp(2\pi i \mathbf{H}\mathbf{r}_j),$$

where the sum is over all the heavy-atom anomalous scatterers in the unit cell, and a_j is an occupancy factor. Then the structure-factor contributions are $\mathbf{f}_m(1+ik)$ for $\mathbf{H}$, and $\mathbf{f}^*_m(1+ik)$ for $-\mathbf{H}$. If there were no errors and the isomorphism were exact, the following relations would hold:†

$$\mathbf{F} = -\mathbf{f}_n + \mathbf{F}_n \tag{6a}$$

$$\mathbf{F} = -\mathbf{f}_m(1+ik) + \mathbf{F}^+_m \tag{6b}$$

$$\mathbf{F} = -\mathbf{f}_m(1-ik) + (\mathbf{F}^-_m)^* \tag{6c}$$

where $\mathbf{F}$, $\mathbf{F}_n$, $\mathbf{F}^+_m$ $[=\mathbf{F}_m(\mathbf{H})]$ and $\mathbf{F}^-_m$ $[=\mathbf{F}_m(-\mathbf{H})]$ are the (true) structure factors for the parent, *n*th normal derivative, and *m*th anomalous derivative.

Fig. 2 is a diagram of equations (6) for one normal and one anomalous derivative, where the observed structure amplitudes and calculated structure-factor contributions are considered to be subject to error. The probability that any complex number (F,φ) is the structure factor for the parent is the probability of having made all the implied errors.‡ The error in F_o would be $F_o - F$, the probability of which is given by equation (7*a*), where δ_0^2 is the applicable variance.

With the assumption that (F,φ) is the parent structure factor, the implied errors in $\mathbf{f}_n$ and F_{no} are interdependent. If the error in $\mathbf{f}_n$ were $\mathbf{z}_n$, then the hypothetical value of F_n would be $F_{nh}(\mathbf{F},\mathbf{z}_n)$, where $\mathbf{F}_{nh} = \mathbf{F} + \mathbf{f}_n - \mathbf{z}_n$ (see Fig. 2), and the error in F_{no} would be $(F_{no} - F_{nh})$. The probability of these two errors occurring together is given by the integrand of equation (7*b*)

† Equation (6*c*) is obtained by taking the complex conjugates of both sides of the equation, $\mathbf{F}^-_m = \mathbf{F}^* + \mathbf{f}^*_m(1+ik)$. It is assumed that anomalous scattering from protein atoms may be neglected.

‡ This statement is a consequence of Bayes' theorem. 'Probability' as used first in the sentence refers to 'posterior probability', a degree of belief. As used the second time with respect to experimental errors, it refers to a 'likehood'. See, for example, Lindley (1965) especially the discussions in section 1.6, part 1, and chapter 5, part 2.

where δ_n^2 and ε_n^2 are the variances, and the assumptions regarding the distributions of the errors are the same as those of Blow & Crick (see § 2). This expression must be integrated over the complex plane.

The errors in $\mathbf{f}_m$, F_{mo}^+, and F_{mo}^- are likewise interdependent. An assumed error $\mathbf{z}_m$ in $\mathbf{f}_m$ would imply the pair of errors $(F_{mo}^+ - F_{mh}^+)$, $(F_{mo}^- - F_{mh}^-)$ in the two amplitudes (see Fig. 2). The probability of these three errors is given by the integrand of equation (7*c*), where the same variance is assumed to apply to both amplitude measurements. This probability must also be integrated over the complex plane. Assuming that p_0 and all p_n and p'_m are independent,† the joint probability $p(F,\varphi)$ is given by their product [equation (7*d*)].

$$p(F) = \exp\left[-(F_o - F)^2/2\delta_0^2\right] \qquad (7a)$$

$$p_n(F,\varphi) = \int_{z_n} \exp\left[-(F_{no} - F_{nh})^2/2\delta_n^2 - z_n^2/2\varepsilon_n^2\right] \mathrm{d}\mathbf{z}_n \qquad (7b)$$

$$p'_m(F,\varphi) = \int_{z_m} \exp\left\{-\left[(F_{mo}^+ - F_{mh}^+)^2 + (F_{mo}^- - F_{mh}^-)^2\right]/2\delta_m^2 - z_m^2/2\varepsilon_m^2\right\} \mathrm{d}\mathbf{z}_m \qquad (7c)$$

$$p(F,\varphi) = p_0(F)\left[\prod_1^N p_n(F,\varphi)\right]\left[\prod_1^{N'} p'_m(F,\varphi)\right]. \qquad (7d)$$

It remains to perform the integrations of equations (7) so as to obtain p_n and p'_m as analytical functions of F and φ; then to reduce the integrations for the centroid (equation 1) to a one-dimensional form.

4. Integrations to obtain p and p'

Equations (7*b*,*c*) for p_n and p'_m cannot be integrated as closed forms, but have been integrated with the use of approximations. Since the derivations are lengthy, they are outlined in Appendix 1, with mathematical details in Appendix 3.

It was assumed for purposes of approximation that δ_j/F_{jo} and ε_j/F_{jo} are small. These are among the usual conditions that moderately good phase information be obtained with derivative j. (Infrequent exceptions are cases in which f_j/F_o is large and F_{jo} small.) If the conditions do not hold, and derivative j provides relatively poor phase information, cruder approximations for p_n and p'_m are acceptable. In the integrations the assumption of Blow & Crick (1959) and Raiz & Andreeva (1970) that $f_j \ll F_{jo}$ was somewhat relaxed. The result for a normal derivative (Appendix 1) is‡

$$p_n(F,\varphi) \sim (1 + \varepsilon_n^2 x_n/2\sigma_n^2 F_{no}) \exp\left(-x_n^2/2\sigma_n^2\right), \qquad (8)$$

where

$$\sigma_n^2 \equiv \delta_n^2 + \varepsilon_n^2.$$

$$x_n(F,\varphi) \equiv F_{no} - |\mathbf{F} + \mathbf{f}_n|.$$

Negative values of p_n, obtained for $x_n < -2\sigma_n^2 F_{no}/\varepsilon_n^2$, are meaningless, and if they cause the joint probability to be negative, it is set to zero (see § 5).

As shown in Fig. 8 for the integration in equation (7*b*), $p_n(F,\varphi)$ varies only with $|\mathbf{F} + \mathbf{f}_n|$, and therefore has circular symmetry about the terminus of $-\mathbf{f}_n$ in the Argand diagram (Fig. 1). As Fig. 1(*b*) roughly illustrates, the function $p_n(F,\varphi)$ resembles a ridge of high probability lying over the circle for the nth derivative in the diagram.

Fig. 3 illustrates comparisons of approximations obtained with equation (8), and of Raiz–Andreeva approximations (equation 5), with 'true' probabilities obtained by numerical integration of equation (7*b*). The agreement for equation (8) is to within a few percent (of the maximum value of p_n) for most of the range of interest, and is best for small ε_n. For large ε_n, p_n is a highly asymmetric function of x_n, to which equation (8) gives at least some degree of fit, whereas the Raiz–Andreeva and Blow–Crick probabilities are less

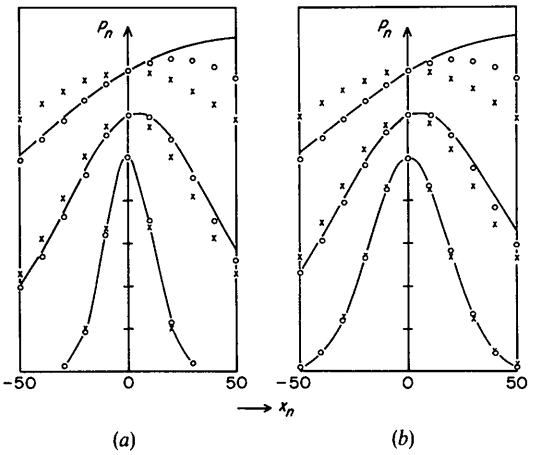

Fig. 3. Probability p_n for a normal derivative as a function of x_n. The curves represent numerical integrations of equation (7*b*). ○ approximations given by equation (8). × Raiz–Andreeva approximations, equation (5). For all calculations $F_{no} = 100$, and a scale factor was applied so that $p_n = 1$ for $x_n = 0$. (*a*) $\delta_n = 5$; (*b*) $\delta_n = 15$. In each case the bottom, middle and top plots are for $\varepsilon_n = 10$, 30 and 70. The marked ordinate (p_n) intervals are 0·2, and the plots are displaced from each other vertically by 0·2.

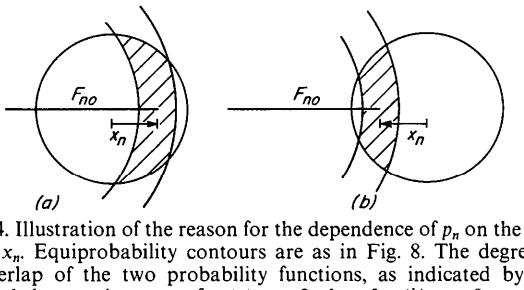

Fig. 4. Illustration of the reason for the dependence of p_n on the sign of x_n. Equiprobability contours are as in Fig. 8. The degree of overlap of the two probability functions, as indicated by the shaded areas, is greater for (*a*) $x_n > 0$, than for (*b*) $x_n < 0$.

† Frequently, several derivatives have heavy-atom sites in common. In such cases, certain systematic errors in the calculated $\mathbf{f}_j$'s – such as those due to departures from isomorphism – may be similar for these derivatives, so that the effects of these errors are enhanced by multiplying the probability distributions. The present treatment does not remove this problem.

‡ It is to be noted that $p_n(F,\varphi)$ does not depend on F_o or δ_0^2.

satisfactory, being symmetric with respect to x_n. The reason for the asymmetry is indicated in Fig. 4.

The same kinds of approximations were used in the integration of equation (7*c*) as for (7*b*) (see Appendices 1 and 3*). The result is

$$p'_m(F,\varphi)\sim(1+\delta^2\varepsilon^2 x/\sigma^2\tau^2 F_{mo}+4k^2\varepsilon^2 s/\tau^2 F_{mo})$$
$$\times\exp\left\{-\frac{x^2}{2\sigma^2}-\frac{w^2}{2\tau^2}\left[1+\frac{8k^2\varepsilon^2}{\tau^2 F_{mo}}\left(s-\frac{\tau^2+2\varepsilon^2}{4\sigma^2}x\right)\right]\right\}, \quad (9)$$

where most subscripts m are omitted for clarity, and where

$$F_{mo}\equiv\tfrac{1}{2}(F^+_{mo}+F^-_{mo})$$
$$\Delta_{mo}\equiv F^+_{mo}-F^-_{mo}$$
$$w_m\equiv\Delta_{mo}+2k_m v_m$$
$$\sigma_m^2\equiv\tfrac{1}{2}\delta_m^2+\varepsilon_m^2$$
$$\tau_m^2\equiv 2\delta_m^2+4k_m^2\varepsilon_m^2.$$

The lack-of-closure error x_m and the parameters s_m, v_m are shown in Fig. 9 and are defined in Table 1. The Bijvoet-difference error w_m is $\Delta_{mo}-(\Delta_m)_{\rm calc}$. Negative values of p'_m are meaningless, and if they cause the joint probability to be negative, it is set to zero (see § 5).

Approximations for p'_m given by equation (9) have been compared with numerical integrations of equation (7*c*) for all combinations of the following values:

$F_{mo}=100$; $k_m=0{\cdot}2$;* $f_m=25,50$;
$\Delta_{mo}=0,10,20,40$; $\delta_m=5,10,30$; $\varepsilon_m=10,30,70$.

F and φ were sampled over the region in which p'_m, evaluated numerically, is greater than 1% of its maximum value. A conventional R index, $\Sigma|p'(\text{numerical})-p'(\text{approx.})|/\Sigma p'(\text{numerical})$, was calculated over this region for each set of parameters. R ranged from 0·7 to 48%, with 8% as the median value. Examples of good, fair and poor agreement are shown in Fig. 5. As for normal derivatives, the poorer agreement for large ε may be accepted, since in these cases the phase-determining power is weak (see Fig. 5*g*,*h*).

R indices were also calculated for $p'=\exp(-x^2/2\sigma^2-w^2/2\tau^2)$, an expression which can be shown to be related to those of North (1965) and Matthews (1966) for the case of a single isomorphous pair [see § 7, equation (17)]. R ranged from 2 to 82%, with 22% as the median value. Examples of p' computed in this way are given in Fig. 5(*c*), (*f*) and (*i*). From geometric considerations it is clear that p', as calculated in this manner, must have a line of mirror symmetry perpendicular to $-\mathbf{f}_m$ and passing through the terminus of this vector. However, as shown in Fig. 5, the 'true' probability distributions as calculated by numerical integration, and the approximations to these distributions from equation (9), may be considerably asymmetric.

* See footnote ‡, p. 78.

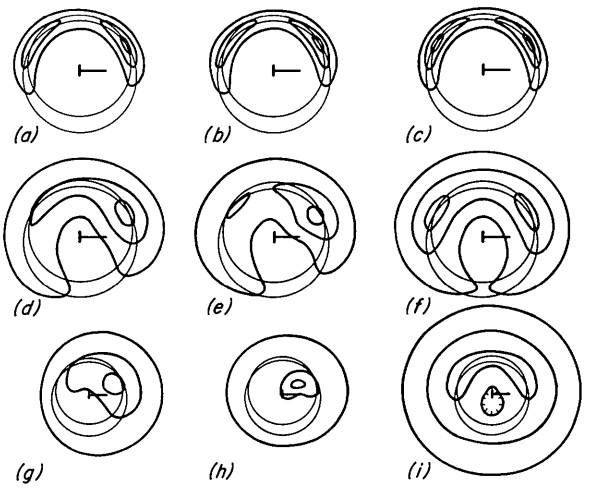

Fig. 5. Examples of probability distributions p'_m over the complex plane for an anomalous derivative by three different methods of calculation. For all calculations, $F_{mo}=100$, $\Delta_{mo}=10$, $f_m=50$, $\alpha_m=0$, $k_m=0{\cdot}2$ and $\delta_m=5$, while $\varepsilon_m=10$ for (*a*), (*b*), (*c*), 30 for (*d*), (*e*), (*f*), and 70 for (*g*), (*h*), (*i*). The approximation of equation (9) was used to calculate (*a*), (*d*), and (*g*). Numerical integrations of equation (7*c*) were used for (*b*), (*e*), and (*h*). The expression of equation (17) was used in the calculations for (*c*), (*f*), and (*i*). The termini of the two vectors $-\mathbf{f}_m(1\pm ik)$ (see equations 6*b*, 6*c*) are the centers of the circles with radii $F_{mo}\pm\Delta_{mo}/2$. Note the change of scale for (*g*), (*h*), and (*i*). The contours for each calculation are drawn at 10, 50, and 90% of the maximum p'_m for that calculation. The R values (see text), obtained by comparison with the corresponding numerical integration (on the same row of the figure) were (*a*) 3, (*c*) 9, (*d*) 9, (*f*) 22, (*g*) 37, and (*i*) 65%.

5. Calculation of centroid

After substitution from equations (7*a*,*d*), (8) and (9) into equation (1), the two-dimensional integrals for the centroid may be reduced to one-dimensional forms by the use of several approximations. As outlined in Appendix 2 and shown in full in Appendix 3,† x_m, s_m, w_m, and F^2 have been approximated as first-degree (two-term) Taylor series in $(F-F_o)$, expanded about the point $F=F_o$. In the resulting expressions all products higher than first degree in $(F-F_o)/F_o$ have been neglected. Furthermore, the lower limit of integration along F has been changed from 0 to $-\infty$. These approximations are valid for reasonably small δ_0/F_o. A poorer degree of approximation is acceptable for reflections with data of very low precision. Then as outlined in Appendix 2, the integrals for the centroid reduce to the one-dimensional forms of equation (10), Table 1. If, because of the various approximations, either integrand of equation (10) should be negative for a given φ, both integrands are set to zero for that φ.

The computing time is not much increased over that required for previous treatments. In a program available from the author, no additional transcendental

* The maximum value for k in current practice is usually about 0·2 (for uranium, Cu $K\alpha$ radiation, $d=2{\cdot}5$ Å), except for some rare-earth elements, for which k approaches 0·3 under the same conditions.

† See footnote ‡, p. 78.

functions are calculated, and most of the arithmetic is done once per reflection for each derivative.

6. Estimation of variances

The variances δ^2 can be estimated in the usual manner for the given method of intensity measurement. If there is at least one centrosymmetric zone, the variances ε^2 can be estimated, as described by Blow & Crick (1959), as follows. If anomalous-scattering effects are neglected, the structure factors for centric reflections are real, and $f_j = |F_{jo} \pm F_o|$ if there are no errors. Let

$$E_j^2 = \overline{(|F_{jo} \pm F_o| - f_j)^2}, \tag{13}$$

where the sign is chosen for each reflection so as to minimize the expression. Then E_j^2 is the sum of the average variances of F_o, F_{jo} and f_j. In the nomenclature of this paper,

$$E_j^2 = \overline{\delta_0^2} + \overline{\delta_j^2} + \varepsilon_j^2, \tag{14}$$

so that

$$\varepsilon_j^2 = E_j^2 - \overline{(\delta_0^2 + \delta_j^2)}. \tag{15}$$

Since ε_j^2 is expected to vary with Bragg angle (Crick & Magdoff, 1956), it should be evaluated for limited Bragg-angle ranges, large enough to contain statistically significant numbers of reflections.

When anomalous scattering is taken into account, the heavy-atom structure factors for centric reflections have imaginary components which, under usual experimental conditions, have amplitudes less than $\frac{1}{4}$ the real components. The relation $f_m \sim |F_m \pm F_o|$ still holds to a good approximation (if there are no errors) for all but a small number of reflections – those for which the real part of $\mathbf{f}_m$ is approximately $-F_o$. Therefore, the effect of the imaginary components may be neglected, or alternatively, reflections for which $f_m \sim F_o$ and F_{mo} is small might be omitted from the calculation.

The heavy-atom parameters and the scale factors for derivative structure amplitudes, used in evaluating E_j^2 (equation 13), can be refined by least-squares procedures (Kartha, 1965; Dickerson, Weinzierl & Palmer, 1968). In the procedure of Dickerson *et al.* (1968) alternate cycles of refinement and phasing (calculation of centroids) are employed, where in the refinement $\Sigma_h w_h x_{jh}^2$ is minimized for each derivative j, the sum being over all reflections h at either the 'best' or most probable phase for the parent. If this procedure is used, the ε_j's would be re-evaluated after each refinement cycle by the use of equations (13) and (15) for centric zones. The δ_j's, however, would remain constant throughout all cycles.

For cases with no centrosymmetric zone, the estimation of the variances ε_j^2 is a more difficult problem, beyond the scope of this paper.

Table 1. *Equations for calculation of the centroid*

For clarity, subscripts n or m have been omitted from the symbols δ, ε, σ, τ, and k; subscripts mo from s, v, Δ, and w; and subscripts no or mo from x.

$$\mathbf{F}_C \sim F_o \frac{\int_0^{2\pi} (\exp i\varphi) A[1 + (2/F_o + B)D/C] \exp(D^2/C + G)\mathrm{d}\varphi}{\int_0^{2\pi} A[1 + (1/F_o + B)D/C] \exp(D^2/C + G)\mathrm{d}\varphi} \tag{10}$$

$$A = \prod_0^{N_t} A_j, \quad B = \sum_0^{N_t} B_j/A_j, \quad C = \sum_0^{N_t} C_j, \quad D = \sum_0^{N_t} D_j, \quad G = \sum_0^{N_t} G_j$$

Normal derivative

$$\begin{aligned} A_n &= 1 + \varepsilon^2 x/2\sigma^2 F_{no} \\ B_n &= -\Omega\varepsilon^2/2\sigma F_{no} \\ C_n &= \Omega^2/2\sigma^2 \\ D_n &= \Omega x/2\sigma^2 \\ G_n &= -x^2/2\sigma^2 \\ \sigma^2 &= \delta^2 + \varepsilon^2 \end{aligned}$$

Parent

$$A_0 = 1;\ B_0 = D_0 = G_0 = 0; \quad C_0 = 1/2\delta_0^2$$

Anomalous derivative

$$\begin{aligned} A_m &= 1 + (\varepsilon^2/\tau^2 F_{mo})(\delta^2 x/\sigma^2 + 4k^2 s) \\ B_m &= -(\Omega\varepsilon^2/\tau^2 F_{mo})(\delta^2/\sigma^2 + 4k^2) \\ C_m &= \Omega^2/2\sigma^2 \\ D_m &= \Omega(x/2\sigma^2 + k^2\varepsilon^4 w^2/\sigma^2\tau^4 F_{mo}) \\ G_m &= -\frac{x^2}{2\sigma^2} - \frac{w^2}{2\tau^2}\left\{1 + \frac{8k^2\varepsilon^2}{\tau^2 F_{mo}}\left[s - \frac{\tau^2 + 2\varepsilon^2}{4\sigma^2}x\right]\right\} \\ \sigma^2 &= \tfrac{1}{2}\delta^2 + \varepsilon^2 \\ \tau^2 &= 2\delta^2 + 4k^2\varepsilon^2 \\ s &= s_{mo} = (f_m/F_{mco})[f_m + F_o\cos(\varphi - \alpha_m)] + x \\ v &\equiv v_{mo} = -f_m F_o \sin(\varphi - \alpha_m)/F_{mco} \\ w &\equiv w_{mo} = \Delta + 2kv \\ F_{mo} &= \tfrac{1}{2}(F_{mo}^+ + F_{mo}^-) \\ \Delta &= F_{mo}^+ - F_{mo}^- \\ k &= f''/f' \end{aligned} \tag{12}$$

All derivatives

$$\left.\begin{aligned} x &\equiv x_{jo} = F_{jo} - F_{jco} \\ F_{jco} &= [F_o^2 + 2f_j F_o\cos(\varphi - \alpha_j) + f_j^2]^{1/2} \\ \Omega_j &= [F_o + f_j\cos(\varphi - \alpha_j)]/F_{jco} \\ (f_j, \alpha_j)&: \text{ see § 3} \end{aligned}\right\} \tag{11}$$

7. Comparisons with previous treatments

For the case of a single isomorphous pair, parent and one normal derivative, the centroid of Table 1 is roughly the same as that of Blow & Crick (equations 3, 4). It can easily be shown by Table 1 and equations (3) and (14) that

$$\exp(D^2/C+G)=\exp[-x_{no}^2/2(\delta_0^2+\delta_n^2+\varepsilon_n^2)]$$
$$=\exp(-x_{no}^2/2E_n^2)=p_n^{BC}(\varphi).$$

Then the integrands of equations (4) and (10) are the same if the nonexponential factors of equation (10) are omitted. For cases with more than one derivative, the approximate correspondence between the treatments breaks down, and the centroids can differ appreciably (see § 9, example 1).

For the case of a parent and one anomalous derivative, with the same value of k for all heavy-atom anomalous scatterers, both North (1965, equations 2, 5) and Matthews (1966, equations 1–7) gave the probability $p(\varphi)$ in forms equivalent to the following:

$$p(\varphi)=\exp(-x_{mo}^2/2E^2)\exp[-(\Delta_H+2\delta\sin\gamma)^2/2E'^2], \quad (16)$$

where by comparison of Fig. 9 with North's Fig. 3 and Matthews's Fig. 1, and by the definitions given for Δ_H,

$$(\Delta_H+2\delta\sin\gamma)^2=(\Delta_{mo}+2k_m v_{mo})^2=w_{mo}^2 \quad \text{(this paper)}.$$

The centroid obtained by using the probability expression of equation (16) in equation (4), as North and Matthews did, is roughly the same as that from Table 1, if appropriate values are assigned to E^2 and E'^2. If terms in k_m^2/F_{mo} are neglected, we obtain from Table 1:

$$\exp(D^2/C+G)\sim\exp[-x_{mo}^2/2(\delta_0^2+\tfrac{1}{2}\delta_m^2+\varepsilon_m^2)]$$
$$\times\exp(-w_{mo}^2/2\tau_m^2),$$

so that if the nonexponential factors of equation (10) are also omitted, the integrands of equations (4) and (10) are the same, with

$$E^2=\delta_0^2+\tfrac{1}{2}\delta_m^2+\varepsilon_m^2 \quad \text{and} \quad E'=\tau_m.$$

These approximations in equation (10) correspond to the approximations: $F\sim F_o$, and (see equation 9)

$$p'_m(F,\varphi)\sim\exp(-x_m^2/2\sigma_m^2-w_m^2/2\tau_m^2). \quad (17)$$

The latter expression was used for the calculations illustrated in Fig. 5(*c*), (*f*) and (*i*). For more than one derivative, the correspondence between the treatments breaks down.

8. Error in electron density

Blow & Crick (1959, equation 13) derived an expression for the expected mean square error in electron density, averaged over the entire unit cell. In the nomenclature of this paper,

$$\langle(\Delta\varrho)^2\rangle=\frac{2}{V^2}\sum_{hkl}\frac{\int\int|\mathbf{F}-\mathbf{F}_C|^2p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi}{\int\int p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi}, \quad (18)$$

where V is the unit-cell volume. They also showed that the contribution from one reflection is

$$\langle(\Delta\varrho)^2\rangle_{hkl}=\frac{2}{V^2}\left\{\frac{\int\int F^2p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi}{\int\int p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi}-F_C^2\right\}. \quad (19)$$

With the Blow–Crick treatment of the probability, equation (19) reduces to

$$\langle(\Delta\varrho)^2\rangle_{hkl}=[2(F_o)_{hkl}^2/V^2](1-m_{hkl}^2), \quad (20)$$

where m, the figure of merit, was defined (Dickerson, Kendrew & Strandberg, 1961) as $m\equiv F_C/F_o$. With the Blow–Crick treatment m must be less than 1, but with the present treatment m may (in unusual cases) be greater than 1 (see example 1). It is therefore not exactly a 'figure of merit' here.

The expected error in the electron density can be estimated by carrying out the integrations in equation (19), using the probability distributions of Table 1 and the same substitutions and approximations as in § 5. The integral in the denominator is the same as that in equation (10), Table 1. For the numerator,

$$\int_0^{2\pi}\int_0^{\infty}F^2p(F,\varphi)F\mathrm{d}F\mathrm{d}\varphi$$
$$\sim F_o^3\int_0^{2\pi}A[1+(3/F_o+B)D/C]\exp(D^2/C+G)\mathrm{d}\varphi.$$

Then

$$\langle(\Delta\varrho)^2\rangle_{hkl}\sim\frac{2(F_o^2)}{V^2}$$
$$\times\left\{\frac{\int_0^{2\pi}A[1+(3/F_o+B)D/C]\exp(D^2/C+G)\mathrm{d}\varphi}{\int_0^{2\pi}A[1+(1/F_o+B)D/C]\exp(D^2/C+G)\mathrm{d}\varphi}-\frac{F_C^2}{F_o^2}\right\}. \quad (21)$$

The integral in the numerator of equation (21) can easily be evaluated numerically at the same time as the integrals of equation (10), Table 1 (see example 1, § 9). A new figure of merit m' can be defined by setting the expression in braces in equation (21) equal to $1-(m')^2$, producing an expression analogous to equation (20).

9. Examples

As stated above, an effect of the simplifying assumptions made by Blow & Crick was to give excessive weight to the observed parent amplitude when $N>1$. Accordingly, consider the following idealized case: circles for three normal derivatives cross at a point off the parent circle (Fig. 6). This is an extreme form of the case of an unusually large error in a parent amplitude. Centroids calculated by the Blow–Crick procedure (equations 3, 4) and by the present one (Table 1) for particular values of the variances are shown in Fig. 6. Note that the centroid from Table 1 lies much closer to the intersection of the derivative circles and to the peak of the probability distribution (calculated as in Table 1) than does the Blow–Crick centroid, which is displaced toward the φ range in which the derivative circles together most closely approach the parent circle. This is a case in which the centroid from Table 1 has an amplitude greater than F_o, and thus corresponds to an apparent figure of merit, F_C/F_o, greater than 1. However, the new figure of merit m' (§ 8) is 0·91.

A second example involving a single isomorphous pair, parent and one anomalous derivative, is shown in Fig. 7. Centroids calculated as in Table 1 are compared with that from equations (4) and (16), where E' is taken as $E/3$, as suggested by Matthews (1966). Values chosen for δ_0 and δ_1 (taken as equal) and for ε_1 are such that E^2 $(=\delta_0^2+\delta_1^2+\varepsilon_1^2)$ remains constant, so that the centroid from equations (4) and (16) is always the same (for the chosen values of F_o, F_{1o}, Δ_1 and $\mathbf{f}_1$). However, the centroids given by Table 1 differ appreciably for the various combinations of variances.

10. Discussion

Theoretically, the use of an improved procedure for combining the data should result in a more nearly error-free Fourier, but it is difficult to predict the extent of improvement. Clearly, and as has been recognized, the Blow–Crick procedure weights too heavily the measurement of the parent amplitude, and the overweighting becomes more pronounced as the number of derivatives increases (there is none for a single isomorphous pair). Therefore, one would expect the present treatment to be most useful in cases with many derivatives to minimize the effect of errors in the parent amplitudes. The better approximations illustrated in Figs. 3 and 5, and the use of variances which depend on the precision of the observations for each reflection, should also give improved results. Even if the improvement in a particular case resulted in the correct placement of only one additional residue of the macromolecule, the small increase in computer time would be justified.

The author is most grateful to Drs H. A. Levy and V. R. R. Uppuluri for helpful criticisms of the manuscript, and to Professor Barbara W. Low for her continued interest. A part of this work, treating crystals assumed to contain only normal scatterers, was done in the Department of Biochemistry, College of Physicians & Surgeons, Columbia University, New York, N.Y. This research was sponsored by the U.S. Energy Research and Development Administration under contract with the Union Carbide Corporation.

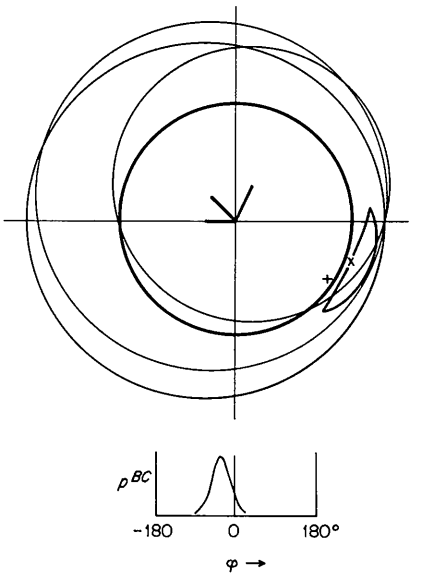

Fig. 6. Probability distribution, centroid from Table 1 (×), and Blow–Crick centroid (equations 3, 4) (+) for first example (see text). Parent amplitude, 100. The vectors $-\mathbf{f}_n$ are shown as heavy lines. The derivative circles cross at (125, 0°). All standard deviations, δ_n and ε_n, are assumed to be 10. The half-maximum contour is shown for $p(F,\varphi)$ as calculated by Table 1. Inset: plot of $p^{BC}(\varphi)$.

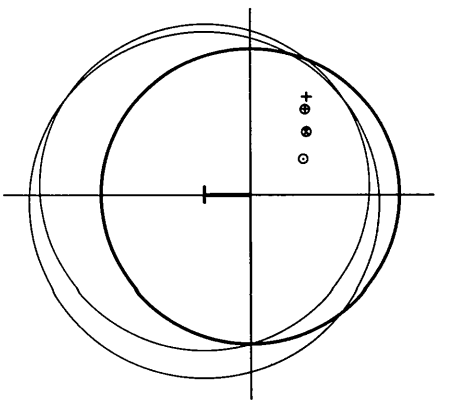

Fig. 7. Centroids from Table 1, and from equations (4) and (16), (Blow–Crick–North–Matthews procedure), for second example (see text). The heavy lines represent $-\mathbf{f}_1$ $(1\pm ik)$. Parent amplitude, 100. The standard deviations for the calculations according to Table 1 were assumed to have the following values, where $\delta=\delta_0=\delta_1$: ⊕ $\delta=3, \varepsilon_1=16{\cdot}8$; ⊗ $\delta=6, \varepsilon=15{\cdot}1$; ⊙ $\delta=10, \varepsilon_1=10$. In all cases $E=17{\cdot}3$ (see text). The symbol + shows the centroid as calculated by equations (5) and (18), with $E'=E/3$, for all three pairs of standard deviations.

APPENDIX 1

Integration to obtain p

Fig. 8, which is directly related to the construction for a normal derivative in Fig. 2, is similar to Fig. 7 of Blow & Crick (1959), except that F does not necessarily equal F_o. The 'lack-of-closure' error, $F_{no}-|\mathbf{F}+\mathbf{f}_n|$, is denoted x_n. Circular equiprobability contours are shown for the probability distributions related to the two errors, $\mathbf{z}_n$ and $(F_{no}-F_{nh})$. The complex error $\mathbf{z}_n$ has Cartesian coordinates $x+x_n$, y with respect to the orthogonal axes shown in Fig. 8.

In integrating equation (7*b*) several approximations will be used, which are valid over a limited region of x,y since the exponential integrand decreases rapidly toward zero outside a central region. Values of the approximate expression thus obtained for the integrals are compared (in § 4) with numerical integrations of equation (7*b*) for wide ranges of the parameters in order to show that the approximations are satisfactory for practical purposes.

Assume that δ_n/F_{no} and ε_n/F_{no} are small (see § 4). Then, as may be seen from Fig. 8, the probability $p_n(F,\varphi)$ will be small except in the region in which $|x|/F_{no}$ and $|y|/F_{no}$ are both small, implying small relative errors in F_{no} and in $\mathbf{f}_n$ respectively. Thus F_{nh} may be approximated by a series which retains only first and second degree terms in x/F_{no} and y/F_{no}, and it may be shown (Appendix 3) that

$$(F_{no}-F_{nh})^2 \sim x^2+xy^2/F_n. \tag{22}$$

By substitution into equation (7*b*),

$$p_n \sim \int_{-\infty}^{\infty} \exp\left[-x^2/2\delta_n^2-(x+x_n)^2/2\varepsilon_n^2\right] \times \int_{-\infty}^{\infty} \exp\left[-(x/2\delta_n^2F_{no}+1/2\varepsilon_n^2)y^2\right]\mathrm{d}y\mathrm{d}x. \tag{23}$$

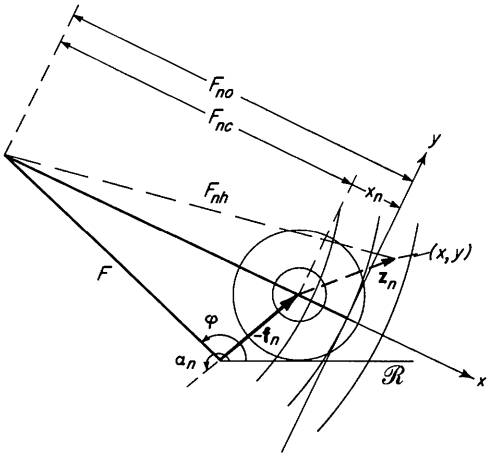

Fig. 8. Construction for integration of p_n. F_{no} is laid off from the point (F,φ) rather than from the terminus of $-\mathbf{f}_n$ as in Fig. 1. Equiprobability contours for the errors $(F_{no}-F_{nh})$ and $\mathbf{z}_n$ are at arbitrary intervals.

The integrals can be evaluated by means of a known formula (Bierens de Haan, 1957):

$$\int_{-\infty}^{\infty} (A+Bx)\exp(-Cx^2+2Dx+G)\mathrm{d}x = (\pi/C)^{1/2}(A+BD/C)\exp(D^2/C+G), \tag{24}$$

where $C>0$. Omitting constant factors as shall be done for all probabilities, the integral over y is $(1+\varepsilon_n^2x/\delta_n^2F_{no})^{-1/2}$, and is defined for $x>-\delta_n^2F_{no}/\varepsilon_n^2$. The integral becomes infinite at this limit only because of the use of the approximation of equation (22). For small x, the integral is approximately $1-\varepsilon_n^2x/2\delta_n^2F_{no}$, an expression which is defined for all x. By substitution in equation (23) and the use of equation (24), one obtains equation (8).

Integration to obtain p'

In Fig. 9, which is directly related to the construction for an anomalous derivative in Fig. 2, the following definitions are used: $F_{mo}\equiv\frac{1}{2}(F_{mo}^{+}+F_{mo}^{-})$, $\Delta_{mo}\equiv F_{mo}^{+}-F_{mo}^{-}$. The lack-of-closure error, x_m, is measured from the terminus of $-\mathbf{f}_m$ (see Fig. 9 and the definition of $\mathbf{f}_m$ in § 3). The origin of the diagram in Fig. 2 has the coordinates $-s$, $-v$ in Fig. 9. With the same kinds of approximations discussed with regard to equation (22), it may be shown (Appendix 3) that

$$[(F_{mo}^{+}-F_{mh}^{+})^2+(F_{mo}^{-}-F_{mh}^{-})^2]/2 \sim x^2+(ky+w/2)^2+[xy^2-ky(2ky+w)(x+s)]/F_{mo}, \tag{25}$$

where w, the Bijvoet-difference error, is given by $w_m\equiv\Delta_{mo}+2k_mv_m$.

In the following, second-degree terms in x/F_{mo} and s/F_{mo}, and $2k^2$, are neglected compared to one. After substitution of the expression above into equation (7*c*) and replacement of z_m^2 by $(x+x_m)^2+y^2$, the integral over y is obtained by the use of equation (24) (see Appendix 3). As in the derivation of p_n, it is ignored that C becomes negative as $x\to-\infty$. Omitting the constant factor, the integral is

$$[1-2\varepsilon^2(x-2k^2s)/\tau^2F_{mo}]\exp\{(k^2\varepsilon^2w^2/\tau^2\delta^2) \times[1-2(\tau^2+2\varepsilon^2)x/\tau^2F_{mo}-4\delta^2s/\tau^2F_{mo}] -(4x^2+w^2)/4\delta^2-(x+x_m)^2/2\varepsilon^2\},$$

where $\tau^2\equiv2\delta^2+4k^2\varepsilon^2$. Again applying equation (24) to integrate over x, one obtains equation (9) (see Appendix 3).

APPENDIX 2

Integration to obtain the centroid

Approximations for x_j, s_m and v_m are obtained as the first two terms of Taylor series, expanded about $F=F_o$. As shown in Appendix 3,

$$\begin{aligned} x_j &\sim x_{jo}-\Omega_Q, \\ v_m &\sim v_{mo}, \\ s_m &\sim s_{mo}-\Omega_Q, \\ w_m &\sim w_{mo}, \\ \Omega_j &\equiv [F_o+f_j\cos(\varphi-\alpha_j)]/F_{jco}, \end{aligned} \tag{26}$$

where x_{jo}, v_{mo}, s_{mo}, w_{mo} are the values at $F=F_o$, and $\varrho \equiv F-F_o$. Values of Ω_j are in the range $\{-1,1\}$, and are close to one for small values of f_j/F_o. After substitution of the approximations, equations (7*a*), (8) and (9) have the following form:

$$p_j(\varrho,\varphi)=(A_j+B_j\varrho)\exp(-C_j\varrho^2+2D_j\varrho+G_j), \quad (27)$$

where $A_j,\ldots,G_j$ are functions of φ but not of ϱ, and are given in Table 1.

Substituting for F by the identity, $F\equiv F_o(1+\varrho/F_o)$, and combining equations (1) and (7*d*), one obtains

$$\mathbf{F}_C=F_o\frac{\int_0^{2\pi}(\exp i\varphi)\int_{-F_o}^{\infty}(1+\varrho/F_o)^2\prod_0^{N_t}p_j(\varrho,\varphi)\mathrm{d}\varrho\mathrm{d}\varphi}{\int_0^{2\pi}\int_{-F_o}^{\infty}(1+\varrho/F_o)\prod_0^{N_t}p_j(\varrho,\varphi)\mathrm{d}\varrho\mathrm{d}\varphi}, \quad (28)$$

where each p_j is given by the appropriate expression in the form of equation (27). The term $(\varrho/F_o)^2$ in the numerator will be neglected in comparison with one, and the lower limit of integration over ϱ will be changed to $-\infty$, since for reasonably small δ_0/F_o, p_0 decreases rapidly with $|\varrho|$, and to a very small value for $\varrho=-F_o$. The product of the nonexponential terms of the p_j's is $\Pi(A_j+B_j\varrho)=\Pi(A_j)\Pi(1+B_j\varrho/A_j)$. The products of the terms $B_j\varrho/A_j$ with each other and with ϱ/F_o may also be neglected. For normal derivatives, $|B_n\varrho/A_n|\sim\varepsilon^2\varrho/2\sigma^2F_{no}<\varrho/2F_{no}$, since $\varepsilon^2<\sigma^2$; and for anomalous derivatives, $|B_m\varrho/A_m|\sim\delta^2\varepsilon^2\varrho/\sigma^2\tau^2F_{mo}<\varrho/2F_{mo}$, since $\varepsilon^2<\sigma^2$ and $\delta^2<\tau^2/2$. Then

$$\mathbf{F}_C\sim F_o\frac{\int_0^{2\pi}(\exp i\varphi)(A\exp G)\int_{-\infty}^{\infty}\left[1+\left(\frac{2}{F_o}+B\right)\varrho\right]\exp(-C\varrho^2+2D\varrho)\mathrm{d}\varrho\mathrm{d}\varphi}{\int_0^{2\pi}(A\exp G)\int_{-\infty}^{\infty}\left[1+\left(\frac{1}{F_o}+B\right)\varrho\right]\exp(-C\varrho^2+2D\varrho)\mathrm{d}\varrho\mathrm{d}\varphi}.$$

$A,\ldots,G$ are as given in equation (10), Table 1, and are functions of φ only. Integrations over ϱ by the formula of equation (24) yield the expression for the centroid given in equation (10), Table 1.

APPENDIX 3

Derivations of some results in Appendices 1 and 2

Derivation of equation (22)

From Fig. 8,

$$F_{nh}^2=(F_{no}+x)^2+y^2,$$
$$F_{nh}^2/F_{no}^2=1+2x/F_{no}+(x^2+y^2)/F_{no}^2.$$

Expressing the square root by a binomial series, and neglecting terms higher than second degree in x/F_{no} and y/F_{no},

$$F_{nh}/F_{no}\sim 1+x/F_{no}+y^2/2F_{no}^2,$$
$$F_{nh}-F_{no}\sim x+y^2/2F_{no}.$$

Then

$$(F_{no}-F_{nh})^2\sim x^2+xy^2/F_{no},$$

neglecting the higher-degree term.

Derivation of equation (25)

From Fig. 9, omitting most subscripts m,

$$x^{\pm}=x\mp k(y+v)$$
$$y^{\pm}=y\pm k(x+s)$$

$$(F_{mh}^{+})^2=(F_{mo}+x^{+})^2+(y^{+})^2.$$

With the same approximation used in the derivation above of equation (22),

$$F_{mh}^{+}\sim F_{mo}+x^{+}+(y^{+})^2/2F_{mo}.$$

Since $F_{mo}^{+}\equiv F_{mo}+\Delta/2$,

$$(F_{mo}^{-}-F_{mh}^{-})^2\sim(x^{-}+\Delta/2)^2+(x^{-}+\Delta/2)(y^{-})^2/F_{mo}.$$

where the term $(y^{+})^4/F_{mo}^2$ is neglected. Similarly,

$$(F_{mo}^{+}-F_{mh}^{+})^2\sim(x^{+}-\Delta/2)^2+(x^{+}-\Delta/2)(y^{+})^2/F_{mo},$$

Now

$$x^{\pm}\mp\Delta/2=x\mp(ky+w/2)$$

where $w\equiv 2kv+\Delta$, and

$$\tfrac{1}{2}[(x^{+}-\Delta/2)^2+(x^{-}+\Delta/2)^2]=x^2+(2ky+w)^2/4. \quad (28)$$

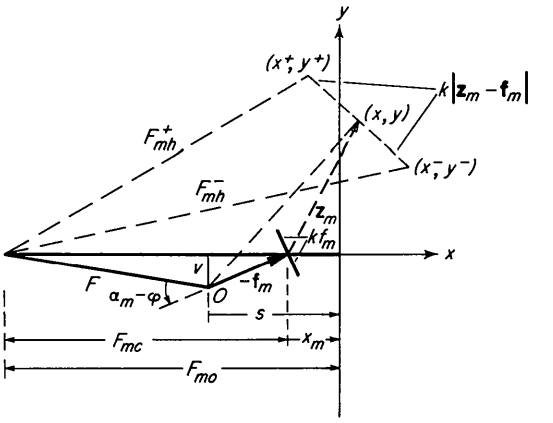

Fig. 9. Construction for integration of p_m'. For equations (12), $F=F_o$.

Also

$$(x^{\pm}\mp\Delta/2)(y^{\pm})^2 = [x\mp(ky+w/2)][y^2\pm 2ky(x+s)+k^2(x+s)^2]$$

$$\tfrac{1}{2}[(x^{+}-\Delta/2)(y^{+})^2+(x^{-}+\Delta/2)(y^{-})^2] = x[y^2+k^2(x+s)^2]-ky(2ky+w)(x+s). \quad (29)$$

Division of equation (29) by F_{mo} and addition to equation (28) yields equation (25), where the term $k^2x(x+s)^2/F_{mo}$ has been omitted because of the factor k^2/F_{mo}. [The term $-2k^2y^2(x+s)/F_{mo}$ has been retained because to do so leads to a neater expression in equation (9).]

Integration of equation (7*c*)

Throughout this derivation, terms higher than first degree in x/F_{mo} and s/F_{mo} are neglected. After substitution from equation (25) into equation (7*c*), the exponent in equation (7*c*) is

$$-\{x^2+(2ky+w)^2/4 + [xy^2-ky(2ky+w)(x+s)]/F_{mo}\}/\delta^2 - [(x+x_m)^2+y^2]/2\varepsilon^2.$$

Expressing the integral over y in the form of equation (24),

$$C=[k^2+x/F_{mo}-2k^2(x+s)/F_{mo}]/\delta^2+1/2\varepsilon^2 \sim(\tau^2/4\delta^2\varepsilon^2)[1+4\varepsilon^2(x-2k^2s)/\tau^2F_{mo}],$$

where $\tau^2\equiv 2\delta^2+4k^2\varepsilon^2$ and $2k^2$ has been neglected compared with one,

$$D=-kw[1-(x+s)/F_{mo}]/2\delta^2,$$
$$G=-(4x^2+w^2)/4\delta^2-(x+x_m)^2/2\varepsilon^2.$$

Expressing C^{-1} and $C^{-1/2}$ as power series, and neglecting higher-degree terms,

$$C^{-1}\sim(4\delta^2\varepsilon^2/\tau^2)[1-4\varepsilon^2(x-2k^2s)/\tau^2F_{mo}],$$
$$C^{-1/2}\sim(2\delta\varepsilon/\tau)[1-2\varepsilon^2(x-2k^2s)/\tau^2F_{mo}].$$

Also,

$$D^2\sim(k^2w^2/4\delta^4)[1-2(x+s)/F_{mo}].$$

As in the derivation of equation (8), it is ignored that C becomes negative as $x\to-\infty$. By equation (24), omitting the constant factor and neglecting higher-degree terms, the integral over y is approximately

$$[1-2\varepsilon^2(x-2k^2s)/\tau^2F_{mo}]\exp\{(k^2\varepsilon^2w^2/\tau^2\delta^2) \times[1-2(\tau^2+2\varepsilon^2)x/\tau^2F_{mo}-4\delta^2s/\tau^2F_{mo}] -(4x^2+w^2)/4\delta^2-(x+x_m)^2/2\varepsilon^2\}.$$

The coefficients for integration over x by equation (24) are:

$$A=1+4k^2\varepsilon^2s/\tau^2F_{mo},$$
$$B=-2\varepsilon^2/\tau^2F_{mo},$$
$$C=1/\delta^2+1/2\varepsilon^2=\sigma^2/\delta^2\varepsilon^2,$$

where $\sigma^2\equiv\frac{1}{2}\delta^2+\varepsilon^2$,

$$D=-k^2\varepsilon^2(\tau^2+2\varepsilon^2)w^2/\tau^4\delta^2F_{mo}-x_m/2\varepsilon^2,$$
$$G=(k^2\varepsilon^2w^2/\tau^2\delta^2)(1-4\delta^2s/\tau^2F_{mo})-w^2/4\delta^2-x_m^2/2\varepsilon^2.$$

Then, with the same kinds of approximations as above, the integral is as given in equation (9).

Derivation of equations (26)

From Figs. 8 and 9,

$$F_{jc}^2=F^2+2Ff\cos\mu+f^2 \quad (30)$$
$$v=-Ff(\sin\mu)/F_{jc}$$
$$s=(f^2-v^2)^{1/2}+x_j,$$

where $x_j\equiv F_{jo}-F_{jc}$ and $\mu\equiv\varphi-\alpha_j$. In order to obtain two-term Taylor series, expanded about the point $F=F_o$, for x_j, v, and s, the first derivatives are required:

$$\partial x_j/\partial F=-\partial F_{jc}/\partial F=-(F+f\cos\mu)/F_{jc}$$
$$\partial v/\partial F=-(f\sin\mu)(F_{jc}^2-F^2-Ff\cos\mu)/F_{jc}^3 =-(Ff^2\sin\mu\cos\mu+f^3\sin\mu)/F_{jc}^3\sim 0$$
$$\partial s/\partial F\sim\partial x_j/\partial F.$$

Here $\partial v/\partial F$ is taken as approximately zero because the expression, obtained by substitution from equation (30), is second degree in f/F. Given that $\partial v/\partial F\sim 0$, the approximation for $\partial s/\partial F$ follows. The Taylor-series approximations (equations 26) follow directly, with $w_m\sim w_{mo}$, since $w_m\equiv\Delta_{mo}+2k_mv_m$.

References

BIERENS DE HAAN, D. (1957). *Nouvelles Tables d'Intégrales Définies* (Edition of 1867, corrected), Table 28, no. 1 and Table 100, no. 7. New York: Hafner.

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

BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.

CRICK, F. H. C. & MAGDOFF, B. S. (1956). *Acta Cryst.* **9**, 901–908.

CULLIS, A. E., MUIRHEAD, H., PERUTZ, M. F. & ROSSMANN, M. G. (1961). *Proc. Roy. Soc.* A**265**, 15–38.

DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188–1195.

DICKERSON, R. E., WEINZIERL, J. E. & PALMER, R. A. (1968). *Acta Cryst.* B**24**, 997–1003.

KARTHA, G. (1965). *Acta Cryst.* **19**, 883–885.

LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint.* Cambridge Univ. Press.

MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 82–86.

NORTH, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.

RAIZ, V. SH. & ANDREEVA, N. S. (1970). *Sov. Phys. Crystallogr.* **15**, 206–210. Translated from *Kristallografiya* (1970). **15**, 246–251.