Acta Cryst. (1987). A43, 370-374

A Weighting Scheme for Tangent Formula Development. III. The Weighting Scheme of the *SIR* Program

BY M. C. BURLA

Dipartimento di Scienze della Terra - Sez. Cristallografia, Università, Piazza Università, 06100 Perugia, Italy

G. CASCARANO AND C. GIACOVAZZO

Dipartimento Geomineralogico, Università, via Salvemini, 70124 Bari, Italy

AND A. NUNZI AND G. POLIDORI

Dipartimento di Scienze della Terra - Sez. Cristallografia, Università, Piazza Università, 06100 Perugia, Italy

(*Received* 24 *February* 1986; *accepted* 17 *November* 1986)

Abstract

In paper I of this series [Giacovazzo (1979). *Acta Cryst.* A35, 757-764] a tangent formula was obtained which is able to take into account uncertainty of the 'known' phases. In paper II [Cascarano, Giacovazzo, Burla, Nunzi & Polidori (1984). *Acta Cryst.* A40, 389-394] the asymptotical distribution of the statistic $\alpha_{\mathbf{h}}$ was derived when the r component vectors $G_j \exp(i\theta_j)$ were distributed according to Von Mises distributions $M(\theta_j; \varphi_{\mathbf{h}}, G_j)$. In the present paper this result is extended to the case in which the vectors $G_j \exp(i\theta_j)$ are distributed according to a more general distribution $M(\theta_j; \varphi_{\mathbf{h}}, \beta_j)$. From the theoretical results a weighting scheme for the tangent procedure is derived which uses the first two moments of the $\alpha_{\mathbf{h}}$ distribution. The scheme has been implemented in the *SIR* program; applications to real structures are presented.

1. Symbols and notation

See papers I and II (Giacovazzo, 1979; Cascarano, Giacovazzo, Burla, Nunzi & Polidori, 1984).

2. Introduction

The conditional distribution of $\varphi_{\mathbf{h}}$, given $\varphi_{\mathbf{k}}$, $\varphi_{\mathbf{h}-\mathbf{k}}$ and $G = 2|E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}|/N^{1/2}$ is given by (Cochran, 1955)

$$P(\varphi_{\mathbf{h}}|\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}}, G) \simeq M(\varphi_{\mathbf{h}}; \theta_{\mathbf{h}}, G) = [2\pi I_0(G)]^{-1} \exp[G\cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})] \quad (1)$$

where $\theta_{\mathbf{h}} = \varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}$. When several pairs of phases $(\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}})$ are known (1) is replaced by

$$P(\varphi_{\mathbf{h}}|\{\varphi_{\mathbf{k}_j}, \varphi_{\mathbf{h}-\mathbf{k}_j}, G_j\}) \simeq M(\varphi_{\mathbf{h}}; \varphi_{\mathbf{h}}, \alpha_{\mathbf{h}}) \quad (2)$$

where $\theta_{\mathbf{h}}$, the most efficient value for $\varphi_{\mathbf{h}}$, is given by (Karle & Hauptman, 1956, Karle & Karle, 1966)

$$\tan\theta_{\mathbf{h}} = \frac{\sum_j |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})} = \frac{T_{\mathbf{h}}}{B_{\mathbf{h}}} \quad (3)$$

and the corresponding variance for $\varphi_{\mathbf{h}}$ is given by

$$\alpha_{\mathbf{h}} = 2N^{-1/2}|E_{\mathbf{h}}|(T_{\mathbf{h}}^2 + B_{\mathbf{h}}^2)^{1/2}. \quad (4)$$

In the practical direct procedures $\varphi_{\mathbf{k}_j}$ and $\varphi_{\mathbf{h}-\mathbf{k}_j}$ are themselves uncertain and are dispersed around their estimates $\theta_{\mathbf{k}_j}$ and $\theta_{\mathbf{h}-\mathbf{k}_j}$. So weighted tangent formulas such as (Germain, Main & Woolfson, 1971)

$$\tan\theta_{\mathbf{h}} = \frac{\sum w_{\mathbf{k}_j} w_{\mathbf{h}-\mathbf{k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \sin(\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j})}{\sum w_{\mathbf{k}_j} w_{\mathbf{h}-\mathbf{k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}| \cos(\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j})} = \frac{T'_{\mathbf{h}}}{B'_{\mathbf{h}}} \quad (5)$$

with

$$\alpha_{\mathbf{h}} = 2N^{-1/2}|E_{\mathbf{h}}|(T_{\mathbf{h}}'^2 + B_{\mathbf{h}}'^2)^{1/2} \quad (6)$$

can usefully replace (3) and (4). The weighting scheme afterwards adopted by Germain, Main & Woolfson (1971) was

$$w_{\mathbf{h}} = \min(0{\cdot}2\alpha_{\mathbf{h}}, 1{\cdot}0). \quad (7)$$

In paper I the reliability of the estimate $\theta_{\mathbf{h}}$ of $\varphi_{\mathbf{h}}$ was theroretically calculated even when $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h}-\mathbf{k}}$ were unknown provided their distributions $M(\varphi_{\mathbf{k}}; \theta_{\mathbf{k}}, \alpha_{\mathbf{k}})$ and $M(\varphi_{\mathbf{h}-\mathbf{k}}; \theta_{\mathbf{h}-\mathbf{k}}, \alpha_{\mathbf{h}-\mathbf{k}})$ were known. It was found that $\varphi_{\mathbf{h}}$ is distributed around $\theta_{\mathbf{h}}$ according to the Von Mises function $M(\varphi_{\mathbf{h}}; \theta_{\mathbf{h}}, \beta)$ where β is given by

$$D_1(\beta) = D_1(G) D_1(\alpha_{\mathbf{k}}) D_1(\alpha_{\mathbf{h}-\mathbf{k}}). \quad (8)$$

0108-7673/87/030370-05$01.50

If $\varphi_{\mathbf{h}}$ is estimated *via r* triplet relationships then $\varphi_{\mathbf{h}}$ is distributed according to $M(\varphi_{\mathbf{h}};\ \theta_{\mathbf{h}}, \alpha_{\mathbf{h}})$ where

$$\tan \theta_{\mathbf{h}} \simeq \sum_{j=1}^{r} \beta_j \sin (\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}) \Big/ \sum_{j=1}^{r} \beta_j \cos (\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}) \tag{9}$$

and

$$\alpha_{\mathbf{h}} = \left\{ \left[\sum_{j=1}^{r} \beta_j \cos (\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}) \right]^2 + \left[\sum_{j=1}^{r} \beta_j \sin (\theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}) \right]^2 \right\}^{1/2}. \tag{10}$$

Equations (9) and (10) can be written in a form which emphasizes weights by assuming $w_j = \beta_j / G_j$; however, the weight w_j is associated with the jth triplet while in (5) the individual phases $\theta_{\mathbf{k}_j}$ and $\theta_{\mathbf{h}-\mathbf{k}_j}$ preserve individual weights.

Unfortunately in some structures (Schenk, 1972) it happens that after tangent refinement three-phase invariants are closer to zero than their true values. Thus the enantiomorph is removed and pseudo-centrosymmetric solutions are found.

Hull & Irwin (1978) devised a successful scheme which attempts to match the $\alpha_{\mathbf{h}}$ calculated by (6) with that expected from the probability distribution of the triplet phase relationships. In practice such a procedure ignores the algebraic form of the distribution $P(\alpha_{\mathbf{h}})$ and uses only its first moment.

If available, the information about the second moment of $P(\alpha_{\mathbf{h}})$ should certainly be useful. The asymptotic (r sufficiently large) probability distribution of $\alpha_{\mathbf{h}}$ was calculated in paper II from this point of view. $\alpha_{\mathbf{h}}$ was considered as the modulus of the resultant of the complex vectors $G_j \exp(i\theta_j)$, $j = 1, 2, \ldots, r$, where $\theta_j = \varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$ are a large random sample of variables which are independently distributed according to the Von Mises functions $M(\theta_j;\ \varphi_{\mathbf{h}},\ G_j)$.

For non-centrosymmetric space groups it was found that

$$P(\alpha_{\mathbf{h}}) \simeq N(\alpha_{\mathbf{h}};\ \langle \alpha_{\mathbf{h}} \rangle,\ \sigma^2_{\alpha_{\mathbf{h}}}), \tag{11}$$

where N denotes the normal distribution and

$$\langle \alpha_{\mathbf{h}} \rangle = \sum_{j=1}^{r} G_j D_1(G_j), \tag{12}$$

$$\sigma^2_{\alpha_{\mathbf{h}}} = \tfrac{1}{2} \sum_{j=1}^{r} G_j^2 [1 + D_2(G_j) - 2D_1^2(G_j)]. \tag{13}$$

Equations (12) and (13) provide both the first and the second moments of $\alpha_{\mathbf{h}}$. However, they cannot be used for the tangent formula without modification in the weighting scheme because $\theta_j = \varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$, $j = 1, \ldots, r$, are unknown. Since only estimates $\theta_{\mathbf{k}_j}$ and $\theta_{\mathbf{h}-\mathbf{k}_j}$ of $\varphi_{\mathbf{k}_j}$ and $\varphi_{\mathbf{h}-\mathbf{k}_j}$ are available during the phasing procedure, a mathematical procedure may be devised aiming at recovering the first moments of $\alpha_{\mathbf{h}}$ in these conditions.

Such an approach is described in § 3 of this paper. In § 4 the weighting criteria and in § 5 the actual weighting procedure are described. Applications and conclusions are discussed in § 6.

3. The distributions of $\alpha_{\mathbf{h}}$ and $\theta_{\mathbf{h}}$ when the distributions of $\theta_{\mathbf{k}_j}$ and $\theta_{\mathbf{h}-\mathbf{k}_j}$ are known

In paper II the probability distributions $P(\alpha_{\mathbf{h}})$ and $P(\varphi_{\mathbf{h}})$ were calculated with the assumption that $\theta_j = \varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$, $j = 1, 2, \ldots, r$, were a random sample of variables distributed around $\varphi_{\mathbf{h}}$ according to $M(\theta_j;\ \varphi_{\mathbf{h}},\ G_j)$.

The problem reduces to that of calculating the distribution of the resultant of the complex vectors $G_j \exp(i\theta_j)$ under the hypothesis that θ_j is distributed according to $M(\theta_j;\ \varphi_{\mathbf{h}},\ G_j)$ [it may be noted that the concentration parameter of the function M is assumed to coincide with the modulus of the complex vector $G_j \exp(i\theta_j)$].

During the phase expansion or the refinement process such an assumption is not satisfied. Indeed, only the estimates $\theta_{\mathbf{k}_j}$ of $\varphi_{\mathbf{k}_j}$ and $\theta_{\mathbf{h}-\mathbf{k}_j}$ of $\varphi_{\mathbf{h}-\mathbf{k}_j}$ are known, which are supposed to be distributed around $\varphi_{\mathbf{k}_j}$ and $\varphi_{\mathbf{h}-\mathbf{k}_j}$ according to $M(\theta_{\mathbf{k}_j};\ \varphi_{\mathbf{k}_j}, \alpha_{\mathbf{k}_j})$ and $M(\theta_{\mathbf{h}-\mathbf{k}_j};\ \varphi_{\mathbf{h}-\mathbf{k}_j},\ \alpha_{\mathbf{h}-\mathbf{k}_j})$ respectively, with known values of $\alpha_{\mathbf{k}_j}$ and $\alpha_{\mathbf{h}-\mathbf{k}_j}$.

In these conditions we have to calculate the probability distributions $P(\alpha_{\mathbf{h}})$ and $P(\varphi_{\mathbf{h}})$ on assuming, in accordance with (8), that $\theta_j = \theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}$ is distributed around $\varphi_{\mathbf{h}}$ according to $M(\theta_j;\ \varphi_{\mathbf{h}},\ \beta_j)$. From a mathematical point of view the problem reduces to that of calculating the distribution of the resultant of the complex vectors $G_j \exp(i\theta_j)$ under the hypothesis that θ_j is distributed according to $M(\theta_j;\ \varphi_{\mathbf{h}},\ \beta_j)$. In accordance with the above remarks the approach described in §§ 4 and 5 of paper II may readily be modified to obtain the wanted probability distributions. For example, for C and S defined by equation (19) of paper II we obtain [to be compared with equation (20a) of paper II]

$$\langle C \rangle = (1/\textstyle\sum_1) \displaystyle\sum_{j=1}^{r} G_j D_1(\beta_j) \cos \varphi_{\mathbf{h}},$$

$$\langle S \rangle = (1/\textstyle\sum_1) \displaystyle\sum_{j=1}^{r} G_j D_1(\beta_j) \sin \varphi_{\mathbf{h}}.$$

After some calculations the resultant formulas for the non-centrosymmetric case are obtained:

$$P(\varphi_{\mathbf{h}}) \simeq M(\varphi_{\mathbf{h}};\ \theta_{\mathbf{h}}, \langle \alpha_{\mathbf{h}} \rangle), \tag{14}$$

where $\theta_{\mathbf{h}}$ is given by (9) and $\langle \alpha_{\mathbf{h}} \rangle$ by

$$\langle \alpha_{\mathbf{h}} \rangle = \sum_{j=1}^{r} G_j D_1(\beta_j). \tag{15}$$

In its turn

$$P(\alpha_{\mathbf{h}}) = N(\alpha_{\mathbf{h}};\ \langle \alpha_{\mathbf{h}} \rangle,\ \sigma^2_{\alpha_{\mathbf{h}}}), \tag{16}$$

where

$$\sigma^2_{\alpha_h} = \tfrac{1}{2} \sum_{j=1}^{r} G_j^2 [1 + D_2(\beta_j) - 2D_1^2(\beta_j)]. \qquad (17)$$

Equations (15) and (17) can usefully be compared with equations (34) and (35) of paper II respectively; in paper II, G_j was both the modulus of the jth component vector and the concentration parameter of its distribution, while in (17), G_j is still the modulus of the jth vector but the concentration parameter of its distribution coincides with β_j.

4. A weighting criterion for the tangent formula in non-centrosymmetric space groups

$P(\alpha_h)$ and $P(\varphi_h)$ rely on the same probabilistic assumptions. Therefore, if the experimental value of α_h for large values of r differs significantly from $\langle \alpha_h \rangle$, then it may be guessed that the θ_j's are not distributed around φ_h according to the theoretical assumptions. Consequently, θ_h appears as a biased estimate of φ_h. Conversely, if $\alpha_h \simeq \langle \alpha_h \rangle$ the phase relationship

$$\theta_h \simeq \varphi_h \qquad (18)$$

is experimentally supported. Such an argument is in practice the statistical basis of Hull & Irwin's (1978) weighting scheme.

Our results in § 3 suggested to us the following weight, which exploits the first two moments of the distribution $P(\alpha_h)$:

$$w_h = [P(\alpha_h)/P(\alpha_h = \langle \alpha_h \rangle)]^n$$
$$= \{\exp[-(\alpha_h - \langle \alpha_h \rangle)^2/2\sigma^2_{\alpha_h}]\}^n, \qquad (19)$$

where n is a rational positive number. Such a scheme has the following properties:

(a)

$$0 \le w_h \le 1;$$

(b)

$$w_h = 1 \quad \text{if} \quad \alpha_h = \langle \alpha_h \rangle, \quad w_h < 1 \text{ otherwise};$$

(c) the observed discrepancy $\Delta_h = \alpha_h - \langle \alpha_h \rangle$ has a different statistical meaning according to the distribution $P(\alpha_h)$. Indeed, if the variance of α_h is small, an experimental large Δ_h may strongly reduce our confidence in θ_h, while the same Δ_h coupled with a large variance in practice preserves it. This agrees well with common sense.

Every weighting scheme allots small weights to phases with small α_h values. Thus the main aim of the above theory is to reduce our confidence in θ_h when $\alpha_h > \langle \alpha_h \rangle$. Accordingly, if $\alpha_h < \langle \alpha_h \rangle$, equations (9) and (10) are used without modifications; if $\alpha_h > \langle \alpha_h \rangle$ then $\alpha_h = \langle \alpha_h \rangle w_h$.

Even if equations (14)-(17) are an indisputable outcome of the theory their use requires some observations:

Table 1. *The weight w in specific cases*

For convenience, in parentheses the values ($\langle \alpha_h \rangle$; $\sigma^2 \alpha_h$) are given.

	$r = 5$	$r = 10$	$r = 20$
$\beta = 0{\cdot}4$, $\Delta_h = 0{\cdot}15$	$w = 0{\cdot}905$ (0·39; 0·38)	$w = 0{\cdot}950$ (0·78; 0·75)	$w = 0{\cdot}970$ (0·16; 1·51)
$\beta = 0{\cdot}4$, $\Delta_h = 0{\cdot}30$	$w = 0{\cdot}672$ (0·39; 0·38)	$w = 0{\cdot}820$ (0·78; 0·75)	$w = 0{\cdot}905$ (0·16; 1·51)
$\beta = 2{\cdot}0$, $\Delta_h = 0{\cdot}30$	$w = 0{\cdot}995$ (6·97; 3·29)	$w = 0{\cdot}998$ (13·95; 6·57)	$w = 0{\cdot}999$ (27·91; 13·14)
$\beta = 2{\cdot}0$, $\Delta_h = 1{\cdot}00$	$w = 0{\cdot}951$ (6·98; 3·29)	$w = 0{\cdot}975$ (13·95; 6·57)	$w = 0{\cdot}987$ (27·91; 13·14)
$\beta = 2{\cdot}0$, $\Delta_h = 3{\cdot}00$	$w = 0{\cdot}633$ (6·98; 3·29)	$w = 0{\cdot}796$ (13·95; 6·57)	$w = 0{\cdot}892$ (27·91; 13·14)

(a) The equations have asymptotical nature. In other words, r is required to be sufficiently large, while in the usual tangent procedures r varies from 1 to (roughly) 30. We calculate w_h only if $r > 4$.

(b) The theory can be successfully applied to finite samples [the available set of the vectors $G_j \exp(i\theta_j)$] provided they are random samples of the entire population. That does not usually occur in practice for various reasons: small sample sizes, structural regularities violating the basic assumptions of the probabilistic method, systematic errors in the estimations of the G's, *etc.*

Such troubles cannot be avoided but they can certainly be reduced by a proper choice of n in (19). Too small values of n would be ineffective, large values would reduce the reliability of the phases having $\alpha_h > \langle \alpha_h \rangle$ too much, so that only phases having $\alpha_h < \langle \alpha_h \rangle$ will pilot the phase-extension procedure, contrary to common sense. We found that $n = \frac{1}{3}$ is a sensible choice.

In order to give some numerical examples, let us suppose that in (17) $\beta_j = \beta$ for $j = 1, 2, \ldots, r$ and $r = 5$, 10, 20. In Table 1 values of $w_h = w$ are given for specific values of Δ_h and β. It may be noted that:

(a) for fixed β and r, w_h decreases if Δ_h increases, according to expectations (compare rows 1 and 2 of the table);

(b) for fixed β and Δ_h, w_h increases with r (our confidence in θ_h has to be larger when Δ_h is observed for large samples, since larger values of $\langle \alpha_h \rangle$ and of $\sigma^2_{\alpha_h}$ arise);

(c) for the same reason, for fixed r and Δ_h, w_h increases with β (compare rows 2 and 3).

5. The weighting scheme

The theory described above has been implemented in the *SIR* program (Cascarano, Giacovazzo, Burla, Nunzi, Polidori, Camalli, Spagna & Viterbo, 1985) according to the following scheme.

5(a) *The weighted convergence method*

The convergence method as described by Germain, Main & Woolfson (1971) calculates α_h as given by

(12) for each reflexion **h** used in the $\sum_2$ list. The reflexion with the lowest $\langle\alpha_{\mathbf{h}}\rangle$ is eliminated together with all the phase relationships in which the reflexion is involved. Then the values $\langle\alpha_{\mathbf{h}}\rangle$ are recalculated again and the reflexion with the smallest $\langle\alpha_{\mathbf{h}}\rangle$ is eliminated together with the phase relationship in which it is involved. The process thus converges on that set of reflexions which are linked together best of all in the probabilistic sense.

The above procedure has been modified by remembering that in accordance with (9) and (10) the concentration parameter $\alpha_{\mathbf{h}}$ of the relation $\theta_{\mathbf{h}} \simeq \varphi_{\mathbf{h}}$ depends on the concentration parameters $\alpha_{\mathbf{k}_j}$ and $\alpha_{\mathbf{h}-\mathbf{k}_j}$, and on $G_j = 2|E_{\mathbf{h}}E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}|/N^{1/2}, j = 1, 2, \ldots r$. Thus, in the absence of phase information, when only the statistic $\langle\alpha_{\mathbf{h}}\rangle$ may be estimated, (12) may be replaced by a formula which takes into account the expected uncertainty of the phases $\theta_j = \theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}$ on which $\theta_{\mathbf{h}}$ depends. If $\theta_{\mathbf{k}_j}$ and $\theta_{\mathbf{h}-\mathbf{k}_j}$ were really dispersed according to concentration parameters $\langle\alpha_{\mathbf{k}_j}\rangle$ and $\langle\alpha_{\mathbf{h}-\mathbf{k}_j}\rangle$ respectively, then

$$\langle\alpha_{\mathbf{h}}\rangle = \sum_{j=1}^{r} G_j D_1(G_j) D_1(\langle\alpha_{\mathbf{k}_j}\rangle) D_1(\langle\alpha_{\mathbf{h}-\mathbf{k}_j}\rangle). \quad (20)$$

Actually $\alpha_{\mathbf{k}_j}$ and $\alpha_{\mathbf{h}-\mathbf{k}_j}$ do not coincide with $\langle\alpha_{\mathbf{k}_j}\rangle$ and $\langle\alpha_{\mathbf{h}-\mathbf{k}_j}\rangle$, but are dispersed around these values. Thus $\langle\alpha_{\mathbf{h}}\rangle$ is overestimated by (20): in the practical procedure we empirically use $\langle\alpha_{\mathbf{h}}\rangle/2$ instead of $\langle\alpha_{\mathbf{h}}\rangle$. After that, the convergence method described by Germain, Main & Woolfson works without further modifications.

It is worth specifying that:

(*a*) Equations (12) and (20) have different statistical meanings. Equation (12) is the expected value of $\alpha_{\mathbf{h}}$ relative to $\varphi_{\mathbf{h}}$ when the phases $\theta_j = \varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$ are considered. Equation (20) is the expected value of $\alpha_{\mathbf{h}}$ relative to $\theta_{\mathbf{h}}$ when the average is taken over independent collections of r elements $\theta_j = \theta_{\mathbf{k}_j} + \theta_{\mathbf{h}-\mathbf{k}_j}$, each element dispersed around $\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}$ according to the concentration parameters $\langle\alpha_{\mathbf{k}_j}\rangle$ and $\langle\alpha_{\mathbf{h}-\mathbf{k}_j}\rangle$.

(*b*) The classical and our convergence procedures usually lead to different starting sets.

5(*b*) *Weighting of the starting set*

The starting set of phases usually contains: (*a*) origin- and enantiomorph-defining phases; (*b*) phases determined by probabilistic formulas (*i.e.* one-phase seminvariants); (*c*) symbolic phases which may be represented by quadrant permutation of magic-integer methods (Main, 1977).

All the phases $\theta_{\mathbf{h}}$ entering the starting set are considered to be distributed around the true values $\varphi_{\mathbf{h}}$ according to the Von Mises distributions $M(\theta_{\mathbf{h}}; \varphi_{\mathbf{h}}, \alpha_{\mathbf{h}})$ (see paper I, § 8).

Phases in category (*a*), which are known without uncertainty, are assumed to be distributed according to $M(\varphi_{\mathbf{h}}; \theta_{\mathbf{h}}, 100)$ which approximates the Dirac function $\delta(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})$.

Table 2. $\alpha_{\mathbf{h}}$ *values associated with starting-set phases represented by magic-integer sequences*

n	Sequence	No. of sets	R.m.s. error (°)	$\alpha_{\mathbf{h}}$
1	1	4	26	5·5
2	2 3	12	29	4·7
3	3 4 5	20	37	3·2
4	5 7 8 9	32	42	2·6
5	8 11 13 14 15	50	45	2·4
6	13 18 21 23 24 25	80	47	2·2
7	21 29 34 37 39 40 41	128	48	2·1
8	34 47 55 60 63 65 66 67	206	49	2·0

The Von Mises parameters for phases in category (*b*) are provided by the subroutine estimating one-phase structure seminvariants *via* the second representation (Cascarano, Giacovazzo, Calabrese, Burla, Nunzi, Polidori & Viterbo, 1984).

Phases in category (*c*) represented by quadrant permutation (including enantiomorph-defining phases whose values are unrestricted by symmetry) are considered to be distributed (see paper I) according to $M(\theta_{\mathbf{h}}; \varphi_{\mathbf{h}}, 5{\cdot}5)$, where $\theta_{\mathbf{h}}$ is the median value in the quadrant. Phases in category (*c*) whose values are restricted by symmetry are distributed according to $M(\varphi_{\mathbf{h}}; \theta_{\mathbf{h}}, 100)$ because all the symmetry-allowed phases are assigned by phase permutation.

Symbolic phases represented by a magic-integer sequence are assumed to be distributed according to $M(\theta_{\mathbf{h}}; \varphi_{\mathbf{h}}, \alpha_{\mathbf{h}})$ where $\alpha_{\mathbf{h}}$ is the concentration parameter of the Von Mises distribution having the square root of the variance equal to the root-mean-square error in the phases expected for that magic-integer sequence. From Table 2 the largest $\alpha_{\mathbf{h}}$ are associated with the smallest integer sequences (with which the smallest expected phase errors are associated).

5(*c*) *The weighted divergence procedure*

The divergence path is fixed by (15), where β_j is calculated according to (8). When $\alpha_{\mathbf{h}}$ is not available, $\langle\alpha_{\mathbf{h}}\rangle$ is used.

5(*d*) *The weighted tangent procedure*

Expressions (9) and (10) are used: if $\alpha_{\mathbf{h}} > \langle\alpha_{\mathbf{h}}\rangle$ then the concentration parameter $\langle\alpha_{\mathbf{h}}\rangle w_{\mathbf{h}}$ is associated with $\theta_{\mathbf{h}}$, where $w_{\mathbf{h}}$ is given by (19) for $n = \frac{1}{3}$.

6. Applications

The *SIR* program has been applied to seven crystal structures, the main data of which are described in Table 3. Corresponding to the different options offered by *SIR* the following procedures have been used:

(*a*) A weighted tangent formula according to (5) and (7) (this is the classical *MULTAN* weighting scheme).

(*b*) A weighted tangent formula according to the new criteria described in § 5.

Table 3. *Relevant data for seven crystal structures*

Data are given in the order: code title (Code), chemical formula (CF), space group (SG), number of chemical formulas in the unit cell (Z), number of nonhydrogen atoms in the unit cell (N). In the last three columns crosses mark successful applications by the *SIR* procedure in options (c) and (d), and by the Hull & Irwin (1978) method (HI).

Code	CF	SG	Z	N	*SIR*(c)	*SIR*(d)	HI
RIFOL†	$C_{39}H_{49}NO_{13}$	$P2_1$	2	106		+	+
PROLIN‡	$C_{26}H_{40}N_4O_7$	$P2_1$	2	74	+	+	
ERGO*	$C_{28}H_{44}O$	$P2_12_12_1$	8	232			
MUNICH1*	$C_{20}H_{16}$	$C2$	8	160			
CEPHAL*	$C_{18}H_{21}O_3N$	$C2$	8	176			
DIOLE*	$C_{10}H_{18}O_2$	$I\bar{4}2d$	16	192	+	+	
APAPA*	$C_{30}H_{37}N_{15}O_{16}P_2.6H_2O$	$P4_12_12$	8	552	+	+	+

* Experimental data and references can be found on the magnetic tape distributed by the York crystallographic group, to which the reader is referred.
† Burla, Cerrini, Lamba, Nunzi & Polidori (1987).
‡ Colapietro, De Santis, Palleschi & Spagna (1987).

(c) First, the reliabilities of the triplets are estimated *via* the second representation formula $P10$ (Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori, 1984). Then, instead of G, the new concentration parameter G' provided by the $P10$ formula is used in (5) and (7).

(d) The concentration parameter G' provided by the $P10$ formula is associated with the new weighting scheme described in § 5.

For every option a default procedure with five magic-integer symbols in the starting set has been used.

Options (c) and (d) proved very much more effective than options (a) and (b). That is not surprising since the $P10$ formula for estimating a single triplet exploits a subset of diffraction intensities much larger than the classical Cochran formula.

The outcome for options (c) and (d) is shown in Table 3: for completeness, in the last column of Table 3 data obtained by application of the Hull & Irwin (1978) weighting scheme are reported (also *via* a default procedure with five symbols in the starting set). Crosses in columns 6, 7 and 8 mark successful trials (*i.e.* the correct solution has been found among the three sets with the largest figures of merit).

From Table 3 it cannot be deduced that the Hull & Irwin (1978) scheme or *SIR* are unable to solve the unlabelled crystal structures, but only that the correct solution has not been found, after a default process, among the three sets marked by the largest figures of merit.

For example, the correct solution for CEPHAL is found by *SIR* in option (d) if two symbols with restricted phase and three magic-integer symbols of general type are used in the starting set instead of five symbols with unrestricted phase (correspondingly, 36 sets per trial instead of 24).

Furthermore, ERGO is solved by *SIR* in option (d) if seven general symbols represented by magic integers are introduced into the starting set. We were not able to find the correct solutions for ERGO and CEPHAL by application of *SIR* in option (c).

Table 3 suggests that *SIR* is a powerful method for solving crystal structures and is an alternative to other packages of proven usefulness.

References

BURLA, M. C., CERRINI, S., LAMBA, D., NUNZI, A. & POLIDORI, G. (1987). *Acta Cryst.* Submitted.

CASCARANO, G., GIACOVAZZO, C., BURLA, M. C., NUNZI, A. & POLIDORI, G. (1984). *Acta Cryst.* A**40**, 389-394.

CASCARANO, G., GIACOVAZZO, C., BURLA, M. C., NUNZI, A., POLIDORI, G., CAMALLI, M., SPAGNA, R. & VITERBO, D. (1985). Proc. Ninth Eur. Crystallogr. Meet., Torino, Abstr. 1-046.

CASCARANO, G., GIACOVAZZO, C., CALABRESE, G., BURLA, M. C., NUNZI, A., POLIDORI, G. & VITERBO, D. (1984). *Z. Kristallogr.* **167**, 37-47.

CASCARANO, G., GIACOVAZZO, C., CAMALLI, M., SPAGNA, R., BURLA, M. C., NUNZI, A. & POLIDORI, G. (1984). *Acta Cryst.* A**40**, 278-283.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.

COLAPIETRO, M., DE SANTIS, P., PALLESCHI, A. & SPAGNA, R. (1987). *Biopolymers.* In the press.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A**27**, 368-376.

GIACOVAZZO, C. (1979). *Acta Cryst.* A**35**, 757-764.

HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* A**34**, 863-870.

KARLE, I. & KARLE, J. (1966). *Acta Cryst.* **21**, 849-859.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635-651.

MAIN, P. (1977). *Acta Cryst.* A**33**, 750-757.

SCHENK, H. (1972). *Acta Cryst.* A**28**, 412-422.