CICCARIELLO, S., GOODISMAN, J. & BRUMBERGER, H. (1988). *J. Appl. Cryst.* **21**, 117-128.
DEBYE, P., ANDERSON, H. R. & BRUMBERGER, H. (1957). *J. Appl. Phys.* **28**, 679-683.
ERDÉLYI, A. (1956). *Asymptotic Expansions*, ch. II. New York: Dover.
GILLE, W. (1987). *Exp. Tech. Phys.* **35**, 93-98.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1970). *Tables of Integrals, Series and Products*, ch. 8.1. London: Academic Press.
GUINIER, A. (1963). *X-ray Diffraction*, chs. 2 and 10. San Francisco: Freeman.
HILBERT, D. & COHN-VOSSEN, S. (1953). *Geometry and the Imagination.* New York: Chelsea.
JONES, D. S. & KLINE, M. (1958). *J. Math. Phys. (Cambridge, Mass.)*, **37**, 1-28.
MILLER, A. & SCHMIDT, P. W. (1962). *J. Math. Phys. (NY)*, **3**, 92-96.
POROD, G. (1952). *Kolloid Z.* **125**, 51-60.
POROD, G. (1982). *Small-Angle X-ray Scattering*, edited by O. GLATTER & O. KRATKY, ch. I. London: Academic Press.
ROJANSKI, D., HUPPERT, D., BALE, H. D., DACAI, X., SCHMIDT, P. W., FARIN, D., SERI-LEVY, A. & AVNIR, D. (1986). *Phys. Rev. Lett.* **23**, 2505-2508.
SCHMIDT, P. W. (1965). *J. Math. Phys. (NY)*, **6**, 424-431.
SCHMIDT, P. W. (1967). *J. Math. Phys. (NY)*, **8**, 475-477.
SMIRNOV, V. I. (1964). *A Course of Higher Mathematics*, Vol. II, ch. 7. Oxford: Pergamon.
WILLIAMS, J. A., RINDONE, G. E. & MCKINSTRY, H. A. (1981). *J. Am. Ceram. Soc.* **64**, 702-709.
WU, H. & SCHMIDT, P. W. (1974). *J. Appl. Cryst.* **7**, 131-146.

Acta Cryst. (1989). A**45**, 99-104

A Reconsideration of the Role of Two-Phase Seminvariants. V. Basic Results

BY M. C. BURLA

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

C. GIACOVAZZO

Dipartimento Geomineralogico, Università, Campus Universitario, 70124 Bari, Italy

AND G. POLIDORI

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

(*Received* 30 *November* 1987; *accepted* 22 *July* 1988)

Abstract

A basic obstacle to the widespread use of two-phase seminvariants of first rank in direct methods is often the large amount of computing time needed for their probabilistic estimation. A new very fast algorithm for identifying such seminvariants and a modified probabilistic formula for their estimation are described. Contrary to common belief, practical tests show that the amount of information contained in two-phase seminvariants is in general not negligible compared with information provided by triplets.

Symbols

N: number of atoms in the cell.
m: number of symmetry operators of the space group.
$\mathbf{C}_s \equiv (\mathbf{R}_s, \mathbf{T}_s)$: sth symmetry operator: $\mathbf{R}_s$ is its rotational and $\mathbf{T}_s$ its translational part.
$E_{\mathbf{h}}$: normalized structure factor with vectorial index $\mathbf{h}$.
$R_{\mathbf{h}} \equiv |E_{\mathbf{h}}|$.
$\varphi_{\mathbf{h}}$: phase of $E_{\mathbf{h}}$.
$\mathbf{I}$: identity 3×3 matrix.

1. Introduction

A first attempt at evaluating two-phase seminvariants was described by Grant, Howells & Rogers (1957). The method (the so-called 'coincidence method') was extended to non-centrosymmetric space groups by Debaerdemaeker & Woolfson (1972), according to the following argument. Let

$$\mathbf{u}_1 = \mathbf{h}_1 - \mathbf{h}_2\mathbf{R}_\beta, \tag{1a}$$

$$\mathbf{u}_2 = \mathbf{h}_2 - \mathbf{h}_1\mathbf{R}_\alpha. \tag{1b}$$

If $|E_{\mathbf{u}_1}|, |E_{\mathbf{u}_2}|, |E_{\mathbf{h}_1}|, |E_{\mathbf{h}_2}|$ are all sufficiently large then

$$\varphi_{\mathbf{u}_1} \simeq \varphi_{\mathbf{h}_1} - \varphi_{\mathbf{h}_2\mathbf{R}_\beta},$$

$$\varphi_{\mathbf{u}_2} \simeq \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1\mathbf{R}_\alpha},$$

so that

$$\Phi_2 = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} \simeq 2\pi(\mathbf{h}_1\mathbf{T}_\alpha + \mathbf{h}_2\mathbf{T}_\beta). \tag{2}$$

Φ_2 is a structure seminvariant and may therefore be estimated by means of (2).

0108-7673/89/010099-06$03.00

The main limit of the above approach is that it cannot provide information on Φ_2 in contradiction with that provided by triplet invariants. For example, in symmorphic space groups the estimate $\Phi_2 \simeq \pi$ is not allowed. Such a result is on the contrary possible if the method of joint probability distribution functions is applied (Giacovazzo 1977*a*, *b*; from now on referred to as papers I and II respectively): in particular Φ_2 was estimated *via* the conditional probability

$$P(\Phi_2 \| E_{\mathbf{h}_1}|, |E_{\mathbf{h}_2}|, |E_{\mathbf{u}_1}|, |E_{\mathbf{u}_2}|).$$

Independently Green & Hauptman (1976, 1978*a*, *b*) and Hauptman & Green (1978), using the neighbourhood concept, derived conditional probability distributions for two-phase seminvariants in space groups $P\bar{1}$ and $P2_1$. The approach was able to exploit larger sets of diffraction magnitudes than before.

A more general point of view was introduced by Giacovazzo (1978), Giacovazzo (1979) (from now on referred to as paper III) and Giacovazzo, Spagna, Vicković & Viterbo (1979) (from now on referred to as paper IV) who applied the representation method for ranking diffraction magnitudes in order of expected effectiveness (in the statistical sense) for the estimation of Φ_2. The algebraic results may be summarized as follows:

(*a*) $\Phi_2 = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$ is a two-phase structure seminvariant of the first rank if at least two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$ and two rotation matrices $\mathbf{R}_\alpha$ and $\mathbf{R}_\beta$ exist such that (1) are satisfied.

(*b*) Solutions $\mathbf{h}_1$ and $\mathbf{h}_2$ of the system (1), for given $\mathbf{u}_1, \mathbf{u}_2, \mathbf{R}_\alpha, \mathbf{R}_\beta$, may be found by first replacing the system (1) by

$$\mathbf{h}_1(\mathbf{I} - \mathbf{R}_\alpha \mathbf{R}_\beta) = \mathbf{u}_1 + \mathbf{u}_2 \mathbf{R}_\beta, \tag{3a}$$

$$\mathbf{h}_2(\mathbf{I} - \mathbf{R}_\beta \mathbf{R}_\alpha) = \mathbf{u}_1 \mathbf{R}_\alpha + \mathbf{u}_2, \tag{3b}$$

and then solving (3).

In order explicitly to obtain $\mathbf{h}_1$ and $\mathbf{h}_2$ the generalized inverses (say $\mathbf{A}^*_{\alpha,\beta}$ and $\mathbf{A}^*_{\beta,\alpha}$) of

$$\mathbf{A}_{\alpha,\beta} = (\mathbf{I} - \mathbf{R}_\alpha \mathbf{R}_\beta) \quad \mathbf{A}_{\beta,\alpha} = (\mathbf{I} - \mathbf{R}_\beta \mathbf{R}_\alpha)$$

respectively have to be calculated, so that

$$\mathbf{h}_1 = (\mathbf{u}_1 + \mathbf{u}_2 \mathbf{R}_\beta) \mathbf{A}^*_{\alpha,\beta} + \mathbf{Z}(\mathbf{I} - \mathbf{A}_{\alpha,\beta} \mathbf{A}^*_{\alpha,\beta}), \tag{4a}$$

$$\mathbf{h}_2 = (\mathbf{u}_1 \mathbf{R}_\alpha + \mathbf{u}_2) \mathbf{A}^*_{\beta,\alpha} + \mathbf{Z}(\mathbf{I} - \mathbf{A}_{\beta,\alpha} \mathbf{A}^*_{\beta,\alpha}). \tag{4b}$$

$\mathbf{Z}$ is a free vector with integer components: consequently $\mathbf{h}_1$ and $\mathbf{h}_2$ represent in general sets of vectors (say $\{\mathbf{h}_1\}$ and $\{\mathbf{h}_2\}$ respectively) rather than single vectors. Equations (4*a*) and (4*b*) have useful integer solutions if systems

$$\left.\begin{aligned} &(\mathbf{u}_1 + \mathbf{u}_2 \mathbf{R}_\beta) \mathbf{A}^*_{\alpha,\beta} \equiv 0 \pmod 1, \\ &(\mathbf{u}_1 \mathbf{R}_\alpha + \mathbf{u}_2) \mathbf{A}^*_{\beta,\alpha} \equiv 0 \pmod 1 \end{aligned}\right. \tag{5}$$

are satisfied respectively.

(*c*) When $\{\mathbf{h}_1\}$ and $\{\mathbf{h}_2\}$ are obtained, the special quartets

$$\varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2 \mathbf{R}_\beta} - \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_1 \mathbf{R}_\alpha \mathbf{R}_\beta}, \tag{6a}$$

$$\varphi_{\mathbf{u}_1 \mathbf{R}_\alpha} + \varphi_{\mathbf{u}_2} - \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_2 \mathbf{R}_\beta \mathbf{R}_\alpha} \tag{6b}$$

can be constructed. The quartet (6*a*) depends on the cross magnitudes with indices

$$\begin{aligned} \mathbf{u}_1 + \mathbf{u}_2 \mathbf{R}_\beta &\equiv \mathbf{h}_1(\mathbf{I} - \mathbf{R}_\alpha \mathbf{R}_\beta), \\ \mathbf{u}_1 - \mathbf{h}_1 &\equiv -\mathbf{h}_2 \mathbf{R}_\beta, \\ \mathbf{u}_2 \mathbf{R}_\beta - \mathbf{h}_1 &\equiv -\mathbf{h}_1(\mathbf{I} + \mathbf{R}_\alpha \mathbf{R}_\beta) + \mathbf{h}_2 \mathbf{R}_\beta; \end{aligned}$$

likewise the quartet (6*b*) depends on the cross magnitudes with indices

$$\begin{aligned} \mathbf{u}_1 \mathbf{R}_\alpha + \mathbf{u}_2 &\equiv \mathbf{h}_2(\mathbf{I} - \mathbf{R}_\beta \mathbf{R}_\alpha), \\ \mathbf{u}_1 \mathbf{R}_\alpha - \mathbf{h}_2 &\equiv \mathbf{h}_1 \mathbf{R}_\alpha - \mathbf{h}_2(\mathbf{I} + \mathbf{R}_\beta \mathbf{R}_\alpha), \\ \mathbf{u}_2 - \mathbf{h}_2 &\equiv -\mathbf{h}_1 \mathbf{R}_\alpha. \end{aligned}$$

Each quartet (6) differs from Φ_2 by a constant arising because of translational symmetry.

(*d*) Some of the magnitudes on which quartets (6*a*) depend influence also the estimates of the quartets (6*b*). Thus the probabilistic estimation of Φ_2 *via* the first representation is obtained by calculating the conditional distribution

$$\begin{aligned} P(\Phi_2 \| &E_{\mathbf{u}_1}|, |E_{\mathbf{u}_2}|, \{|E_{\mathbf{h}_1}|\}, \{|E_{\mathbf{h}_2}|\}, |E_{\mathbf{h}_1(\mathbf{I} - \mathbf{R}_\alpha \mathbf{R}_\beta)}|, \\ &|E_{\mathbf{h}_2(\mathbf{I} - \mathbf{R}_\beta \mathbf{R}_\alpha)}|, \{|E_{-\mathbf{h}_1(\mathbf{I} + \mathbf{R}_\alpha \mathbf{R}_\beta) + \mathbf{h}_2 \mathbf{R}_\beta}|\}, \\ &\{|E_{\mathbf{h}_1 \mathbf{R}_\alpha - \mathbf{h}_2(\mathbf{I} + \mathbf{R}_\beta \mathbf{R}_\alpha)}|\}). \end{aligned} \tag{7}$$

If more pairs $(\mathbf{R}_\alpha, \mathbf{R}_\beta)$ satisfy (1) then (7) may be easily generalized. The above procedure is quite general and may be applied to any space group: however the only program so far available [the *SIR* program (Cascarano *et al.*, 1985)] works only up to the orthorhombic system (see paper IV). In addition the procedure is rather time consuming: indeed for each pair of reflexions $(\mathbf{u}_1, \mathbf{u}_2)$ chosen among the NRIF (say NRIF ≤ 500) largest reflexions, all possible pairs of matrices $(\mathbf{R}_\alpha, \mathbf{R}_\beta)$ have to be introduced in (7) in order to check if integer solutions $\{\mathbf{h}_1\}$ and $\{\mathbf{h}_2\}$ are allowed. Even if the procedure may be sped up by storing all $m(m-1)$ generalized inverses $\mathbf{A}^*_{\alpha,\beta}$, solving (4) and checking (5) up to

$$\mathrm{NRIF}(\mathrm{NRIF} - 1)m(m-1)/4$$

times may be a big job even for fast computers, particularly for high-symmetry space groups (incidentally, just those space groups provide the largest information on structure seminvariants). The computing time is almost doubled in non-centrosymmetric space groups because we have to check which of $\varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$ and/or $\varphi_{\mathbf{u}_1} - \varphi_{\mathbf{u}_2}$ are seminvariants of first rank.

In order to reduce computing time, NRIF may be reduced (in the *SIR* program NRIF ≤ 50), but not too much because two-phase seminvariants are phase relationships of order $1/N$ so that only a small percentage of them can be reliably estimated. The low average reliability level of the probabilistic estimates is also due to the fact that quartets (6) are of special type, and they are on average less reliable than general quartets. On the other hand the use of two-phase seminvariants in direct procedures is encouraged because several related special quartets are often available to estimate single two-phase seminvariants, particularly in high-symmetry space groups.

It may be concluded that the most important obstacle to the wider use of two-phase seminvariants in direct procedures is their computing cost. To overcome this drawback a new algorithm is described in § 2 and in Appendix *A* (deposited) which drastically reduces computing time. The probabilistic background and the conclusive formulae estimating seminvariants are described in § 3. Some practical applications are described in § 4.

2. A fast algorithm for finding two-phase structure seminvariants of the first rank

A new algorithm singling out two-phase seminvariants of first rank is here described and has been implemented into the new release of the *SIR* program.

Usually the first phase relationships estimated by *SIR* are the one-phase structure seminvariants of first rank $\varphi_{\mathbf{H}}$. They are characterized by the following property: at least one rotation matrix $\mathbf{R}_n$ and one vector $\mathbf{h}_1$ may be found such that

$$\mathbf{H}=\mathbf{h}_1(\mathbf{I}-\mathbf{R}_n). \tag{8}$$

Once reflexions $\mathbf{H}$ have been identified by the program (say the *SEM*1 subroutine) the pairs $(\mathbf{R}_n, \{\mathbf{h}_1\})$ which satisfy (8) are stored for every $\mathbf{H}$. This information, which is essential for the estimation of $\varphi_{\mathbf{H}}$ *via* its first and/or second representation (Cascarano, Giacovazzo, Calabrese, Burla, Nunzi, Polidori & Viterbo, 1984), is later supplied by *SEM*1 to the program which estimates two-phase structure seminvariants of first rank (say *SEM*2).

*SEM*2 works according to the following steps. A $\sum_2$ list is calculated relative only to triplets containing one-phase structure seminvariants of first rank. A typical row of the list looks like

$$n_i, +n_j-n_k, \Delta_3; \quad -n_s+n_t, \Delta_3; \ldots \tag{9}$$

where n_i is the code number of $\mathbf{H}$, (n_j, n_k), $(n_s, n_t), \ldots$, are code numbers of usually general reflexions, and Δ_3 is the phase shift of the triplet.

Let us rewrite the generical triplet in (8) in terms of indices:

$$\mathbf{H}=\mathbf{h}_1(\mathbf{I}-\mathbf{R}_n)=s_1\mathbf{u}_{a_1}\mathbf{R}_p+s_2\mathbf{u}_{a_2}\mathbf{R}_q \tag{10}$$

where $(\mathbf{u}_{a_1}, \mathbf{u}_{a_2})$ are the standard indices of the reflexions (n_2, n_3), $\mathbf{R}_p, \mathbf{R}_q$ are the known rotation matrices involved in the triplet and s_1, s_2 are the signs for which $\mathbf{u}_{a_1}$ and $\mathbf{u}_{a_2}$ satisfy (10); then

$$\Delta_3=-2\pi(s_1\mathbf{u}_{a_1}\mathbf{T}_p+s_2\mathbf{u}_{a_2}\mathbf{T}_q).$$

It is immediately seen that (10) plays the same role as (3*a*): therefore our reduced $\sum_2$ list directly provides all pairs $(\mathbf{u}_1, \mathbf{u}_2)$ which form two-phase seminvariants of first rank. In addition it may be noted that starting from (3*a*) we were obliged [*via* (4) and (5)] to seek the integer solutions $\{h_1\}$ for all pairs $(\mathbf{R}_\alpha, \mathbf{R}_\beta)$. On using (10) *SEM*2 directly receives from *SEM*1 the pairs $(\mathbf{R}_n, \{\mathbf{h}_1\})$ to be used for constructing the quartets

$$\Phi_4=s_1\varphi_{\mathbf{u}_{a_1}\mathbf{R}_p}+s_2\varphi_{\mathbf{u}_{a_2}\mathbf{R}_q}-\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_1\mathbf{R}_n} \tag{11}$$

which coincide with quartets (6*a*): by comparison it may be assumed that

$$s_1\mathbf{u}_{a_1}\mathbf{R}_p=\mathbf{u}_1=\mathbf{h}_1-\mathbf{h}_2\mathbf{R}_\beta, \tag{12a}$$

$$s_2\mathbf{u}_{a_2}\mathbf{R}_q=\mathbf{u}_2\mathbf{R}_\beta=(\mathbf{h}_2-\mathbf{h}_1\mathbf{R}_\alpha)\mathbf{R}_\beta, \tag{12b}$$

$$\mathbf{R}_n=\mathbf{R}_\alpha\mathbf{R}_\beta. \tag{12c}$$

It apparently seems that in order to obtain indices of the reflexions involved in (6*b*) we have to solve the system of equations (12) with respect to $\mathbf{R}_\alpha, \mathbf{R}_\beta$ and $\mathbf{h}_2$ [$\mathbf{u}_{a_1}\mathbf{R}_p$ and $\mathbf{u}_{a_2}\mathbf{R}_q$ are provided by the $\sum_2$ list and the pair $(\mathbf{h}_1, \mathbf{R}_n)$ is supplied from *SEM*1]. Solving (12) for every pair $(\mathbf{u}_{a_1}\mathbf{R}_p, \mathbf{u}_{a_2}\mathbf{R}_q)$ and for every $\mathbf{h}$ is a very time-consuming job. However that may be avoided too, provided we do not determine $\mathbf{R}_\alpha, \mathbf{R}_\beta$ and $\mathbf{h}_2$ explicitly, but seek only the product $\mathbf{h}_2\mathbf{R}_\beta$.

Indeed, if the indices in (6*b*) are multiplied by $\mathbf{R}_\beta$, after application of (12) the quartet (6*b*) becomes

$$s_1\varphi_{(\mathbf{u}_{a_1}\mathbf{R}_p)\mathbf{R}_\alpha\mathbf{R}_\beta}+s_2\varphi_{\mathbf{u}_{a_2}\mathbf{R}_q}-\varphi_{\mathbf{h}_2\mathbf{R}_\beta}+\varphi_{\mathbf{h}_2\mathbf{R}_\beta\mathbf{R}_\alpha\mathbf{R}_\beta},$$

which may be rewritten as

$$\Phi_4'=s_1\varphi_{(\mathbf{u}_{a_1}\mathbf{R}_p)\mathbf{R}_n}+s_2\varphi_{\mathbf{u}_{a_2}\mathbf{R}_q}-\varphi_{\mathbf{h}_2\mathbf{R}_\beta}+\varphi_{\mathbf{h}_2\mathbf{R}_\beta\mathbf{R}_n}. \tag{13}$$

Since $-\mathbf{h}_2\mathbf{R}_\beta$ is just the second cross of the quartet (6*a*) the indices of (13) instantaneously follow from (11).

We now quote some additional characteristics of the algorithm which further speeds up the procedure:

(*a*) According to (11) Φ_4 reduces to

$$\begin{aligned}\Phi_4&=\Psi_2-2\pi(s_1\mathbf{u}_{a_1}\mathbf{T}_p+s_2\mathbf{u}_{a_2}\mathbf{T}_q)-2\pi\mathbf{h}_1\mathbf{T}_n\\&=\Psi_2+\Delta_3-2\pi\mathbf{h}_1\mathbf{T}_n,\end{aligned}$$

where

$$\Psi_2=s_1\varphi_{\mathbf{u}_{a_1}}+s_2\varphi_{\mathbf{u}_{a_2}} \tag{14}$$

is the quantity to be estimated, and Δ_3 is given by the $\sum_2$ list. Likewise Φ_4' is equal to

$$\Phi_4'=\Psi_2+\Delta_3-2\pi s_1\mathbf{u}_{a_1}\mathbf{R}_p\mathbf{T}_n-2\pi\mathbf{h}_2\mathbf{R}_\beta\mathbf{T}_n. \tag{15}$$

Since, according to (12*a*), $s_1\mathbf{u}_{a_1}\mathbf{R}_p=\mathbf{h}_1-\mathbf{h}_2\mathbf{R}_\beta$, then

$$\Phi_4'=\Psi_2+\Delta_3-2\pi\mathbf{h}_1\mathbf{T}_n=\Phi_4.$$

Table 1. *Structure codes, space groups, chemical data and some numerical results (see text) are shown for 23 test structures*

In order to save space references are not given. For most of them the reader is referred to magnetic tapes distributed by crystallographic groups in York, England and Göttingen, Federal Republic of Germany.

Code	Space group	Formula	Number of non-H atoms in the cell	NRIF	NTRIP	n.s.s.	(n.o.p.)
INOS	$P2_1/n$	$C_6H_{12}O_6.H_2O$	104	304	3508	>8000	234
ERGO	$P2_12_12_1$	$C_{28}H_{44}O$	232	382	>8000	7964	181
RIFOLO	$P2_1$	$C_{39}H_{49}NO_{13}$	106	362	2890	2223	333
GRA4	$P\bar{1}$	$C_{30}H_{22}N_2O_4$	72	394	5618	7494	>400
QUINO	$R\bar{3}$	C_6O_2	432	296	5877	>8000	>400
DIOLE	$I\bar{4}2d$	$C_{10}H_{18}O_2$	192	182	6428	>8000	42
APAPA	$P4_12_12$	$C_{30}H_{35}N_{15}O_{16}P_2$	504	426	>8000	>8000	66
PROLINA	$P2_1$	$C_{26}H_{40}N_4O_7$	74	298	1741	889	190
CEPHAL	$C2$	$C_{18}H_{21}NO_3$	176	326	3572	4354	106
NEWQB	$P\bar{1}$	$C_{24}H_{20}N_2O_5$	124	473	5073	7139	79
TPH	$C222_1$	$C_{24}N_2H_{20}$	312	306	3946	7994	335
BED	$I4$	$C_{26}H_{26}N_4O_4$	272	286	4254	>8000	369
MGHEX	$P3_1$	$C_{48}H_{68}N_{12}O_{12}\cdot$ $Mg(ClO_4)_2(CH_3CN)_4$	285	489	>8000	3392	104
GOLDMAN	Cc	$C_{28}H_{16}$	224	374	4730	65	18
MUNICH1	$C2$	$C_{20}H_{16}$	160	310	3708	5296	173
DIAM	$P4_2/n$	$C_{14}H_{20}O$	120	260	6472	>8000	>400
TUR10	$P6_322$	$C_{15}H_{24}O_2$	204	218	>8000	>8000	256
TPALA	$P2_1$	$C_{28}H_{42}N_4O_7$	78	306	3571	1255	209
BOBBY	$P2_13$	$CaNaN(CH_2CO_2)_3$	60	68	2062	5508	>400
NO55	$Fdd2$	$C_{20}H_{24}N_4$	384	246	4538	>8000	100
LOGANIN	$P2_12_12_1$	$C_{17}H_{26}O_{10}$	68	258	2916	2619	252
SELENID	$P2_1$	$C_{22}H_{28}O_2Se$	50	254	4736	689	226
LITHO	$P2_12_12_1$	$C_{24}H_{40}O_3$	108	258	4297	3922	209

Thus both the quartets Φ_4' and Φ_4 are referred to Ψ_2 by the same phase shift.

(*b*) The first cross term of all the quartets (11) relative to the various seminvariants $(+n_j, -n_k)$, $(-n_s, +n_t), \ldots$ in (9) coincides constantly with n_i whose $|E|$ is supplied by the $\sum_2$ list.

(*c*) No search of signs s_1 and s_2 which make $s_1\varphi_{\mathbf{u}_1} + s_2\varphi_{\mathbf{u}_2}$ a seminvariant is needed: indeed they are directly supplied by the special $\sum_2$ list.

It may be concluded that, once one-phase structure seminvariants of first rank have been identified (that is a very fast job), singling out two-phase seminvariants by means of the algorithm described above requires almost the same computing time as a usual $\sum_2$ list. That allows us to employ, for searching seminvariant pairs $E_{\mathbf{u}_1}$ and $E_{\mathbf{u}_2}$, the same subset of NRIF reflexions commonly used for the $\sum_2$ list (while **H** varies over all the subset of the one-phase seminvariants). In Table 1 we show, for 19 crystal structures of different complexity and covering a variety of space groups, the NRIF values, the triplet invariants (NTRIP) found among the NRIF reflexions, and the number of two-phase seminvariants of first rank (n.s.s.) found by the procedure.

3. The probabilistic estimation of Ψ_2

Every Ψ_2 found by the previous procedure may be estimated *via* its first representation according to probabilistic formulas described in papers III and IV of this series. Their different algebraic forms are a consequence of the various approximations introduced into the probabilistic approaches and of the particular algorithms chosen for the computation. In the present paper a further modification of the original formulas is used.

For centrosymmetric space groups the evaluation of the sign of cos Ψ_2 is performed by means of a sign probability distribution of type

$$P_+ \simeq 0{\cdot}5 + 0{\cdot}5 \tanh(\alpha_2/2); \tag{16a}$$

for non-centrosymmetric space groups the distribution of Ψ_2 is given by the von Mises distribution

$$P(\Psi_2) \simeq [2\pi I_0(\alpha_2)]^{-1} \exp[\alpha_2 \cos(\Psi_2 - \theta_2)]. \tag{16b}$$

α_2 and θ_2 are parameters which can be calculated from the contributions Q arising from the various pairs of quartets (11) and (13) obtained when $\mathbf{h}_1$ varies into $\{\mathbf{h}_1\}$ and $\mathbf{R}_n$ into the set of matrices satisfying (8).

For abbreviation let us define

$$\varepsilon = |E|^2 - 1, \quad \varepsilon_1 = \varepsilon_{\mathbf{u}_{a_1}}, \quad \varepsilon_2 = \varepsilon_{\mathbf{u}_{a_2}},$$

$$\varepsilon_{3,1} = \varepsilon_{\mathbf{h}_1}, \quad \varepsilon_{12,1} = \varepsilon_{s_1\mathbf{u}_{a_1}\mathbf{R}_p + s_2\mathbf{u}_{a_2}\mathbf{R}_q}, \quad \varepsilon_{23,1} = \varepsilon_{s_2\mathbf{u}_{a_2}\mathbf{R}_q - \mathbf{h}_1},$$

$$\varepsilon_{3,2} = \varepsilon_{\mathbf{h}_2\mathbf{R}_\beta}, \quad \varepsilon_{12,2} = \varepsilon_{(s_1\mathbf{u}_{a_1}\mathbf{R}_p)\mathbf{R}_n + s_2\mathbf{u}_{a_2}\mathbf{R}_q},$$

$$\varepsilon_{13,2} = \varepsilon_{(s_1\mathbf{u}_{a_1}\mathbf{R}_p)\mathbf{R}_n - \mathbf{h}_2\mathbf{R}_\beta}.$$

Then

$$Q = (a + b) \tag{17}$$

Table 2. *QUINOL: the numbers of triplet* ($n = 3$) *and of two-phase seminvariants* ($n = 2$) *relationships having reliability parameters larger than fixed ARG values are given* (*in parentheses are the number of wrong relationships*)

ARG	$n = 3$	$n = 2$
0·4	5816 (932)	400 (22)
0·8	2847 (285)	317 (10)
1·2	1216 (69)	181 (3)
1·6	529 (16)	112 (1)
2·4	130 (1)	57 (0)
3·2	37 (0)	36 (0)
5·0	2 (0)	14 (0)

where

$$a = \varepsilon_{3,1} + \varepsilon_{3,2},$$

$$b = \varepsilon_{3,1}\varepsilon_{12,1} + \varepsilon_{3,2}\varepsilon_{12,2} + 2\varepsilon_{3,1}\varepsilon_{3,2} + 2\varepsilon_{3,1}\varepsilon_{23,1} + 2\varepsilon_{3,2}\varepsilon_{13,2}.$$

In order to take into account the contributions of the various pairs of quartets the parameters α_2 and θ_2 have to be calculated:

$$\alpha_2 = (|E_1E_2|/N) \times \left\{\left[\sum_j Q_j \sin \Delta_j\right]^2 + \left[\sum_j Q_j \cos \Delta_j\right]^2\right\}^{1/2}, \quad (18)$$

$$\tan \theta_2 = \left(\sum_j Q_j \sin \Delta_j\right) \Big/ \left(\sum_j Q_j \cos \Delta_j\right) \quad (19)$$

where

$$\Delta_j = (-\Delta_3 + 2\pi \mathbf{h}_1 \mathbf{T}_n).$$

Seminvariants with restricted phase values can be estimated by slightly modifying (16) and (18). Let $(\theta_{2r}, \theta_{2r} + \pi)$ be the allowed phase values and $\alpha_{2r} = \alpha_2 \cos(\theta_2 - \theta_{2r})$. If $\alpha_{2r} \geq 0$ then $\Psi_2 = \theta_{2r}$ with reliability parameter α_{2r}; if $\alpha_{2r} < 0$ then $\Psi_2 = \theta_{2r} + \pi$ with reliability parameter $|\alpha_{2r}|$.

Equations (16) may be easily compared with corresponding distributions given in papers III and IV. In particular Q, as defined by (IV.43), almost reduces to our Q term [as in (17)] if $C = 0$. Such an assumption is not theoretically justified; however, our experimental tests clearly showed that $|C|$ in (IV.43) was too large when $|a + b|$ was large, so that reliable phase indications were often weakened because of large denominators $(1 + C)$. We have thus preferred to assume $C = 0$ always and empirically rescale α_2 parameters on triplet reliability parameters. In the last column of Table 1 for each structure the number (n.o.p.) of two-phase seminvariants with α_2 larger than 0·4 is given (only a maximum of 400 seminvariants is stored by *SIR*).

To be successful the above procedure requires some special care in order to avoid the repeated use in (18) and (19) of the same information. A typical situation occurs when some of the cross terms of the pair of quartets (11) and (13) coincide because of special algebraic conditions among $\mathbf{u}_{a_1}, \mathbf{u}_{a_2}, \mathbf{h}_1, \mathbf{h}_2$: in such cases (17)-(19) do not hold. In accordance with Appendix *A** such special quartets are readily recognized by some algebraic tests and probabilistic formulas are suitably modified.

Table 3. *Number of correct solutions and number of trials are given for the* 23 *crystal structures quoted in Table* 1, *according to conditions defined in the text*

Different numbers of trials for the same structure may be obtained when the *CONVERGENCE* procedure chooses different starting sets according to the different prior information relative to PROT 1 and PROT 2.

	PROT 1	PROT 2
INOS	4 (32)	10 (32)
ERGO	— (30)	— (36)
RIFOLO	13 (30)	24 (30)
GRA4	8 (32)	8 (32)
QUINO	5 (32)	14 (32)
DIOLE	— (36)	4 (30)
APAPA	1 (72)	2 (72)
PROLINA	— (24)	7 (24)
CEPHAL	— (24)	1 (24)
NEWQB	1 (32)	1 (32)
TPH	— (56)	— (36)
BED	7 (36)	10 (36)
MGHEX	— (48)	— (48)
GOLDMAN	1 (24)	1 (24)
MUNICH1	— (24)	— (24)
DIAM	4 (32)	20 (32)
TUR10	12 (56)	8 (32)
TPALA	— (24)	1 (24)
BOBBY	28 (72)	72 (72)
NO55	8 (56)	13 (56)
LOGANIN	2 (24)	8 (36)
SELENID	10 (24)	8 (24)
LITHO	1 (30)	8 (30)

Table 2 is a typical example of how triplet and seminvariant relationships may be ranked according to $G = 2|E_1E_2E_3|/N^{1/2}$ and to α_2 parameters. The table clearly suggests that information provided by two-phase seminvariants is not negligible with respect to that contained in triplet relationships.

4. Experimental results and conclusions

The 23 crystal structures quoted in Table 1 have been processed by *SIR* according to the following default conditions: (*a*) five starting set phases are represented by magic integers (Main, 1978); (*b*) triplet invariants are estimated *via* the P_{10} formula (Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori, 1984).

For each structure the number of correct solutions and the number of trials (in parentheses) are shown in Table 3 (column PROT 1). No solution is found in these conditions for eight structures.

---

* Appendix *A* has been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 51178 (3pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

The same 23 structures were later processed by *SIR* using one-phase (according to Cascarano, Giacovazzo, Calabrese, Burla, Nunzi, Polidori & Viterbo, 1984) and two-phase seminvariants in addition to conditions (*a*) and (*b*). In Table 3 for each structure the number of correct solutions is given (column PROT 2) together with the number of trials (in parentheses).

Only four structures remain unsolved by *SIR* [however, it cannot be concluded that *SIR* is unable to solve those structures using non-default conditions; see for example, Burla, Giacovazzo & Polidori (1987)]. Very often the ratio (number of correct solutions/number of trials) is larger in PROT 2 than in PROT 1. Thus Table 3 provides clear evidence of how relevant phase seminvariants may be for the success of direct phasing procedures.

Two further observations can be made.

(1) The search of seminvariant pairs is made among the largest NRIF reflections commonly used for the $\sum_2$ list, while **H** varies over the complete subset of one-phase seminvariants. Limiting the cross term **H** to the strongest reflexions reduces both the computing time of the procedure and (dramatically) the number of available two-phase seminvariants. A large subset of reliable two-phase seminvariants can be more easily found when **H** is unrestricted.

(2) The factor SC used to rescale α_2 on the triplet reliability parameter α_3 ($\alpha_3=2|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}|/N^{1/2}$) was empirically chosen so as to satisfy

$$0{\cdot}6\times \mathrm{SC}\times\sum\alpha_3=\sum\alpha_2,$$

where the summations are over the most reliable 400 triplets and two-phase seminvariants respectively. The scaling scheme is unsatisfactory from the theoretical point of view, even if it works quite well in practice for all our test structures. It is hoped that future probability distributions will make available more efficient formulas for estimating seminvariants.

References

BURLA, M. C., GIACOVAZZO, C. & POLIDORI, G. (1987). *Acta Cryst.* A**43**, 797-802.

CASCARANO, G., GIACOVAZZO, C., BURLA, M. C., NUNZI, A., POLIDORI, G., CAMALLI, M., SPAGNA, R. & VITERBO, D. (1985). IX Eur. Crystallogr. Meet., Torino. Abstract 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., CALABRESE, G., BURLA, M. C., NUNZI, A., POLIDORI, G. & VITERBO, D. (1984). *Z. Kristallogr.* **167**, 37-47.

DEBAERDEMAEKER, T. & WOOLFSON, M. M. (1972). *Acta Cryst.* A**28**, 477-481.

GIACOVAZZO, C. (1977*a*). *Acta Cryst.* A**33**, 531-538.

GIACOVAZZO, C. (1977*b*). *Acta Cryst.* A**33**, 539-547.

GIACOVAZZO, C. (1978). *Acta Cryst.* A**34**, 27-30.

GIACOVAZZO, C. (1979). *Acta Cryst.* A**35**, 296-305.

GIACOVAZZO, C., SPAGNA, R., VICKOVIĆ, I. & VITERBO, D. (1979). *Acta Cryst.* A**35**, 401-412.

GRANT, D., HOWELLS, R. & ROGERS, D. (1957). *Acta Cryst.* **10**, 489-497.

GREEN, E. A. & HAUPTMAN, H. (1976). *Acta Cryst.* A**32**, 940-944.

GREEN, E. A. & HAUPTMAN, H. (1978*a*). *Acta Cryst.* A**34**, 216-223.

GREEN, E. A. & HAUPTMAN, H. (1978*b*). *Acta Cryst.* A**34**, 230-241.

HAUPTMAN, H. & GREEN, E. A. (1978). *Acta Cryst.* A**34**, 224-229.

MAIN, P. (1978). *Acta Cryst.* A**34**, 31-38.

Acta Cryst. (1989). A**45**, 104-109

Phase Dependence of Kikuchi Patterns. I. Theory

BY D. M. BIRD AND A. G. WRIGHT

School of Physics, University of Bath, Bath BA2 7AY, England

(*Received* 6 *May* 1988; *accepted* 22 *July* 1988)

Abstract

Thermal diffuse scattering in high-energy electron diffraction is analysed using the Einstein model of lattice vibrations. An expression for the intensity distribution in the Kikuchi pattern is obtained which includes thickness-dependent terms (*i.e.* a dependent-Bloch-wave theory is used) and the effects of a general crystal structure. The corresponding two-beam limit is shown to consist of four distinct terms, two of which depend on the phase of the structure factor. One of these is found to be non-zero only for non-centrosymmetric crystals and for relatively thin crystals. It leads to an asymmetric Kikuchi band, even in a symmetrical scattering geometry. This asymmetry may be used to determine the polarity of non-centrosymmetric crystals.

1. Introduction

The theory of thermal diffuse scattering in high-energy electron diffraction and the associated process of Kikuchi pattern formation have been extensively studied over the last thirty years. Many of the earlier

0108-7673/89/010104-06$03.00