

## References

- BARRON, T. H. K. (1977). *Acta Cryst.* **A33**, 602-604.  
 BEG, M. M. (1976). *Acta Cryst.* **A32**, 154-156.  
 DARWIN, C. G. (1922). *Philos. Mag.* **43**, 800-829.  
 GROENEWEGEN, P. P. M. & HUISZON, C. (1972). *Acta Cryst.* **A28**, 164-169.  
 HAMILTON, W. C. (1957). *Acta Cryst.* **10**, 629-634.  
 HEWAT, A. W. (1979). *Acta Cryst.* **A35**, 248-249.  
 HILL, R. J. & HOWARD, C. J. (1987). *J. Appl. Cryst.* **20**, 467-474.  
 HOWARD, C. J. & SABINE, T. M. (1974). *J. Phys. C*, **7**, 3453-3465.  
 KOESTER, L. & YELON, W. B. (1982). *Summary of Low-Energy Neutron Scattering Lengths and Cross-Sections*. IUCr Commission on Neutron Diffraction.  
 LARSON, A. C. & VON DREELE, R. B. (1986). *GSAS - Generalised Crystal Structure Analysis System*. Los Alamos Report LAUR 86-748, Los Alamos National Laboratory, USA.  
 LAWRENCE, J. L. (1973). *Acta Cryst.* **A29**, 94-95.  
 MARSHALL, W. & LOVESEY, S. W. (1971). *Theory of Thermal Neutron Scattering*. Oxford: Clarendon Press.  
 OLEKHOVICH, N. M. & OLEKHOVICH, A. I. (1978). *Acta Cryst.* **A34**, 321-326.  
 RIETVELD, H. M. (1969). *J. Appl. Cryst.* **2**, 65-71.  
 SABINE, T. M. (1985). *Aust. J. Phys.* **38**, 507-518.  
 SABINE, T. M. (1988). *Acta Cryst.* **A44**, 368-373.  
 SANGER, P. L. (1969). *Acta Cryst.* **A25**, 694-702.  
 TOGAWA, S. (1965). *J. Phys. Soc. Jpn*, **20**, 742-752.  
 VON DREELE, R. B., JØRGENSEN, J. D. & WINDSOR, C. G. (1982). *J. Appl. Cryst.* **15**, 581-589.  
 WALKER, D. G. & HICKMAN, B. S. (1965). *Philos. Mag.* **12**, 445-451.  
 WERNER, S. A., BERLINER, R. R. & ARIF, M. (1986). *Physica (Utrecht)*, **137B**, 245-255.

*Acta Cryst.* (1988). **A44**, 379-382

## Combining Direct Methods with Isomorphous Replacement or Anomalous Scattering Data. VI. Incorporation of Heavy-Atom Information into Hauptman's Distributions\*

BY HAO QUAN AND FAN HAI-FU

*Institute of Physics, Academia Sinica, Beijing, China*

(Received 23 September 1987; accepted 15 January 1988)

### Abstract

Heavy-atom information has been incorporated into Hauptman's distributions of three-phase structure invariants for both isomorphous replacement and anomalous scattering cases. Reliable estimates of individual phases can be obtained by introducing the phase doublet expression  $\varphi_H = \varphi'_H \pm |\Delta\varphi_H|$ . A test calculation with error-free data of insulin showed results better than previous methods.

### Introduction

In recent years, approaches based on the combination of direct methods with SIR (single isomorphous replacement) or OAS (one-wavelength anomalous scattering) have been well developed. Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing (1984) proposed that in the case of SIR or OAS the phase of a structure factor can be expressed as  $\varphi_H = \varphi'_H \pm |\Delta\varphi_H|$ , where  $\varphi'_H$  can be calculated from the heavy-atom sites and  $|\Delta\varphi_H|$  can be derived from the experimental diffraction data. The phase problem is thus reduced to a matter of making a sign choice. Satisfactory

estimates of individual phases were obtained by combining this phase-difference relation with Cochran's (1955) probability distribution. Hauptman (1982*a, b*) derived new probability distributions for three-phase structure invariants in the SIR and OAS cases. The formulas proved to be more reliable than Cochran's distribution. However, there is still the potential to improve Hauptman's formulas by making use of heavy-atom information. In this context Fortier, Moore & Fraser (1985) obtained the full range (-1 to 1) of estimates of cosine invariants. However, the procedure yields a twofold ambiguity which would lead to difficulties in the derivation of individual phases. In this paper, improved Hauptman distributions are given, which make full use of the heavy-atom information. These distributions are then used instead of Cochran's distribution as the foundation of the individual phase derivation. The concept of 'best phase' (Fan Hai-fu, Han Fu-son & Qian Jin-zi, 1984) is also used for error treatment.

### Theoretical basis

#### 1. The probability distribution of three-phase structure invariants

According to Hauptman (1982*a*), there are four kinds of three-phase structure invariants in the SIR

\* Part of this paper was presented at the International Symposium on Molecular Structure, Beijing, China, 15-19 September 1986.

case, classified as

$$\begin{aligned}\omega_0 &= \varphi_{\text{H}} + \varphi_{\text{K}} + \varphi_{\text{L}} \\ \omega_1 &= \varphi_{\text{H}} + \varphi_{\text{K}} + \Psi_{\text{L}} \\ \omega_2 &= \varphi_{\text{H}} + \Psi_{\text{K}} + \Psi_{\text{L}} \\ \omega_3 &= \Psi_{\text{H}} + \Psi_{\text{K}} + \Psi_{\text{L}},\end{aligned}\quad (1)$$

where  $\varphi_j$  and  $\Psi_j$  ( $j = \text{H, K or L}$ ) represent the phases of the structure factors from the native protein and the heavy-atom derivative respectively. With the similar approach of Hauptman, the probability distribution for all kinds of three-phase structure invariants can be obtained with the following unified expression:

$$P_i(\omega_i | R_1, R_2, R_3, S_1, S_2, S_3) = (1/K_i) \exp(A_i \cos \omega_i) \quad (i = 0, 1, 2, 3), \quad (2)$$

where  $R_j$  and  $S_j$  ( $j = 1, 2, 3$ ) are the magnitudes of normalized structure factors associated with the native protein and heavy-atom derivative respectively. The notation system used here is the same as Hauptman's except  $A_i$ . With Hauptman's (1982a) method, in order to calculate  $A_i$  one has to use the probability distribution of the two-phase invariant  $\varphi_{\text{H}} - \Psi_{\text{H}}$ , whose most probable value is  $\varphi_{\text{H}} - \Psi_{\text{H}} = 0$ . This will seriously weaken the difference between the native and derivative structure; consequently, it tends to a degeneration towards the SIR phases leaving the phase ambiguity unresolved.

In order to overcome this disadvantage, more accurate values of  $\cos |\varphi - \Psi|$  should be used. This can be accomplished by calculating  $\cos |\varphi - \Psi|$  from experimental data rather than by making a probabilistic estimation (Fortier, Moore & Fraser, 1985). Letting  $F_{\text{P}}$ ,  $F_{\text{PQ}}$ ,  $F_{\text{Q}}$  denote the structure factors contributed from the native protein, the derivative and the heavy-atom substructure respectively, we have

$$\cos |\varphi - \Psi| = (F_{\text{P}}^2 + F_{\text{PQ}}^2 - F_{\text{Q}}^2) / 2F_{\text{P}}F_{\text{PQ}}. \quad (3)$$

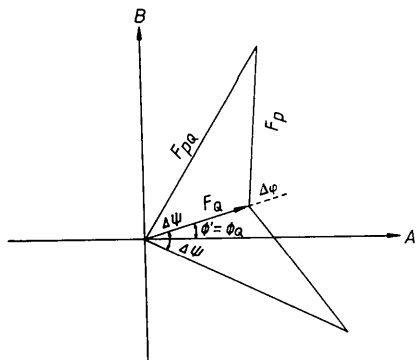


Fig. 1. Phase-vector diagram showing the enantiomorphous phase doublet arising from the SIR method.

This can be used to replace  $T(z)$  in the expression for the  $A_j$ 's in equations (3.14), (3.18), (3.21) and (3.24) of Hauptman's paper.

In order to reduce the phase problem to a matter of sign choice, the following relations are substituted into Hauptman's formulas:

$$\varphi_{\text{H}} = \varphi'_{\text{H}} + \Delta\varphi_{\text{H}}, \quad (4)$$

$$\Psi_{\text{H}} = \varphi'_{\text{H}} + \Delta\Psi_{\text{H}}, \quad (5)$$

where  $\varphi'_{\text{H}}$  is the phase of  $F_{\text{Q}}$  while  $|\Delta\varphi_{\text{H}}|$ ,  $|\Delta\Psi_{\text{H}}|$  can be calculated from the phase-vector diagram (Fig. 1). Finally, for further simplification, the triplet phase invariants of the heavy-atom substructure,  $\varphi'_{\text{H}} + \varphi'_{\text{K}} + \varphi'_{\text{L}}$ , are made equal to zero. This is a good approximation since the heavy-atom substructure is in fact a simple small structure. The Hauptman distribution now becomes

$$P_i(\Delta\omega_i | R_1, R_2, R_3, S_1, S_2, S_3) = (1/K_i) \exp(A_i \cos \Delta\omega_i) \quad (i = 0, 1, 2, 3), \quad (6)$$

where

$$\begin{aligned}\Delta\omega_0 &= \Delta\varphi_{\text{H}} + \Delta\varphi_{\text{K}} + \Delta\varphi_{\text{L}} \\ \Delta\omega_1 &= \Delta\varphi_{\text{H}} + \Delta\varphi_{\text{K}} + \Delta\Psi_{\text{L}} \\ \Delta\omega_2 &= \Delta\varphi_{\text{H}} + \Delta\Psi_{\text{K}} + \Delta\Psi_{\text{L}} \\ \Delta\omega_3 &= \Delta\Psi_{\text{H}} + \Delta\Psi_{\text{K}} + \Delta\Psi_{\text{L}}.\end{aligned}\quad (7)$$

In the OAS case, the Hauptman distribution can be improved similarly. We have

$$\cos |\varphi_{\text{H}} + \varphi_{-\text{H}}| = (F_{\text{H}}^2 + F_{-\text{H}}^2 - 2F_{\text{Q}}'^2) / 2F_{\text{H}}F_{-\text{H}}, \quad (8)$$

where  $F_{\text{H}}$  and  $\varphi_{\text{H}}$  are the magnitude and phase of the structure factor contributed from both normal scattering and the real part of anomalous scattering, while  $F_{\text{Q}}'$  is the magnitude of the imaginary part of anomalous scattering from the heavy atoms (Fig. 2). The phase doublets are of the same form as (3). Incorporating (8) into Hauptman's formulas instead

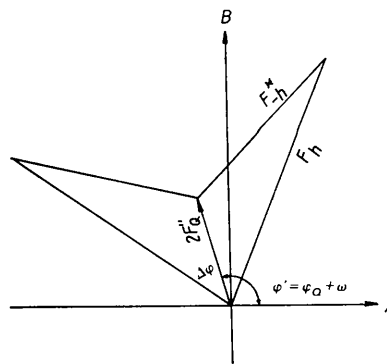


Fig. 2. Phase-vector diagram showing the enantiomorphous phase doublet arising from the OAS method.  $\omega$  is the phase difference between  $F_{\text{Q}}'$  and  $F_{\text{Q}}$ , which equals  $\pi/2$  if all the anomalous scattering atoms are of the same kind.

of  $\tau_i$  [see equations (3.39) (3.54) of Hauptman (1982b)], one obtains

$$P_i(\Omega_i) = (1/K_i) \exp [A_i \cos (\Omega_i - \omega_i)]$$

$$(j = 0, 1, 2, 3, -0, -1, -2, -3). \quad (9)$$

The notations of these distributions are similar to those of Hauptman.

### 2. Individual phase determination

The Cochran distribution and the distributions of (2) and (9) have a common form:

$$P(\varphi_H) = (1/K) \exp [B \cos (\varphi_H - \varphi_{H'} - \varphi_{H-H'} - \xi_{H,H'})]. \quad (10)$$

In a practical iterative procedure the calculation can be made much simpler if we can obtain the total distribution of  $\varphi_H$  by the product of individual  $P(\varphi_H)$ 's corresponding to different  $H'$ . However this is in conflict with Hauptman's theory since (2) and (9) are valid only for single  $\Sigma_2$  relationships. A compromise way to use (10) is to use (2) or (9) to estimate the value of  $\xi_{H,H'}^*$  while the Cochran distribution is used to obtain the value of  $B$ .

Replacing  $\varphi$  by  $\varphi' + \Delta\varphi$  in (10) one obtains the conditional probability for  $\Delta\varphi_H$  to be positive as

$$P_+(\Delta\varphi_H) = \frac{1}{2} + \frac{1}{2} \tanh \left[ k R_H \sin |\Delta\varphi_H| \sum_{H'} R_{H'} R_{H-H'} \times \sin (\Phi'_3 + \Delta\varphi_{H'} + \Delta\varphi_{H-H'} - \xi_{H,H'}) \right], \quad (11)$$

where

$$\Phi'_3 = -\varphi'_H + \varphi'_{H'} + \varphi'_{H-H'}.$$

In order to deal with the experimental SIR or OAS data, the concept of 'best phase' should be introduced (Fan Hai-fu, Han Fu-san & Qian Jin-zi, 1984). The 'best' value of the normalized structure factor can be

Table 1. Average errors in estimated values of 140 000 three-phase invariants arranged in descending order of  $|A_i|$  and cumulated in 14 groups (the number of invariants in each group is 10 000)

Gn: Group number. I: Results from (6). II: Results from Hauptman's formula. Min: Minimum value of  $|A_i|$  in the group. Per: Percentage of correctly estimated signs of cosine invariants. Er: Average magnitude of the error of invariants in degrees.

Gn	I			II		
	Min	Per (%)	Er (°)	Min	Per (%)	Er (°)
1	2.60	96.0	35.0	2.53	95.8	36.1
2	2.10	94.6	37.8	2.05	93.0	40.5
3	1.84	94.7	37.6	1.80	92.7	41.0
4	1.67	94.1	38.8	1.64	90.7	43.6
5	1.54	93.7	39.2	1.52	89.5	44.6
6	1.44	92.5	41.0	1.42	88.2	46.5
7	1.36	93.3	39.5	1.34	87.1	47.4
8	1.29	93.0	40.4	1.28	86.9	47.8
9	1.23	92.5	41.1	1.22	86.2	49.1
10	1.18	92.7	40.6	1.17	85.3	49.6
11	1.13	91.9	42.1	1.13	84.6	50.4
12	1.09	91.8	41.8	1.08	84.3	51.0
13	1.05	91.6	42.2	1.04	82.9	52.3
14	1.02	91.5	42.6	1.01	82.6	53.1

expressed as

$$E_{H_{best}} = m_H R_H \exp (i\alpha_{H_{best}}), \quad (12)$$

where  $\alpha_{H_{best}}$  and  $m_H$  are known as the best phase and the figure of merit in protein crystallography. Defining  $\Delta\varphi_{H_{best}} = \alpha_{H_{best}} - \varphi'_H$ , one obtains

$$\tan (\Delta\varphi_{H_{best}}) = 2(P_+ - \frac{1}{2}) \sin |\Delta\varphi_H| / \cos \Delta\varphi_H, \quad (13)$$

$$m_H = \exp (-\sigma_H^2/2) [P_+^2 + P_-^2 + 2P_+P_- \cos 2\Delta\varphi_H]^{1/2}, \quad (14)$$

where  $\sigma_H$  can be obtained from the standard deviation  $D$  of the 'lack of closure error' (Blow & Crick, 1959). Replacing  $E_{H'}$  and  $E_{H-H'}$  by their 'best' values, (11) becomes

$$P_+ = \frac{1}{2} + \frac{1}{2} \tanh \left[ k R_H \sin |\Delta\varphi_H| \sum_{H'} m_{H'} R_{H'} m_{H-H'} R_{H-H'} \times \sin (\Phi'_3 + \Delta\varphi_{H'_{best}} + \Delta\varphi_{H-H'_{best}} - \xi_{H,H'}) \right]. \quad (15)$$

Starting phases can be obtained by setting initial values of all the  $P_+$ 's equal to 0.5. The flow chart of the iterative process is given in Fig. 3. For the sake of simplicity,  $\Phi'_3$  is set to zero.

\* In the SIR case  $\xi_{H,H'}$  equals 0 or  $\pi$  according as  $A_i$  in (2) is positive or negative. In the OAS case  $\xi_{H,H'}$  equals  $\omega_i$  in (9).

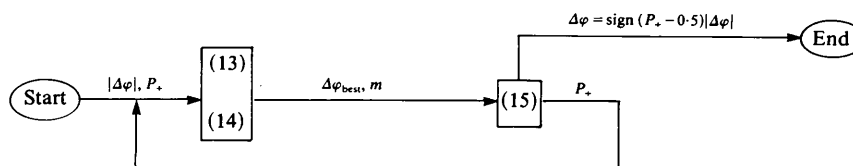


Fig. 3. Flow chart of iteration.

Table 2. Errors in the estimation of signs of 1000  $\Delta\varphi_{\text{H}}$ 's arranged in descending order of  $|P_+ - \frac{1}{2}| + \frac{1}{2}$ 

Procedure 1: Test calculation for formulas of this paper. Procedure 2: Abstracted Table 2 of Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing (1984). Procedure 3: Results of Langs (1986). Gn: Group number. Ngr: Number of reflections in the group.  $P$  (%): Minimum value of  $|P_+ - \frac{1}{2}| + \frac{1}{2}$  in the group. Per (%) Percentage of the correctly estimated signs of  $\Delta\varphi$ . Er (°): Average magnitude of the error of phases in degrees. C1: Results from Fig. 3 starting from  $P_+ = 0.5$ . C2: Results from Fig. 3 using results of C1 as input.

Gn	Ngr	Procedure 1						Procedure 2						Procedure 3
		C1			C2			I			II			Er
		$P$	Per	Er	$P$	Per	Er	$P$	Per	Er	$P$	Per	Er	Er
1	200	90.7	94.5	6.6	99.6	96.5	3.7	88.2	94.5	7.0	98.3	97.5	2.0	
2	400	75.5	90.0	10.2	94.4	93.5	6.9	74.6	86.3	15.0	89.7	93.3	7.0	
3	600	63.3	82.7	15.7	78.7	85.7	11.4	63.2	80.7	20.0	74.8	86.0	13.0	
4	800	54.2	77.6	17.7	61.5	79.3	14.7	55.3	72.5	24.0	59.4	78.4	17.0	
5	1000	50.0	73.1	17.5	50.0	75.3	15.7	50.0	67.6	25.0	50.0	73.8	18.0	50.0

Table 3. Errors in the estimation of signs of 2000  $\Delta\varphi_{\text{H}}$ 's arranged in descending order of  $|P_+ - \frac{1}{2}| + \frac{1}{2}$ 

All the notations are the same as in procedure 1 of Table 2.

	Gn	Ngr	$P$ (%)	Per (%)	Er (°)
C1	1	2000	50.0	74.0	17.2
	1	200	100.0	98.5	1.6
	2	400	100.0	95.2	5.1
	3	600	100.0	92.2	8.2
	4	800	100.0	88.2	12.3
C2	5	1000	99.7	85.9	13.5
	6	1200	97.4	83.1	14.8
	7	1400	89.7	79.9	15.8
	8	1600	76.9	77.4	16.6
	9	1800	60.6	75.0	17.0
	10	2000	50.0	74.9	16.3

### Test calculation

All the test calculations below are based on error-free SIR data from the native insulin (molecular weight  $\sim 12\,000$ ) and its Pb derivative. Crystals of insulin belong to space group  $R3$  with unit-cell parameters  $a = 82.5$ ,  $c = 34.0$  Å,  $\gamma = 120^\circ$  and  $Z = 9$ . The data were calculated from the known atomic parameters. There are 6371 independent reflections within the resolution limit of 1.9 Å.

#### 1. Test of equation (6)

2000 largest  $E$ 's of the native protein and 2000 largest  $G$ 's of the Pb derivative were selected to test equation (6). Meanwhile, as a comparison, triplet structure invariants were also estimated by Hauptman's distribution. The results are summarized in Table 1, in which the  $\Sigma_2$  relationships are arranged and grouped in descending order of calculated  $|A|$ 's. As anticipated the incorporation of heavy-atom information led to better results than those from the original Hauptman formula.

#### 2. Individual phase derivation according to flow chart of Fig. 3

Both Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing (1984) and Langs (1986) employed error-free

data for insulin in their test calculations. For comparison with their results, we use the 1000 largest  $E$ 's and the corresponding  $G$ 's to estimate individual phases. Only the largest 56 000  $\Sigma_2$  relationships are included in the calculation. The results are listed in Table 2. The column headed 'C1' shows that, with Fig. 3, it is possible to obtain a very large starting set of good quality without any preliminary knowledge of the sign of  $\Delta\varphi_{\text{H}}$ . After one cycle of iteration (see Fig. 3), the column headed 'C2' shows considerable improvement in the starting phases. As can be seen the calculated probabilities  $P$  from Fig. 3 are in good agreement with the percentages of the correctly estimated signs of  $\Delta\varphi$  and the results are much more accurate than those from other procedures.

For solving the phase problem of macromolecules, much larger starting sets will be required. We have also performed the test calculation with 2000 largest  $E$ 's and  $G$ 's, and 470 000  $\Sigma_2$  relationships. The results are summarized in Table 3. This shows that very large sets of phases can be evaluated initially and improved iteratively using the procedure shown in flow chart of Fig. 3. By comparison with Table 2, it can also be seen that the participation of more  $\Sigma_2$  relationships can lead to better results.

HQ is indebted to Professor H. Hauptman for helpful discussions and to Drs Yao Jia-xing, Gu Yuan-xin and Qian Jin-zi for their help in computational work. FHF would like to thank Professor G. Dodson and E. Dodson for making available the insulin data.

### References

- BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.  
 COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.  
 FAN HAI-FU, HAN FU-SON & QIAN JIN-ZI (1984). *Acta Cryst.* **A40**, 495-498.  
 FAN HAI-FU, HAN FU-SON, QIAN JIN-ZI & YAO JIA-XING (1984). *Acta Cryst.* **A40**, 489-495.  
 FORTIER, S., MOORE, N. J. & FRASER, M. E. (1985). *Acta Cryst.* **A41**, 571-577.  
 HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289-294.  
 HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632-641.  
 LANGS, D. A. (1986). *Acta Cryst.* **A42**, 362-368.

## Streaks of Instrumental Origin Near a Bragg Point

BY S. A. WERNER AND M. ARIF

Department of Physics and Astronomy, University of Missouri-Columbia, Columbia, MO 65211, USA

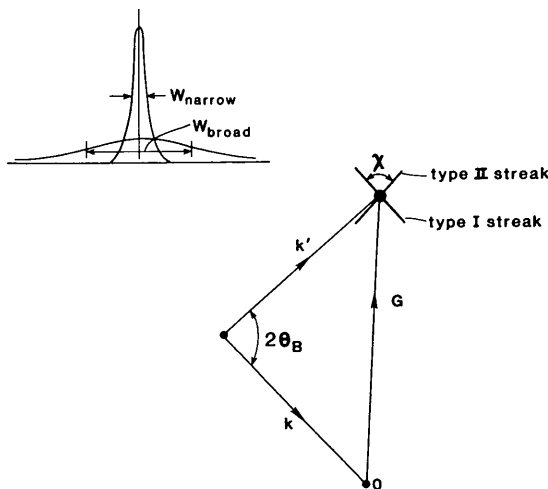
(Received 8 September 1987; accepted 25 January 1988)

### Abstract

It is predicted that, at some sufficiently high level of sensitivity, two streaks intersecting at the Bragg point and symmetrically situated about the reciprocal-lattice vector  $G$  should always be observable in a diffraction experiment.

### Introduction

We show here that the close examination of a Bragg point  $G$  in any diffraction experiment will reveal the presence of two streaks, intersecting at  $G$  and symmetrically situated about the radial line leading from the origin of reciprocal space to the point  $G$ , as shown in Fig. 1. The angle  $\chi$  between the two streaks will depend upon the instrumental parameters as discussed below, but can be expected to approach  $2\theta_B$  in most cases. We believe that the occurrence of these streaks is a very general instrumental phenomenon, and will be observed at some level of sensitivity in all experiments. Our thinking here is guided by



$$\tan \frac{\chi}{2} = \left( \frac{1 - \rho_0^2}{1 + \rho_0^2} \right) \tan \theta_B$$

$$\rho_0 = W_{\text{narrow}} / W_{\text{broad}}$$

Fig. 1. Orientation of instrumentally generated streaks in reciprocal space.

experience with high-resolution high-sensitivity triple-axis neutron diffraction experiments, which can routinely be carried out in the sensitivity range of  $10^{-4}$  to  $10^{-6}$  of the Bragg point intensity. The reasoning and analysis given here apply equally well to high-sensitivity X-ray diffraction experiments, such as those designed to examine Huang scattering or thermal diffuse scattering (TDS).

We label the two streaks type I and type II. The occurrence of the type I streak is well known, although its physical origin is rarely examined in any detail. It occurs when the Bragg point  $G$  falls on the sphere of reflection, giving a strong reflected beam at a scattering angle  $2\theta = 2\theta_B$ , but the detector is situated at an angle  $\gamma$  away from the nominal scattering angle  $2\theta_B$  that is required to fully accept this Bragg beam (Fig. 2). Normally, if  $\gamma$  is several times the resolution

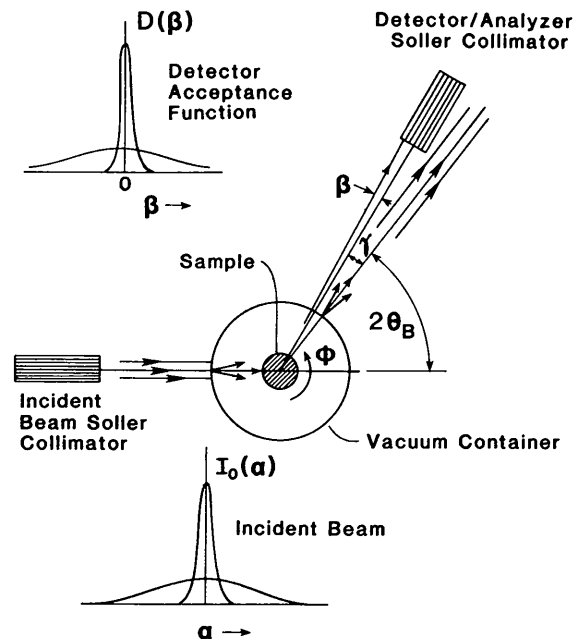


Fig. 2. Schematic diagram of typical diffraction experiment, showing the possibility of small-angle scattering of the incident and diffracted beams by the vacuum container surrounding the sample. The small-angle scattering has the effect of creating an incident beam  $I_0(\alpha)$  having a broad component in addition to the original narrow component. Small-angle scattering in the diffracted beam leads to a detector acceptance function  $D(\beta)$  having a broad component in addition to a narrow component.