

Note that the determined best fit is independent of the particular pair of base vectors. The resulting vector meets the requirements of case 24. The result vector is converted to the reduced cell (8.45, 9.98, 9.98, 112.4, 106.4, 106.4) which is the best least-squares fit of the original cell to case 24. The corresponding unreduced rhombohedral cell has  $a = 9.98 \text{ \AA}$ ,  $\alpha = 112.4^\circ$ .

The error of the fit can be computed in two ways: either the reduced vector can be projected onto a set of  $G^6$  base vectors normal to the base vectors of the particular case (the length of the sum of the projections being the error), or the length of the difference between the best-fit vector and the  $G^6$  vector before projection can be computed. The latter difference is

$$\begin{pmatrix} -0.69 \\ 0.43 \\ 0.43 \\ 0.87 \\ 0.30 \\ -0.90 \end{pmatrix} = \begin{pmatrix} 70.70 \\ 100.00 \\ 100.00 \\ -74.90 \\ -47.30 \\ -48.50 \end{pmatrix} - \begin{pmatrix} 71.39 \\ 99.57 \\ 99.57 \\ -75.77 \\ -47.60 \\ -47.60 \end{pmatrix}$$

with a length of  $1.6 \text{ \AA}$  (from the reduced vector to the best fit in case 24). In practice, this length should be compared with the error computed from the determination of the unit-cell parameters. For small molecules, diffractometers normally produce error vectors with lengths in the range  $0.05$  to  $0.4 \text{ \AA}^2$ . On the basis of this rule of thumb, the proposal that the original cell is rhombohedral should probably be rejected in the absence of other evidence. Certainly the differences in the original angles are outside the usual error bounds for diffractometers.

*Acta Cryst.* (1988). **A44**, 1018-1021

## The Use of $E$ -Magnitude Filters in Direct Methods

BY CHRISTOPHER J. GILMORE AND STEPHEN R. BROWN

*Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland*

(Received 18 April 1988; accepted 26 May 1988)

### Abstract

In direct methods difficulties can be experienced in solving structures in situations where data of high resolution are being used very early in the phasing process; in real space, this tends to build too much atomic detail before the molecular outline is fully

### References

- ANDREWS, L. C., BERNSTEIN, H. J. & PELLETIER, G. A. (1980). *Acta Cryst.* **A36**, 248-252.
- BLAHUT, R. E. (1985). *Proc. IEEE*, **73**, 874-893.
- BUERGER, M. J. (1957). *Z. Kristallogr.* **109**, 42-60.
- BURZLAFF, H., ZIMMERMANN, H. & DE WOLFF, P. M. (1983). In *International Tables for Crystallography*, Vol. A, edited by TH. HAHN. Dordrecht: Reidel. (Present distributor: Kluwer Academic Publishers, Dordrecht.)
- CHOW, T.-S. (1983). *Matrices and Linear Algebra*. In *Handbook of Applied Mathematics*, edited by C. E. PEARSON, pp. 878-927. New York: Van Nostrand Reinhold.
- CLEGG, W. (1981). *Acta Cryst.* **A37**, 913-915.
- DELAUNAY, B. (1933). *Z. Kristallogr.* **84**, 109-149.
- FERRARIS, G. & IVALDI, G. (1983). *Acta Cryst.* **A39**, 595-596.
- GRUBER, B. (1973). *Acta Cryst.* **A29**, 433-440.
- GRUBER, B. (1978). *Acta Cryst.* **A34**, S2.
- HIMES, V. L. & MIGHELL, A. D. (1982). *Acta Cryst.* **A38**, 748-749.
- HIMES, V. L. & MIGHELL, A. D. (1987). *Acta Cryst.* **A43**, 375-384.
- KATAYAMA, C. (1986). *J. Appl. Cryst.* **19**, 69-72.
- KNUTH, D. E. (1973). *The Art of Computer Programming*, Vol. 1: *Fundamental Algorithms*, pp. 4-6. Reading, MA: Addison Wesley.
- KŘIVÝ, I. & GRUBER, B. (1976). *Acta Cryst.* **A32**, 297-298.
- LAWTON, S. L. & JACOBSON, R. A. (1965). *The Reduced Cell and its Crystallographic Applications*. Report IS-1141, Ames Laboratory, Ames, Iowa 50010, USA.
- LE PAGE, Y. (1982). *J. Appl. Cryst.* **15**, 255-259.
- MIGHELL, A. D. (1976). *J. Appl. Cryst.* **9**, 491-498.
- MIGHELL, A. D. & HIMES, V. L. (1986). *Acta Cryst.* **A42**, 101-105.
- MIGHELL, A. D. & RODGERS, J. R. (1980). *Acta Cryst.* **A36**, 321-326.
- NIGGLI, P. (1928). *Handbuch der Experimentalphysik*, Vol. 7, part 1, pp. 108-176. Leipzig: Akademische Verlagsgesellschaft.
- ROOF, R. B. JR (1969). *A Theoretical Extension of the Reduced-Cell Concept in Crystallography*. Los Alamos Report LA-4038, Clearinghouse for Federal Scientific and Technical Information, National Bureau of Standards, US Dept of Commerce, Springfield, VA, USA.
- ZIMMERMANN, H. & BURZLAFF, H. (1985). *Z. Kristallogr.* **170**, 241-246.

problems. A simple weighting scheme and a Bragg-angle cut-off procedure are presented which alleviate these problems.

### 1. Introduction

In direct methods difficulties can be experienced when solving structures in situations where data of high resolution are being used early in the phasing process. There are two reasons for this:

(1) Use of high-angle reflections in this way has the effect of attempting to build very precise atomic detail in real space before the molecular outline has been fully defined and fixed by the lower-order reflections. We are currently carrying out research into the maximum entropy (ME) method as a technique for *ab initio* phasing of small molecules. Although there are highly significant differences, at one level this formalism can be considered as a real-space equivalent of traditional reciprocal-space direct methods. In the ME method it is important to work in resolution shells starting with low-resolution data; any attempt to use high-angle reflections early in the process builds atomic resolution into the method too early and causes instability. A similar phenomenon occurs in reciprocal space with traditional methods, often manifesting itself as a single very large peak in the final  $E$  map(s).

(2) The  $E$  magnitude ( $|E_{\mathbf{h}}|$ ) used in direct methods is defined as

$$E_{\mathbf{h}}^2 = kF_{\mathbf{h}}^2/\varepsilon \sum_{j=1}^N f_j^2 \exp[-2B(\sin^2 \theta)/\lambda^2] \quad (1)$$

where  $k$  is a scale factor,  $\varepsilon$  a correction for point-group symmetry,  $F_{\mathbf{h}}$  is the observed structure factor,  $f_j$  the scattering factor for atom  $j$ ,  $N$  is the number of atoms in the unit cell, and  $B$  is the overall temperature factor, assumed isotropic.

The normalization procedure usually estimates  $k$  and  $B$  via a Wilson plot (Wilson, 1949). This method gives only approximate values of these parameters. Whereas the error in  $k$  is uniformly distributed throughout the  $E$  magnitudes independently of Bragg angle, any error in  $B$  will give rise to an error which increases with angle. This can be seen clearly when standard deviations are calculated for the  $E$  magnitudes. The alternative  $K$ -curve method (Karle & Hauptman, 1953) will also give  $E$  magnitudes whose variance increases with angle because the average structure factor has a higher variance at high angle.

Direct methods use the three-phase structure invariants

$$\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}} = \Phi_3 \quad (2)$$

and the four-phase structure invariants (quartets):

$$\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{-\mathbf{h}-\mathbf{k}-\mathbf{l}} = \Phi_4 \quad (3)$$

where  $\varphi_{\mathbf{h}}$  is the phase angle of reflection  $\mathbf{h}$ ,  $\Phi_3$  is

usually assumed to be distributed around zero, whilst for the quartets the distribution of  $\Phi_4$  depends on the cross terms  $\mathbf{h} + \mathbf{k}$ ,  $\mathbf{k} + \mathbf{l}$  and  $\mathbf{l} + \mathbf{h}$  (Schenk, 1973).

The usual measure of triplet reliability is  $\kappa_{\mathbf{h}\mathbf{k}}$  defined as

$$\kappa_{\mathbf{h}\mathbf{k}} = 2|E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|/N^{1/2}. \quad (4)$$

This is used either directly or in modified form in all direct-methods programs throughout the phasing process. Any error in the  $E$  magnitudes will generate a corresponding error in  $\kappa_{\mathbf{h}\mathbf{k}}$ , and this will increase with the number of high-angle reflections involved in the invariant, so that relationships involving two or more high-resolution reflections are inherently more unreliable than those involving lower-order reflections. The  $\kappa_{\mathbf{h}\mathbf{k}}$  value will not reflect this. A similar argument applies to the quartets.

Problems also exist with very low-angle reflections. They are prone to error arising from solvent effects, extinction and measurement errors, so that the invariants involving them are similarly subject to error.

There are methods for estimating  $\cos \Phi_3$ , usually via a quintet extension (Hauptman, 1972, 1985; Gilmore & Hauptman, 1985), quadrupoles (Viterbo & Woolfson, 1973), representation theory (Giacovazzo, 1976, 1977; Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori, 1984) or determinantal methods (Messenger & Tsoucaris, 1972; Karle, 1979, 1980). All these methods have many features in common: they need empirical scale factors, they are very approximate and slow to compute, and they do not treat the fundamental difficulties described above. We describe here a method which avoids all these problems, and is more reliable.

### 2. Filtering structure invariants

Three filtering methods have been used:

(1) A notch filter consisting of defining a maximum value of  $c = (\sin^2 \theta)/\lambda^2$ . Any triplets or quartets which involve two or more reflections above this limit are excluded from the phasing process (both convergence mapping and tangent refinement). The value of  $c$  is user adjustable but a value corresponding to the Cu sphere for high-resolution data sets or  $0.95[(\sin^2 \theta)/\lambda^2]_{\max}$  in other cases is a suitable default. A similar filter can be applied to very low-order reflections.

(2) A weighting scheme in which each  $E$  magnitude is given an associated weight  $w_{\mathbf{h}}$ , where

$$w_{\mathbf{h}} = 1.0 \quad (\sin^2 \theta)/\lambda^2 \leq c$$

$$w_{\mathbf{h}} = \{(\sin^2 \theta)/\lambda^2 - [(\sin^2 \theta)/\lambda^2]_{\max}\}^{-1} \\ \times \{[c - (\sin^2 \theta)/\lambda^2]_{\max}\}^{-1} \\ (\sin^2 \theta)/\lambda^2 > c.$$

Table 1. *Crystallographic data for the seven structures selected for study*

Structure	Space group	Molecular formula	Z	Reference
AZET	<i>Pca</i> 2 <sub>1</sub>	C <sub>21</sub> H <sub>16</sub> ClNO	8	Colens, Declercq, Germain, Putzeys & Van Meerssche (1974)
PGE2	<i>P</i> 1	C <sub>20</sub> H <sub>32</sub> O <sub>5</sub>	1	DeTitta, Langs, Edmonds & Duax (1980)
GOLDMAN2	<i>Cc</i>	C <sub>28</sub> H <sub>16</sub>	8	Iringarter, Reibel & Sheldrick (1981)
BED2	<i>I</i> 4	C <sub>26</sub> H <sub>26</sub> N <sub>4</sub> O <sub>2</sub>	8	Sheldrick, Davison & Trotter (1978)
NO55	<i>Fdd</i> 2	C <sub>20</sub> H <sub>24</sub> N <sub>4</sub>	16	Sheldrick & Trotter (1978)
APAPA	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	C <sub>30</sub> H <sub>37</sub> N <sub>15</sub> O <sub>16</sub> P <sub>2</sub> ·6H <sub>2</sub> O	8	Suck, Manor & Saenger (1976)
TVAL	<i>P</i> 1	C <sub>54</sub> H <sub>90</sub> N <sub>6</sub> O <sub>18</sub>	2	Karle (1975)
WINTER2	<i>P</i> 2 <sub>1</sub>	C <sub>52</sub> H <sub>83</sub> N <sub>11</sub> O <sub>16</sub> ·3CH <sub>2</sub> Cl <sub>2</sub>	2	Butters, Hütter, Jung, Pauls, Schmitt, Sheldrick & Winter (1981)

Table 2. *Figures of merit and the number of atoms found for solutions which yield the best E map*

Missing values for PGE2 and TVAL arise from a lack of suitable relationships to give reliable figures of merit.

Structure	ABSFOM	Psi-zero	Resid	Nqest	Number of atoms/total	Rank of solution
AZET	1.84	3.23	49.0	-0.07	16/48	1
PGE2	1.33	-	25.3	-0.09	12/25	1
GOLDMAN2	1.15	2.10	28.3	-0.05	17/56	2
BED	1.12	1.19	23.0	-0.16	19/38	2
NO55	-	-	-	-	0/24	None
APAPA	1.07	1.16	26.7	-0.09	26/63	1
TVAL	1.25	-	22.8	-	80/156	1

Table 3. *Figures of merit and the number of atoms found when using the cut-off method or the weighting scheme*

Structure	ABSFOM	Psi-zero	Resid	Nqest	Number of atoms/total	Rank of solution
AZET	1.58	2.71	37.9	-0.06	31/48	1
PGE2	1.33	-	25.3	-0.08	15/25	1
GOLDMAN2*	1.14	0.86	18.8	0.01	56/56	1
BED	1.14	1.18	22.7	-0.17	21/38	2
NO55	0.66	1.70	28.3	-0.03	21/24	1
APAPA	1.09	1.10	25.1	-0.05	20/63	1
TVAL*	1.28	-	23.1	-	80/156	1

\* Cut-off method used.

The constant  $c$  is defined above. This function is shown diagrammatically in Fig. 1. The weight is then incorporated in the  $k_{hk}$  value as

$$k_{hk} = 2w_h w_k w_{-h-k} |E_h E_k E_{-h-k}| / N^{1/2}$$

and this  $k_{hk}$  is used throughout convergence mapping, tangent formula refinement and any other method that utilizes  $k_{hk}$ 's.

(3) A step filter which removes very low-angle reflections.

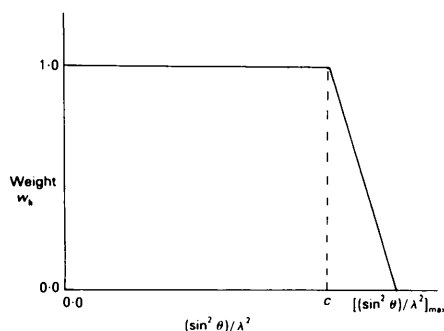


Fig. 1. The weighting scheme. Dashed line shows the cut-off scheme.

All these options are available in a new version of the *MITHRIL* program (Gilmore, 1984).

### 3. Application to known and unknown structures

Both methods were applied to seven structures in the database of difficult structures compiled by Sheldrick. Table 1 tabulates the salient feature of these structures. Table 2 tabulates the results of the best solutions which could be found using *MITHRIL*. In all cases but one a solution was possible, but sometimes difficult to detect - in particular AZET and GOLDMAN2 - and often the figures of merit were sufficiently poor to discourage the calculation and investigation of the corresponding *E* map. The application of cut-offs and weighting schemes is listed in Table 3. All the structures are now readily solvable, although in the case of APAPA there is a small deterioration in *E*-map quality. The figures of merit are also better behaved, giving values which are more in line with those expected for correct solutions, and thus encouraging the investigation of the corresponding *E* maps.

The method has also been applied to a hitherto unsolved structure. The polypeptide MORNON has

the formula  $C_{91}H_{158}N_{23}O_{24}$  and crystallizes in space group  $C222_1$  with  $Z = 8$ . There are 138 non-hydrogen atoms in the asymmetric unit and the data extend to 0.8 Å resolution. Exhaustive trials with various direct-methods techniques were unsuccessful. The weighting scheme was used for both triplets and negative quartets; magic-integer phase permutation was used with 11 reflections in the starting set. A total of 640 phase sets was generated, and the best solution yielded an  $E$  map in which 52 atoms were clearly visible. The structure was completed by standard Fourier techniques.

The filtering of very low-order reflections is, in general, much less successful, largely because we are trying to encourage direct methods to build a molecular shape as early as possible. However, in the case of WINTER2 (Table 1), removing all reflections with  $(\sin^2 \theta)/\lambda^2 < 0.016$  produced an  $E$  map in which 50 out of 88 non-hydrogen atoms were located. It is worth noting that this is the only structure tested that has a disordered solvent of appreciable scattering power. This type of filter could be of general applicability in these circumstances.

#### 4. Concluding remarks

This method is simple to apply and can be very effective. It requires no extra computing time except in cases when the scheme filters so many reflections that the convergence map becomes fragmented with many gaps. In these circumstances a larger starting set is necessary with a corresponding increase in computer time. There is no clear distinction between the efficacy of the weighting scheme or the cut-off method – sometimes one technique works and sometimes the other, as is the nature of direct methods.

*Acta Cryst.* (1988). **A44**, 1021–1028

## Analytical X-ray Line Profile Analysis Based Upon Correlated Dislocations

BY SATISH RAO AND C. R. HOUSKA

*Department of Materials Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*

(Received 8 February 1988; accepted 31 May 1988)

### Abstract

Recent advances describing X-ray line profiles analytically, in terms of a minimum number of parameters, are related to a theory based upon correlated dislocations. It is shown that a multiple convolution approach, based upon the Warren–Averbach (W–A) analysis, leads to a form that closely approximates

the strain coefficient obtained by Krivoglaz, Martynenko & Ryaboshopka [*Phys. Met. Metall.* (1983), **55**, 1–12]. This connection enables one to determine the dislocation density and the ratio of the correlation range parameter to the mean particle size. These two results are obtained most accurately from previous analytical approaches which make use of a statistical least-squares analysis. The W–A Fourier

### References

- BUTTERS, T., HÜTTER, P., JUNG, G., PAULS, P., SCHMITT, H., SHELDRIK, G. M. & WINTER, W. (1981). *Angew. Chem.* **93**, 904–905.
- CASCARANO, G., GIACOVAZZO, G., CAMALLI, M., SPAGNA, R., BURLA, M. C., NUNZI, A. & POLIDORI, G. (1984). *Acta Cryst.* **A40**, 278–283.
- COLENS, A., DECLERCO, J. P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.
- DETTA, G. T., LANGS, D. A., EDMONDS, J. W. & DUAX, W. L. (1980). *Acta Cryst.* **B36**, 638–645.
- GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 967–976.
- GIACOVAZZO, C. (1977). *Acta Cryst.* **A33**, 527–531.
- GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42–46.
- GILMORE, C. J. & HAUPTMAN, H. A. (1985). *Acta Cryst.* **A41**, 457–462.
- HAUPTMAN, H. A. (1972). *Crystal Structure Analysis: the Role of the Cosine Seminvariants*. New York: Plenum Press.
- HAUPTMAN, H. A. (1985). *Acta Cryst.* **A41**, 454–456.
- IRNGARTINGER, H., REIBEL, W. R. K. & SHELDRIK, G. M. (1981). *Acta Cryst.* **B34**, 1387–1389.
- KARLE, I. L. (1975). *J. Am. Chem. Soc.* **97**, 4379–4386.
- KARLE, J. (1979). *Proc. Natl Acad. Sci. USA*, **76**, 2089–2093.
- KARLE, J. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 5–9.
- KARLE, J. & HAUPTMAN, H. A. (1953). *Acta Cryst.* **6**, 473–476.
- MESSAGER, J. C. & TSOUCARIS, G. (1972). *Acta Cryst.* **A28**, 482–484.
- SCHENK, H. (1973). *Acta Cryst.* **A29**, 77–82.
- SHELDRIK, G. M., DAVISON, B. E. & TROTTER, J. (1978). *Acta Cryst.* **B34**, 1387–1389.
- SHELDRIK, G. M. & TROTTER, J. (1978). *Acta Cryst.* **B34**, 3122–3124.
- SUCK, D., MANOR, P. C. & SAENGER, W. (1976). *Acta Cryst.* **B32**, 1727–1735.
- VITERBO, D. & WOOLFSON, M. M. (1973). *Acta Cryst.* **A29**, 205–208.
- WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.