

used in another case were the original functions with exponents multiplied by 0.85 and 1.5. They result in agreement factors of 0.0046 and 0.0048, respectively, again confirming that the application of this algorithm is not highly dependent upon having a good basis. These results indicate there is a possibility that the solution will depend upon the initial guess matrix, but there is not an infinity of solutions as would be expected in a typical unconstrained problem. To eliminate any possible arbitrariness, we suggest using a free variational calculation, either Hartree–Fock or some approximation to Hartree–Fock as an initial guess. This idea is an extension of the work of Henderson & Zimmerman (1976) which indicated the existence of continuous, noncrossing, idempotent, energy-property hypersurfaces. This means that if one starts with an idempotent matrix of lowest possible energy and changes its fit to X-ray scattering data in a slow quasi-continuous way, the result will be closest to the original matrix. To investigate a problem for which the exact solution is not idempotent, the spherical atomic scattering factors for hydrogen calculated by Stewart *et al.* (1965) were used to construct a set of molecular scattering factors at an interatomic distance of 0.81×1.4009 atomic units. Scattering factors for nineteen values of $\sin \theta/\lambda$ from 0.0215 to 0.6011 \AA^{-1} were used in four planes whose angles with the internuclear bond were 0, 20, 50 and 90° , to make a total of seventy-six scattering factors. Huzinaga's (1965) Gaussian expansions of STO's with variable exponential factors were used to expand the density matrix. In accord with previous results, good fits to the density were found to be not very sensitive to the details of the basis

functions. Limited experimentation with three atomic basis functions generally refined to $R = 0.023$ regardless of exponents or type (1s, 2s, 3s). Presented in Fig. 3 is a typical result.

We thank Barbara Rubensohn for constructing Fig. 1, and Eric Unhjem and Toni Pezone for constructing Figs. 2 and 3, and CUNY-FRAP for grant support.

References

- CLINTON, W. L., FRISHBERG, C. A., MASSA, L. J. & OLDFIELD, P. A. (1973). *Int. J. Quantum Chem. Symp.* **7**, 505–514.
- CLINTON, W. L., FRISHBERG, C., MASSA, L. J. & OLDFIELD, P. A. (1981). *J. Chem. Phys.* Submitted.
- CLINTON, W. L. & MASSA, L. J. (1972). *Phys. Rev. Lett.* **20**, 1363–1366.
- COPPENS, P. (1975a). *International Reviews of Science: Phys. Chem. Ser. 2*, vol. 11, pp. 21–56. Baltimore: Univ. Park Press.
- COPPENS, P. (1975b). *Electronic Structure of Polymers and Molecular Crystals*, edited by J.-M. ANDRÉ & J. LADIK, pp. 227–257. New York: Plenum.
- COPPENS, P., WILLOUGHBY, T. V. & CSONKA, L. N. (1971). *Acta Cryst.* **A27**, 248–256.
- HENDERSON, G. A. & ZIMMERMAN, R. K. (1976). *J. Chem. Phys.* **65**, 619–622.
- HUZINAGA, S. (1965). *J. Chem. Phys.* **42**, 4, 1293–1302.
- MASSA, L. J. & CLINTON, W. L. (1972). *Trans. Am. Crystallogr. Assoc.* **8**, 149–153.
- MATTHEWS, D. A., STUCKY, G. D. & COPPENS, P. (1972). *J. Am. Chem. Soc.* **94**, 8001–8008.
- STEWART, R. F. (1969). *J. Chem. Phys.* **51**, 4569–4577.
- STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 9, 3175–3187.

Acta Cryst. (1982). **A38**, 98–102

On the Estimation of $|E|$ Values of the Squared Structure on the Basis of Quartet Invariants

BY N. VAN DER PUTTEN AND H. SCHENK

Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, 1018WV Amsterdam, The Netherlands

AND G. TSOUCARIS

Laboratoire de Physique, Centre Universitaire Pharmaceutique, 92290 Chatenay-Malabry, France

(Received 9 July 1980; accepted 2 July 1981)

Abstract

Expressions for the estimation of the $|G_H|$ values, the $|E_H|$ values of the squared structure, on the basis of all quartets in which H is a cross-term vector are presented

for the space groups $P1$ and $P\bar{1}$. A reliable estimation procedure was developed employing the quartets with highest quartet product only. It appears that in this way $|E|$ values of strong or weak reflections outside the limiting sphere could be predicted. An implication of

this is that for quartets with only two cross terms in the measured area of reciprocal space, the E 's of which are both either large or small, the $|E|$ of the third cross term is more likely to be large or small respectively. It is further shown that the estimated $|G|$ values sharpen the Patterson synthesis; however, at its present state this technique does not offer advantages over other sharpening procedures.

Introduction

Quartet phase sums

$$\varphi_4 = \varphi_H + \varphi_K + \varphi_L + \varphi_{-H-K-L}$$

can be estimated in a reliable way with the $|E|$ values of the reflections H, K, L and $-H-K-L$ and those of $H+K, H+L$ and $K+L$ (Schenk, 1973; Hauptman, 1975; Hauptman & Green, 1976; Gilmore, 1977; van der Putten & Schenk, 1979) and in particular extreme values of φ_4 may be predicted with the required accuracy for not too large structures. A quartet has a large probability to be positive ($\varphi_4 \approx 0$) if $E_4 = N^{-1}|E_H E_K E_L E_{H+K+L}|$ and the cross-term magnitudes $|E_{H+K}|, |E_{H+L}|, |E_{K+L}|$ are large (Schenk, 1973). If, on the other hand, E_4 is large and the three cross term $|E|$'s are small a quartet has a large probability to be negative: $\varphi_4 \approx \pi$ (Hauptman, 1974; Schenk, 1974).

This paper presents expressions for the estimation of $|G_H|$, the $|E|$ value of a reflection H for the squared structure. The expressions make use of all quartets for which H is a cross term and they show that if H occurs many times as a cross term in positive quartets, there is a large probability that $|G_H|$ is large. The reflection H may be either inside or outside the limiting sphere.

By application to five structures the relevance of the $|G|$ estimates for crystal structure determinations is evaluated. For two equal-term structures with $|E_H| \approx |G_H|$, the $|G|$ estimates are used in order to predict $|E|$ values outside the limiting sphere and for the other three, in which heavy atoms were present, the $|G|$'s were used to sharpen the Patterson synthesis.

Expressions for estimation of $|G_H|$ in the space groups $P1$ and $P\bar{1}$

For structures with non-equal atoms, Sayre's equation for normalized structure factors may be written as:

$$G_H = C_H \sum_K E_K E_{H-K}, \quad (1)$$

where G_H is the E value of reflection H of the squared structure and C_H is a positive constant which may be readily determined (see, for example, Cochran, 1955).

Multiplying both sides by their complex conjugates leads to

$$|G_H|^2 = C_H^2 \left| \sum_K E_K E_{H-K} \right|^2 \quad (2)$$

or

$$|G_H|^2 = C_H^2 \sum_K \sum_L |E_K E_{H-K} E_{-L} E_{L-H}| \cos \varphi_4 \quad (3)$$

in which

$$K + (H - K) + (-L) + (L - H) = 0 \quad (4)$$

and

$$\varphi_4 = \varphi_K + \varphi_{H-K} + \varphi_{-L} + \varphi_{L-H}. \quad (5)$$

The $\cos \varphi_4$ value in (3) may be approximated by the expected value found from the joint probability distribution of φ_4 given the $|E|$ values of the seven reflections $H, K, H - K, -L, L - H, K - L$ and $H - K - L$ (Hauptman, 1975). In view of the fact that one of the aims of this paper is the estimation of $|G_H|$ outside the limiting sphere, for which $|E_H|$ is not available, (3) is here approximated by an analogous expression:

$$|G_H|^2 = C_H^2 \langle |E_K E_{H-K} E_{-L} E_{L-H}| (\cos \varphi_4 |6) \rangle_{K,L}, \quad (6)$$

where $(\cos \varphi_4 |6)$ is the expected value of $\cos \varphi_4$ given the $|E|$ values of the six reflections $K, H - K, -L, L - H, K - L$, and $H - K - L$.

As was pointed out by Giacovazzo in a number of papers (see, for example, Giacovazzo, 1976), six magnitude expressions for quartets in fact use $|E|^2 = 1$ for the missing cross term, which implies that our practical approach to be described in the next section in fact starts from an estimate $|E_H|^2 = 1$ and ends with a better estimate for $|G_H|^2$. In analogy with Hauptman (1975), the joint probability distribution of φ_4 given six magnitudes is then

$$P(\varphi_4 |6) = C'' \exp(-B \cos \varphi_4) I_0(2|E_{K-L}| Z_{13}/N^{1/2}) \times I_0(2|E_{H-K-L}| Z_{23}/N^{1/2}), \quad (7)$$

where C'' is a suitable normalizing constant,

$$B = 2N^{-1}|E_K E_{H-K} E_{-L} E_{L-H}|, \quad (8)$$

$$Z_{13} = (E_K^2 E_L^2 + E_{H-K}^2 E_{L-H}^2 + 2|E_K E_{-L} E_{H-K} E_{L-H}| \cos \varphi_4)^{1/2}, \quad (9)$$

$$Z_{23} = (E_{H-K}^2 E_L^2 + E_K^2 E_{L-H}^2 + 2|E_K E_{-L} E_{H-K} E_{L-H}| \cos \varphi_4)^{1/2}, \quad (10)$$

and I_0 is the modified Bessel function of zeroth order.

In further analogy with Hauptman (1975) the conditional expected value of $\cos \varphi_4$, $(\cos \varphi_4 |6)$, is then

found from (7) as

$$(\cos \varphi_4 |6) = \frac{\sum_{\mu, \nu}^{\infty} (-1)^{\mu+\nu+1} I_{\mu\nu} I_{\mu+\nu+1}(B)}{\sum_{\mu, \nu}^{\infty} (-1)^{\mu+\nu} I_{\mu\nu} I_{\mu+\nu}(B)} \quad (11)$$

in which I is a modified Bessel function and $I_{\mu\nu}$ is defined by

$$\begin{aligned} I_{\mu\nu} = & I_{\mu}(2N^{-1/2}|E_{K-L} E_K E_{-L}|) \\ & \times I_{\nu}(2N^{-1/2}|E_{K-L} E_{H-K} E_{L-H}|) \\ & \times I_{\nu}(2N^{-1/2}|E_{H-K-L} E_{H-K} E_{-L}|) \\ & \times I_{\nu}(2N^{-1/2}|E_{H-K-L} E_K E_{L-H}|). \end{aligned} \quad (12)$$

Of course $(\cos \varphi_4 |6)$ may be estimated from (11), but it takes less computing time to calculate

$$(\cos \varphi_4 |6) \simeq \int_0^{\pi} [P(\varphi_4 |6) \cos \varphi_4 d\varphi_4]. \quad (13)$$

A numerical approximation of (13), involving a summation at five $|\varphi_4|$ values has been used throughout our calculations.

In space group $P\bar{1}$ (6) reduces to

$$|G_H|^2 = C_H'^2 \langle |E_K E_{H-K} E_{-L} E_{L-H}| S \rangle_{K,L}, \quad (14)$$

in which $S (= \pm 1)$, the sign of the quartet, may be estimated with P_6^{\pm} (Giacovazzo, 1976):

$$\begin{aligned} P_6^{\pm} = & C'' \exp(\mp B/2) \\ & \times \cosh[|E_{H-L}| N^{-1/2} (|E_K E_{-L}| \pm |E_{H-K} E_{L-H}|)] \\ & \times \cosh[|E_{H-K-L}| N^{-1/2} (|E_{H-K} E_{-L}| \pm |E_K E_{L-H}|)]. \end{aligned} \quad (15)$$

A better expression including weights for the individual terms is given by

$$|G_H|^2 = C_H'''^2 \langle |E_K E_{H-K} E_{-L} E_{L-H}| S / \text{VAR} \rangle_{K,L}, \quad (16)$$

with the variance VAR defined by

$$\text{VAR} = \frac{4P_6^+ P_6^-}{(P_6^+ + P_6^-)^2}.$$

Numerical estimation of $|G|^2$ values

In our experiments (7) and (15) were used rather than the corresponding formulae for the unequal-atom case, derived by Hauptman (1976) and Hauptman & Green (1976). The latter expressions, which involve the terms $(3\sigma_3^2 - \sigma_2 \sigma_4) / \sigma_3^3$ and $\sigma_3 / \sigma_2^{3/2}$ instead of $2/N$ and $1/N^{1/2}$ respectively ($\sigma_n = \sum_{j=1}^N f_j^n$), appear to overestimate negative values of $\cos \varphi_4$ for structures with

heavy atoms. In testing (6) and (14) we made the following observations:

(1) A given reflection may occur as a cross term in a very large number of quartets depending on the length of its reciprocal vector. For a structure with 4000 reflections this number may be of the order of 1 000 000. Thus the estimation of one $|G|$ value will take large amounts of computing time, if all quartets are employed.

(2) The estimate of $\cos \varphi_4$ in (6) and of S in (14) is unreliable if either B , defined by (8), is small or the $|E|$ values of one or two cross terms $K-L$ and $H-K-L$ are unknown. Of course the contributions of those quartets to (6) and (14) are small in most cases.

In view of this we have developed the following procedure. To start with, 10 000 to 50 000 quartets with largest E_4 values are generated for the 500 to 1000 strongest reflections. Then in the centrosymmetric case for each H the sum

$$\text{SUMC} = \sum_K \sum_L |E_K E_{H-K} E_{-L} E_{L-H}| S / \text{VAR}, \quad (17)$$

and in the non-centrosymmetric case the sum

$$\text{SUMN} = \sum_K \sum_L |E_K E_{H-K} E_{-L} E_{L-H}| (\cos \varphi_4 |6), \quad (18)$$

is calculated. Subsequently, those sums are selected for further calculations which include at least 20 quartets, as the others do not give reliable results. In order to correct for the variation in the number of quartets, in which a given reflection appears as a cross term, a scaling factor A can be applied (see the Appendix for its derivation and form). Because the summations (17) and (18) do not include all quartets, (16) and (6) reduce to

$$|G_H|^2 \propto A \times \text{SUMC} \quad (19)$$

and

$$|G_H|^2 \propto A \times \text{SUMN}. \quad (20)$$

From tests it is found that large values of (19) and (20) correspond to large $|G_H|^2$ values and likewise for small values. The latter, however, are less reliable than large $|G_H|$ estimates and are very scarce, because the requirement that at least 20 quartets are used in the summations (17) and (18) is seldom fulfilled. In order to arrive at estimated $|G_H|^2$ values the data obtained from (19) and (20) have to be scaled. This is done from the distribution of the known $|E|$ values as a function of $\sin \theta / \lambda$ by the following procedure.

A list is made of the reflections, arranged in decreasing order of SUMC or SUMN. This list is divided into groups of 100–300 reflections. Then the known $|E|$ values of each group are divided into four equivalent subgroups, according to their $\sin \theta / \lambda$ value. For each of the subgroups the $\langle |E| \rangle$ and $\langle \sin \theta / \lambda \rangle$ are calculated. Finally, these values are plotted and a

straight line is fitted. The lines provides an estimate for $|G_H|$ at a given value of $\sin(\theta_H)/\lambda$.

Calculating $|E|$'s outside the limiting sphere

For equal-atom structures $|G_H| = |E_H|$ and thus $|E|$ values of reflections outside the limiting sphere can be estimated. This application has been tested with two structures: prostaglandin E_2 (PGE₂), C₂₀H₃₂O₅, $P1$, $Z = 1$, $N = 25$ (Edmonds & Duax, 1974*a,b*) and monomorphine, C₁₉H₂₇N₄O₇, $P1$, $Z = 4$, $N = 120$ (Overbeek, unpublished).

For the 1554 measured reflections of PGE₂ $|E|$'s were calculated and about 20 000 quartets were generated for the 500 reflections with largest $|E|$. On the basis of the quartets 99 $|E|$'s of reflections outside the limiting sphere were estimated to be at least 1.5. A comparison of these estimates with the calculated $|E|$'s showed that 62 out of the 99 are larger than 1.5, 11 have an $|E|$ between 1.5 and 1.2, 13 an $|E|$ between 1.2 and 0.8 and only three have an $|E|$ lower than 0.8. In the set of 1554 measured reflections only 63 have an $|E|$ value larger than 1.5 and thus there is a great extension of the number of strong reflections for this difficult structure. No small $|E|$ values could be estimated because of the requirements that SUMN (18) is built up from at least 20 contributors and $SUMN \approx 0$. Relaxing the first requirement to three contributors, 14 $|E|$ values of reflections outside the limiting sphere could be indicated to be small. For these reflections the actual average $|E|^2$ value was 0.67.

The above calculations were repeated using the measured $|E|$ values instead of the calculated ones. These experiments showed similar results.

For the measured 3981 reflections of monomorphine the $|E|$ values were calculated and about 27 000 quartets were generated for the 1000 strongest reflections. With the quartets, 66 $|E|$ values of reflections outside the Ewald sphere were estimated to be 2.0 or higher, and a comparison with their calculated $|E|$ values showed that 31 were above 2.0, 14 were between 2.0 and 1.5, 10 between 1.5 and 1.0 and the other 11 smaller than 1.0. Again it appears that most of the strong $|E|$ estimates correspond to strong $|E|$ values. With the requirements that the number of contributors to SUMC is at least 3 and $SUMC \approx 0$, 176 $|E|$ values of reflections outside the limiting sphere could be indicated to be small. Table 1 shows that the smaller the SUMC value, the more reliable the estimated $|E|$ value.

Again experiments using the actual $|E|$ values instead of the calculated ones show similar results.

From the two examples it may be concluded that a number of large $|E|$'s outside the limiting sphere can be estimated. In fact, this applies also to the weak $|E|$'s. In the quartet theories $|E|^2$ estimates of 1 are used in case

Table 1. Average actual $|E|^2$ value for monomorphine of the number (NR) of reflections outside the limiting sphere with the smallest $A \times SUMS$'s (19)

NR	$\langle E _{\text{real}}^2 \rangle_{NR}$
10	0.34
20	0.37
30	0.50
40	0.57
50	0.67
75	0.79
100	0.78
125	0.79
150	0.86
175	0.94

a reflection outside the limiting sphere occurs in a quartet as a cross vector, thus eliminating the effect of the cross term on the probability distribution of the quartet. Our examples show that this leads to results which are less accurate than necessary, since, for example, a quartet with two large cross $|E|$'s has a sizeable probability to have the third cross $|E|$ large also. A better procedure should therefore include an estimation of the strong $|E|$ values of reflections outside the limiting sphere, before the phase sums of the individual quartets are estimated on the basis of main and cross terms. Of course, the above results cannot imply that the technique described improves the resolution; this is only achieved by measuring data for larger reciprocal vectors. On the other hand, our technique enables a reorganization of the structure-factor information and thus of the phase relationships; it acts as a sort of filter for useful information.

Sharpening the Patterson function

For non-equal atom structures the $|G|$ values were used as coefficients in a Patterson synthesis in order to sharpen the peaks of the heavy-atom-heavy-atom structures. This application will be described for three structures: a sulphur steroid (SULSTE), C₂₀H₂₄O₅S, $P2_1/c$, $Z = 8$, $N = 204$ (Kops, Bode & Schenk, 1977); norgalanthamine chlorohydrate (NORG), C₁₆H₂₀NO₃Cl, $P1$, $Z = 2$, $N = 40$, (Germain, unpublished); and a randomly generated structure C₉₉Br (RANBRO), $P2_1$, $Z = 2$, $N = 200$.

In general it can be stated that Patterson maps calculated with $|G_H|^2$ values preserve the general features of the correct maps; the sharpening effects will be dealt with separately for each structure.

For the 500 strongest reflections of SULSTE 25 000 quartets were generated, and these were used to estimate 767 $|G|$ values with a lower limit of 0.8. Then the Harker section $y = 0.500$ was calculated with the $|G|$ values. In this section two S-S peaks are present, one is the strongest peak in the section, the other, how-

ever, is peak number 16 in decreasing order of height. For comparison four other Patterson syntheses were calculated with 2403 $|F|^2$ values, 1727 $|F^2|$ values with $\sin \theta/\lambda > 0.3$, 2403 $|E|^2$ values and 767 $(|G| + |E|)^2/4$ values respectively and here the two S-S Harker vectors were represented by the peaks numbered 2 and 14, 1 and 3, 2 and 5 and 1 and 11 in decreasing order of height respectively.

For the structure NORG, 1445 $|G|$ values with $|G| > 0.8$ were estimated with 32 000 quartets. The Cl-Cl peak was the third highest in the Patterson function based on these $|G|$ values. Other sharpening techniques such as the use of $|F|$'s with $\sin \theta/\lambda > 0.3$ or the use of $|E|$'s revealed the Cl-Cl peak as the highest maximum.

The third experiment was carried out with the generated $P2_1$ structure RANBRO ($C_{99}Br$). On the basis of 26 000 quartets 656 $|G|$'s were estimated and the Harker section $y = 0.5$ showed the Br-Br peak at a height of 625, whilst the second peak was 350. For comparison two maps were calculated with the calculated $|E|$'s and $|G|$'s, in which the Br-Br vectors were 350 and 790 respectively, whereas in both cases the second highest peak was 150. Again the use of $|G|$'s certainly sharpens the Patterson synthesis; however, the effect is not dramatically better in terms of the signal-to-noise ratio.

The results of these three structures show that sharpening the Patterson function with $|G|$ estimates is possible, but it does not offer great advantages over other sharpening techniques. Perhaps the procedure for the estimation of $|G|$ values may be improved by:

(1) The use of formulae for the estimation of $\cos \theta_4$ and S , employing the seven magnitude expressions (Hauptman, 1975; Hauptman & Green, 1976).

(2) The incorporation of a second cycle, in which the $|G|$ values estimated in the first cycle, are treated as known values.

(3) The summation over a much larger number of reliably estimated quartets in (17) and (18).

Of course computing time (now about 750 CPU s for one structure on a CYBER 73) will rise steeply.

The authors thank Professor B. O. Loopstra and Dr C. H. Stam for their helpful comments with respect to

the manuscript and Professor H. Hauptman for valuable discussions.

APPENDIX

The derivation of A in (19) and (20)

The theoretically possible number of triplets, for which the reflection H is one of the three terms, is proportional to

$$\frac{V_{\text{com}}}{V_{\text{lim}}} = [2 \sin(\theta_{\text{max}})/\lambda - \sin(\theta_H)/\lambda]^2 \\ \times (2\pi/3)[4 \sin(\theta_{\text{max}})/\lambda + \sin(\theta_H)/\lambda] \\ \times \{(4/3)\pi[2 \sin(\theta_{\text{max}})/\lambda]^3\}^{-1} \quad (A1)$$

in which V_{com} equals the overlapping volume of two limiting spheres of radius $2 \sin(\theta_{\text{max}})/\lambda$ whose centres have a separation equal to the distance, $2 \sin(\theta_H)/\lambda$, of the point H from the origin of reciprocal space. V_{lim} equals the volume of the limiting sphere.

The theoretically possible number of quartets, in which the reflection H occurs as a cross term, is proportional to:

$$A^{-1} = \frac{V_{\text{com}}^2}{V_{\text{lim}}^2} \quad (A2)$$

References

- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.
 EDMONDS, J. W. & DUAX, W. L. (1974a). *Prostaglandins*, **5**, 275-281.
 EDMONDS, J. W. & DUAX, W. L. (1974b). *Am. Crystallogr. Assoc. Spring Meeting, Univ. California, Abstract A5*.
 GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 958-966.
 GILMORE, C. (1977). *Acta Cryst.* **A33**, 712-716.
 HAUPTMAN, H. (1974). *Acta Cryst.* **A30**, 472-476.
 HAUPTMAN, H. (1975). *Acta Cryst.* **A31**, 680-687.
 HAUPTMAN, H. (1976). *Acta Cryst.* **A32**, 877-882.
 HAUPTMAN, H. & GREEN, E. A. (1976). *Acta Cryst.* **A32**, 45-49.
 KOPS, R. T., BODE, J. & SCHENK, H. (1977). *Cryst. Struct. Commun.* **6**, 823-827.
 PUTTEN, N. VAN DER & SCHENK, H. (1979). *Acta Cryst.* **A35**, 381-387.
 SCHENK, H. (1973). *Acta Cryst.* **A29**, 77-81.
 SCHENK, H. (1974). *Acta Cryst.* **A30**, 477-481.