and type of anomalously scattering atoms and can readily incorporate information from isomorphous replacement. If a heavy-atom structure is not known, it is still possible to evaluate triplet phase invariants (Karle, 1984*e*) from the exact algebraic analysis.

The algebraic analysis presented here involves somewhat different quantities. In its range of applicability, individual phase information would also be derivable from the present analysis if the heavy-atom structure were known since, then, $\varphi_{2,\mathbf{h}}^{n}$ in (24) and (25) would be known. The two kinds of algebraic analyses are expected to yield closely equivalent results in their areas of common applicability.

In order to apply the results of this paper, which concern one predominant type of anomalous scatterer, it is only necessary to know the chemical identity of this type of anomalous scatterer.

I wish to thank Mr Stephen Brenner for writing the appropriate programs and making the computations reported here.

This work was supported in part by USPHS grant GM30902.

References

GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585-592.
HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 289-294, 632-641.
KARLE, J. (1980). *Int. J. Quantum Chem.* **7**, 357-367.
KARLE, J. (1983). *Acta Cryst.* **A39**, 800-805.
KARLE, J. (1984*a*). *Acta Cryst.* **A40**, 1-4.
KARLE, J. (1984*b*). *Acta Cryst.* **A40**, 4-11.
KARLE, J. (1984*c*). *Acta Cryst.* **A40**, 366-373.
KARLE, J. (1984*d*). *Acta Cryst.* **A40**, 374-379.
KARLE, J. (1984*e*). *Acta Cryst.* **A40**, 526-531.
PONTENAGEL, W. M. G. F., KRABBENDAM, H., PEERDEMAN, A. F. & KROON, J. (1983). Eighth Eur. Crystallogr. Meet., 8-12 August, Abstract 4.01-P, p. 257.
TIMKOVITCH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **251**, 4033-4046.

Acta Cryst. (1985). **A41**, 189-195

Intensities, Structure Factors and their Variances

BY W. GONSCHOREK

Institut für Kristallographie der RWTH Aachen, D-5100 Aachen, Federal Republic of Germany

(*Received* 30 *January* 1984; *accepted* 18 *October* 1984)

Abstract

Hypothetical variances are checked and improved and observations are tested for normal and independent distribution by means of χ^2 distributions. For the conversion of σ_I^2 into σ_F^2 and of I into F the quantity $X(I)=\operatorname{sign}(I)|I|^{1/2}$ instead of $F=pI^{1/2}$ ($p=$ Lorentz correction *etc.*) is considered. With the probability density function $f(I)$ for I the variance of F is

$$\sigma_F^2=p^2\left\{\int_{-\infty}^{+\infty} X^2(I)f(I)\,\mathrm{d}I-\left[\int_{-\infty}^{+\infty} X(I)f(I)\,\mathrm{d}I\right]^2\right\}.$$

This formula is also valid if the observed intensity is zero or negative. Unexpectedly, the approximate $\sigma'_F=p\sigma_I/(2I^{1/2})$ are in most cases smaller than the correct ones. The simple formula $F=pI^{1/2}$ yields F that in the mean are too low. The correction formula valid for all ratios I/σ_I is $F_{\text{corr}}=p(2I^{1/2}-A)$, where I is the observed intensity and A the second integral in the above formula for σ_F^2. Since the expressions σ_F^2 and F_{corr} are known, there is no other reason to discard weak intensities than the fact that this saves computing time. Examples are given and the normal probability plot is applied.

Introduction

Hypothetical variances σ_I^2 of integrated intensities I of the rotating-crystal method can be obtained by means of Poisson's formula:

$$I=IT-\alpha(BL+BH) \qquad (1)$$

$$\sigma_I^2=IT+\alpha^2(BL+BH). \qquad (2)$$

IT is the sum of all X-ray quanta recorded by the counter during the rotation of the crystal through the reflecting orientation. BL and BH are the low- and high-angle background counts and α is the ratio of the times that were spent for the measurements of IT and of $(BL+BH)$. It is not essential that I and σ_I^2 are obtained by as simple formulae as (1) and (2); the arguments given in the following apply also if more sophisticated measuring procedures and formulae for the determination of I and σ_I^2 are used as, for example, proposed by Shoemaker (1968), Blessing, Coppens & Becker (1974) or Lehmann (1975).

The necessity and procedures to correct pure Poisson variances have been discussed by many authors, *e.g.* Jeffery (1964), Jeffery & Rose (1964),

0108-7673/85/020189-07$01.50

Abrahams (1964), Schulz (1971),* Abrahams (1974), McCandlish, Stout & Andrews (1975), Gonschorek (1981). Dunitz (1979) is not convinced of the effectiveness of any of the proposed error estimates. Apparently he would accept repeated measurements as a mean 'to get some reasonable statistical estimates of the errors' but does not think this to be feasible in practice. Indeed, a statistical estimate of the variance σ^2 is the sample variance s^2 and this is obtained from *repeated* measurements. A practicable way of checking and modifying hypothetical variances by repeated measurements was proposed by Gonschorek (1982*a*) and examples are given by Alte da Veiga, Andrade & Gonschorek (1982), Gonschorek (1982*b*) and Gonschorek & Feld (1982). The data are also checked, whether they are normally and independently distributed or not. In the present paper, first an improvement of this procedure is described and in another section the formulae for the derivation of F and σ_F from I and σ_I are given.

Checking for normal and independent distribution and improving hypothetical variances

As measures of how well symmetry-equivalent intensities and intensities measured with different crystals agree with each other, the internal (R_i, wR_i) and mutual (R_{ij}, wR_{ij}) consistencies, respectively, have been used by Abrahams, Hamilton & Mathieson (1970).

The disadvantage of these quantities is that they do not give information about the statistical behaviour of the data such as variances or systematic errors. Further, the numerical values of R_i and R_{ij}, as those of the R factor in least-squares refinements, depend upon whether the data set consists primarily of strong intensities or if there are many weak reflections.

These disadvantages can be avoided if the following relationships are taken into account. Consider a sample of n_j observations x_{ji}, $i = 1, \ldots, n_j$, with the population mean $\hat{x}_j$ and the variances σ_{ji}^2. The weighted mean

$$\bar{x}_j = \sum_i^{n_j} w_i x_{ji} \tag{3}$$

and the quantity

$$s_j^2 = (n_j - 1)^{-1} \sum_i^{n_j} w_i (x_{ji} - \bar{x}_j) \tag{4}$$

with weights

$$w_i = \sigma_{ji}^{-2} \Big/ \sum_k^{n_j} \sigma_{jk}^{-2} \tag{5}$$

are taken. The variance of $\bar{x}_j$ is derived from (3):

$$\sigma_{\bar{x}_j}^2 = \sum_i^{n_j} w_i^2 \sigma_{ji}^2 = \left(\sum_i^{n_j} \sigma_{ji}^{-2} \right)^{-1}. \tag{6}$$

If the observations x_{ji} are normally and independently distributed, the quantity

$$\chi_j^2 = (n_j - 1) s_j^2 / \sigma_{\bar{x}_j}^2 \tag{7}$$

is χ^2 distributed with $\nu = (n_j - 1)$ degrees of freedom, *cf.* Gonschorek (1982*a*).* Instead of χ_j^2, the quantity

$$\lambda_j^2 = \chi_j^2 / (n_j - 1) = s_j^2 / \sigma_{\bar{x}_j}^2 \tag{8}$$

is considered. It has the population mean unity and the variance $2/(n_j - 1)$, *cf.* van der Waerden (1971), p. 94. If there are $j = 1, \ldots, n$ observables, for example intensities of different *hkl*, the weighted mean of all λ_j^2

$$Q = \overline{\lambda^2} = \sum_j^n w_j' \lambda_j^2 \tag{9}$$

with weights

$$w_j' = (n_j - 1) \Big/ \sum_k^n (n_k - 1) \tag{10}$$

has the population mean unity and the variance

$$\sigma_Q^2 = 2 \Big/ \sum_j^n (n_j - 1). \tag{11}$$

Also, the empirical variances s_Q^2 of Q can be evaluated:

$$s_Q^2 = (n - 1)^{-1} \sum_j^n w_j' (\lambda_j^2 - Q)^2. \tag{12}$$

The quantity Q, (9), taking (10), (8), (4), (5) and (6) into account can be rewritten:

$$Q = \sum_j^n \sum_i^{n_j} \sigma_{ji}^{-2} (x_{ji} - \bar{x}_j)^2 \Big/ \sum_j^n (n_j - 1). \tag{13}$$

Suppose now, out of a set of integrated intensities of a crystal, $n = 1000$ intensities have been measured twice, *i.e.* $n_j = 2$ for thousand j or thousand *hkl*! Hypothetical variances σ_{ji}^2 can be derived, assuming the intensities to be Poisson distributed, so that Q can be evaluated. If Q differs from unity by more than some fraction of its standard deviation σ_Q, (11), one may decide to modify the variances σ_{ji}^2. A simple modification as proposed by Gonschorek (1982*a*) is the multiplication by a constant factor v:

$$\sigma_{ji}'^2 = v \sigma_{ji}^2. \tag{14}$$

The factor v was determined by fitting the observed to the theoretical χ^2 distribution and this has the

---

* Schulz (1971) in his expression for the variance of the corrected intensity took into account a term representing the uncertainty of the scaling parameter deduced from control or standard reflections.

* On request, an English translation of this paper can be obtained from the author.

disadvantage that the numbers n_j of repeated measurements must be the same for all n reflections *hkl*. Now, from (13) it is seen that $v = Q$ makes the new Q equal to unity and the numbers n_j need not be the same. Another modification of hypothetical variances uses the so called instability constant p, which was discussed in detail by McCandlish, Stout & Andrews (1975). This modification is equivalent to

$$\sigma'^2_{ji} = \sigma^2_{ji} + p\bar{x}^2_j. \qquad (15)$$

Again, the factor p can be determined *via* (13) in a few steps regarding Q as a function of p and demanding $Q = 1$.

As soon as the variances are modified in (13), so that $Q = 1$ is valid, the quantity Q loses its capability to give information about the normal and independent distribution of the data. Therefore, instead of Q the ratio

$$r = s^2_Q / \sigma^2_Q \qquad (16)$$

with s^2_Q from (12) and σ^2_Q from (11) may be used for this purpose. The closer r comes to unity, the surer the data can be assumed to be normally and independently distributed.

Systematic errors

Systematic errors, since they are not random, cannot be treated statistically but they may be revealed by statistical methods. Abrahams (1969) gives a list of error sources separated as random and systematic errors and earlier Abrahams (1964) divides systematic errors into anisotropic and isotropic classes. Anisotropic systematic errors are detectable by comparison of symmetry-equivalent intensities, certain kinds of isotropic ones are detectable by comparison of data, which were measured under different conditions, *e.g.* different crystals, different scan ranges, monochromatized or β-filtered radiation.

To reveal systematic errors, data that are supposed to have the same mean are checked by the same criteria as repeatedly measured data; but now the variances must not be changed! The quantities Q and r, defined by (13) and (16), are both expected to come close to unity.

If their actual numerical values indicate systematic errors not to be negligible, one may apply a χ^2 test as described by Gonschorek (1982*a*) and reject those (equivalent data) whose χ^2 value is larger than is acceptable from a statistical point of view. Before the comparison, data measured under different conditions must be brought onto the same scale. This can be achieved by a procedure described by Hamilton, Rollett & Sparks (1965) and modified by Gonschorek (1982*b*).

If a set of data could successfully pass the described statistical examination, the goodness-of-fit parameter in a least-squares refinement should not be different from unity. If it is, then most probably (i) there are undetected systematic errors, or (ii) the model is wrong. Examples both for (i) and (ii) are given by Rollet (1984).

As an alternative mean, to check for systematic errors in structural crystallography, the Durbin–Watson d statistic has been proposed by Flack, Vincent & Vincent (1980), *cf.* also Flack (1984).

Converting I into F and σ^2_I into σ^2_F

It is now assumed that the integrated intensities I are normally and independently distributed with variances σ^2_I and that there are no systematic errors. If I is normally distributed, then $|F| = pI^{1/2}$ is not normally distributed and in a least-squares refinement based on $|F|$ the expectation value of the goodness-of-fit parameter is in principle unknown. One may therefore prefer to refine on I and knowledge of σ^2_F is not necessary. Refinement on I is compared with that on $|F|$ and F by Wilson (1976) and under certain conditions it may be desirable to refine on $|F|$. Further, for the estimation of errors in electron or deformation densities as discussed by Rees (1976, 1978), the structure factor variances must be known. In the following again F instead of $|F|$ is written. For the derivation of the variance σ^2_F, instead of F the quantity $X(I) = \operatorname{sign}(I)|I|^{1/2}$ is considered. The variance of $X(I)$ is

$$\sigma^2_X = \int_{-\infty}^{+\infty} X^2(I) f(I)\,\mathrm{d}I - \left[\int_{-\infty}^{+\infty} X(I) f(I)\,\mathrm{d}I\right]^2 \qquad (17)$$

with $f(I) = (\sigma^2_I 2\pi)^{-1/2}\exp[-(I-\hat{I})^2/2\sigma^2_I]$, so that

$$\sigma^2_F = p^2\sigma^2_X \qquad (18)$$

is found. Here $\hat{I}$ is the population mean, whereas $\bar{I}$ is the sample mean. A similar equation to (17) was used by Rees (1976), equation (4). But Rees integrates from zero to infinity and does not consider negative values of I, although I can become negative. Suppose a reflection is systematically absent [$\hat{I} = 0$ in $f(I)$] and one wants to check by repeated measurements of I if this is the case or not; then I from (1) scatters about zero and the mean $\bar{I}$, (3), can be negative. In this case it is reasonable to put $\hat{I} = 0$ with the variance σ^2_I as defined in (6). Compare, however, Hirshfeld & Rabinowich (1973) and French & Wilson (1978).

The conversion of I into F was discussed by Ibers & Hamilton (1964), Rees (1977*b*) and Wilson (1979). The formula given by Wilson fails just in those cases where a correction of the simple equation (19) seems to be necessary, namely for small I/σ_I. The amounts by which the values of F are in the mean too low, if (19) is applied, were given by Rees (1977*b*). But his error estimates are too low compared with those obtained, if negative values of I are also taken into account as is seen by comparison of curve (*b*) in Fig. 2 with Fig. 4 of Rees (1977*b*).

For the evaluation of $F = pI^{1/2}$ the square root of the mean rather than the mean of the square roots of equivalent intensities should be taken. Even then, the observed F is too low, *cf.* equation (27) of Wilson (1979).

The 'true' $F = p\hat{I}^{1/2}$ is unknown. From experiment

$$F = pI^{1/2} \tag{19}$$

is found; its expectation value is

$$\hat{F} = pA, \tag{20}$$

where A is the second integral in (17). Hence, the expectation value of $F' = p[I^{1/2} + (\hat{I}^{1/2} - A)]$ is $p\hat{I}^{1/2}$, the 'true' F. Replacing $\hat{I}$ by I in this equation leads to the correction formula

$$F_{\text{corr}} = p(2I^{1/2} - A). \tag{21}$$

In the next section it is seen that the approximate structure factor variance

$$\sigma_F'^2 = p^2\sigma_I^2/4I \tag{22}$$

deviates from the correct one, (18), only for small I/σ_I as does F, (19), compared with F_{corr}, (21). One may therefore decide to discard all reflections with intensities less than some limit (*e.g.* $I < 3\sigma_I$) and to use (19) and (22). In this case, the *mean* of equivalent intensities must be checked if it is larger than the rejection threshold. If, otherwise, equivalent intensities are checked individually, some weaker ones may be discarded and the stronger retained so that the mean of the remaining intensities is too large; an example is given by Alte da Veiga, Andrade & Gonschorek (1982). Knowing (18) and (21), there is no other reason to discard weak intensities than the fact that this saves computing time. Note that (18) is valid also for $I = 0$, when (22) cannot be applied. In some cases, weak intensities must not be discarded, *e.g.* if a superlattice with many weak intensities is studied or if the errors of a difference density are estimated according to Rees (1976, 1977*a*, 1978). Since the variances of weak intensities are missing, the estimated errors are too low.

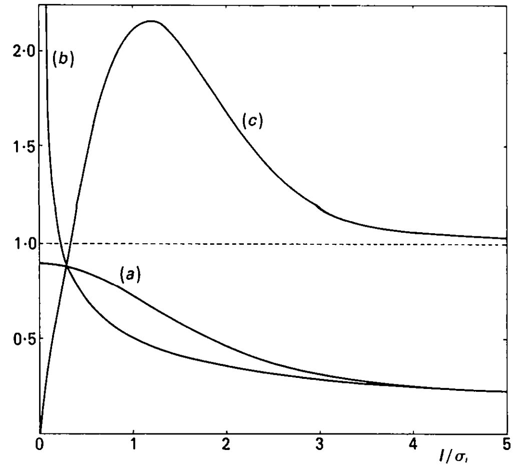


Fig. 1. The ratios $\sigma_F/\sigma_I^{1/2}$: (*a*) for correct σ_F derived *via* (17) and (18); (*b*) for approximate σ_F', (22). In curve (*c*) the ratio $(\sigma_F/\sigma_F')^2$ is plotted.

Numerical examples and discussion

For the numerical evaluation of the integrals in (17) the Gaussian quadrature formula with 26 grid points on either side of $\hat{I}$ was used. The integrations were extended to $\hat{I} \pm 6\sigma_I$. The results are represented in Figs. 1 and 2. Curve (*b*) in Fig. 1 is identical with curve (*b*) in Fig. 1 of Rees (1976, 1977*a*). The (*a*) curves correspond to each other but they are different, since Rees does not consider negative values of I. Curve (*c*) in Fig. 1 shows that weights $w \sim \sigma_F'^{-2}$ using approximate variances, (22), are too large for $I > 0\cdot3\sigma_I$ with a maximum error of a factor two for $I \simeq \sigma_I$ and of 18 and 1% for $I = 3\sigma_I$ and $10\sigma_I$, respectively. The variances $\sigma_F'^2$, if they are used for the error estimation of a Fourier transform, are too low by the same factor. Curve (*a*) in Fig. 2 shows that the maximum correction of F amounts to $0\cdot34\sigma_F$ and occurs at $I/\sigma_I = 0\cdot45$, curve (*b*) is given for comparison with the corresponding curve in Fig. 4 of Rees (1977*b*).

The result of the statistical treatment of three data sets is summarized in Table 1.* Two VO_2 crystals were used, one with dimensions 35 × 47 × 170 μm for measurements at 363 K in the tetragonal phase and one with dimensions 17 × 30 × 144 μm for measurements at 293 K in the monoclinic phase. The two data sets VO_2 mon. I and VO_2 mon. II were both measured with the second crystal in almost the same orientation and with the same CAD-4 diffractometer. The time interval between the two measurements was 15 months. The Ag $K\alpha$ radiation was monochromatized by a 50 μm Rh β filter for the first data set and by a

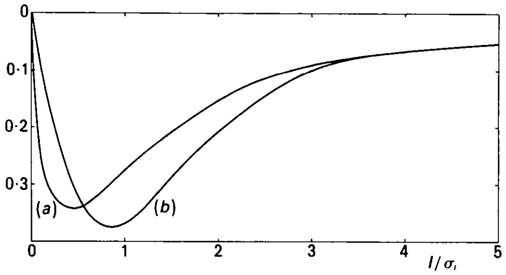


Fig. 2. The influence of the correction (21) on F: (*a*) $(F - F_{\text{corr}})/\sigma_F$, *cf.* (19) and (21); (*b*) $x_{\text{obs}}^{1/2} - \bar{x}_{\text{obs}}$, with $x_{\text{obs}} = I/\sigma_I$.

---

* For one more data set with 1472 twice measured intensities the factor $v = 1\cdot94$ and the values $Q = 159$, $\sigma_Q = 0\cdot04$, $s_Q = 137$, $R_i = 0\cdot02$ and $wR_i = 0\cdot09$ were found for the 1472 pairs of intensities. This complete data set is discarded.

Table 1. *Statistics of three data sets of integrated intensities*

Scan widths $\Delta\theta = D \tan\theta + E$. For symmetry-equivalent intensities n_j is between 2 and 16 (tetrag.) and 2 and 4 (mon.).

$$R_i = \sum_h \sum_e |I_{he} - \bar{I}_h| / \sum_h \sum_e |I_{he}|, \qquad wR_i = \left[\sum_h \sum_e \sigma_{he}^{-2} (I_{he} - \bar{I}_h)^2 / \sum_h \sum_e \sigma_{he}^{-2} I_{he}^2 \right]^{1/2},$$

$$R = \sum |F - |F_{\text{calc}}|| / \sum F, \qquad wR = [\sum \sigma_F^{-2} (F - |F_{\text{calc}}|)^2 / \sum \sigma_F^{-2} F^2]^{1/2},$$

$$S^2 = (n-r)^{-1} \sum \sigma_F^{-2} (F - |F_{\text{calc}}|)^2. \quad F \text{ stands for } F_{\text{obs}}.$$

The numbers $n(I \leq 0)$, $n(\sigma_F/\sigma'_F > 1{\cdot}1)$ and $n(\Delta F/\sigma F > 0{\cdot}2)$ give the numbers of unique intensities for which the relation given in brackets is valid [$\Delta F = F_{\text{corr}} - F$, *cf.* (21) and (19)]. The number $n(I \leq 0)$ does not include systematically absent reflections.

Crystal and scan parameters	Control reflections			Intensity statistics*								Refinement				
	$\sin\theta/\lambda$	I	v	n	p	v	Q	σ_Q	s_Q	R_i	wR_i	n	R	wR	S^2	σ_{S^2}
VO_2 tetrag.	0·16 Å^{-1}	1241	13·42	128	0	1·52	0·96	0·13	0·13	0·03	0·04					
$D = 0{\cdot}9°$, $E = 0{\cdot}33°$	0·47	295	4·26	136	0	1·52	3·81	0·04	1·53	0·04	0·05					
$\sin\theta/\lambda < 1{\cdot}0$ Å^{-1}	0·70	103	1·08	106	0	1·52	0·98	0·05	0·06	0·03	0·03	72	0·029	0·015	2·24	0·18
	Number of repetitions: 34			$n(I \leq 0) = 0$		$n(\sigma_F/\sigma'_F > 1{\cdot}1) = 1$			$n(\Delta F/\sigma_F > 0{\cdot}2) = 0$							
VO_2 mon. I	0·30 Å^{-1}	534	1·25	1031	6×10^{-4}	1·02	0·99	0·04	0·04	0·03	0·04					
$D = 0{\cdot}1°$, $E = 1{\cdot}5°$	0·38	261	1·10	341	6×10^{-4}	1·02	1·14	0·05	0·06	0·03	0·03					
$\sin\theta/\lambda < 0{\cdot}75$ Å^{-1}	0·43	29	1·48	322	6×10^{-4}	1·02	0·96	0·05	0·05	0·03	0·04	322	0·024	0·016	1·93	0·09
	Number of repetitions: 156			$n(I \leq 0) = 0$		$n(\sigma_F/\sigma'_F > 1{\cdot}1) = 17$			$n(\Delta F/\sigma_F > 0{\cdot}2) = 7$							
VO_2 mon. II	0·31	524	1·43	1067	0	1·00	0·97	0·04	0·04	0·04	0·05					
$D = 0{\cdot}1°$, $E = 1{\cdot}5°$	0·31	107	0·87	657	0	1·00	1·13	0·03	0·05	0·06	0·06					
$\sin\theta/\lambda < 0{\cdot}9$ Å^{-1}	0·45	44	1·00	636	0	1·00	0·98	0·03	0·03	0·06	0·06	636	0·062	0·018	1·43	0·06
	Number of repetitions: 104			$n(I \leq 0) = 45$		$n(\sigma_F/\sigma'_F > 1{\cdot}1) = 94$			$n(\Delta F/\sigma_F > 0{\cdot}2) = 72$							

* First line: repeatedly measured intensities, $n_j = 2$; second line: symmetry-equivalent intensities before χ^2 test; third line: symmetry-equivalent intensities after χ^2 test.

graphite monochromator for the other two data sets. Three control reflections were measured for all data sets and the factors v given in the second block of Table 1 are the ratios $s_j^2/\sigma_{\bar{x}j}^2$, *cf.* (4) and (6). In the third block, p denotes the instability constant, (15). The factors v, (14), given in this block were obtained by fitting the observed to the theoretical χ^2 distribution for twice ($n_j = 2$) measured intensities.

The quantities R_i and wR_i are as used by Abrahams, Hamilton & Mathieson (1970). Here, the symbol e means either repeatedly measured or symmetry equivalent. The values of agreement, R and wR, are as defined by Hamilton & Abrahams (1970). The goodness-of-fit parameter S^2 has for large $(n-r)$ the expectation value 1 and the standard deviation $\sigma_{S^2} = [2/(n-r)]^{1/2}$, n is the number of observations and r the number of parameters.

These quantities, expectation value and standard deviation, are unknown for R and wR so that S^2 is to be preferred against R and wR as a measure of agreement as was pointed out by Hahn (1960).

The first two data sets of Table 1 do not contain those intensities for which $I/\sigma_I < 1{\cdot}7$ was found in a short prescan. Taking these intensities into account increases drastically the values of R_i and R and decreases the e.s.d.'s of the refined parameters. The third data set comprises the intensities of all reflections with $\sin\theta/\lambda$ as given in Table 1. The factor p was varied or put equal to zero depending upon whether s_Q came closer to its expectation value or not. For $I > 10\sigma_I$, (22) was used for the derivation of σ_F and (29) of Wilson (1979) for the derivation of F. The other intensities and their variances were transformed *via* (21) and (18).

All three data sets exhibit numerical values of Q and s_Q^2 (for symmetry-equivalent data after the χ^2 test), which indicate the data to be normally and independently distributed and the variances to be properly chosen. Nevertheless, the goodness-of-fit parameters adopt values much larger than acceptable from a statistical point of view.

For the tetragonal phase, several reasons can be given to explain this discrepancy: many intensities, mainly strong ones, were rejected by the χ^2 test indicating strong extinction effects and errors caused by different absorption for different orientations of the crystal ($\mu = 37{\cdot}7$ cm^{-1} for both phases of VO_2, none of the three data sets was corrected for absorption). During the refinement, an additional 34 reflections all with $F > 9{\cdot}0$ were discarded. For most of them F was much less than $|F_{\text{calc}}|$ (F stands for F_{obs}, $F_{\text{max}} \simeq 38$), again showing the influence of extinction.

With the data of the monoclinic phase multipole population parameters also were refined. Here the large value of S^2 to some extent may be caused by some few variances supposedly chosen too low compared with the corresponding sample variances. [For the control reflections, s_j^2 from (4) was taken as the variance of $\bar{x}_j$, (3).] One may therefore try some more modifications of σ_{ji}^2, for example, by dividing all data in groups according to intensities or to $\sin\theta/\lambda$ and by determining p and/or v separately for each group. Another modification, similar to that proposed by Nielson (1977) is $\sigma_{ji}'^2 = \sigma_{ji}^2 + bx_{ji} + px_{ji}^2 + d(\sin\theta/\lambda)$.

Table 2. *Comparison of the two data sets for the monoclinic phase of* VO_2

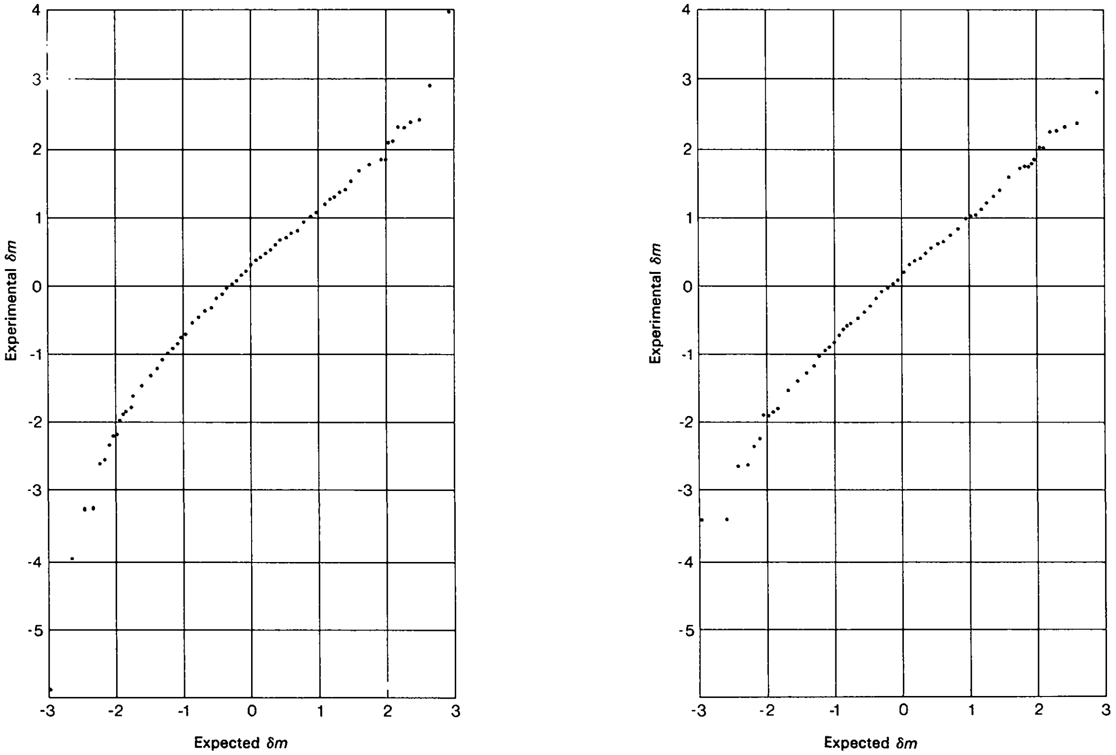
E.s.d.'s for the least significant digits are given in parentheses.

χ^2 test	Number of intensities	k	Q	σ_Q	s_Q	a	b
Before	359	1·573 (5)	1·14	0·075	0·135	0·20 (5)	1·03 (5)
After	312	1·579 (5)	0·93	0·080	0·085	0·10 (6)	0·95 (6)

Here again the constants b, p, d are to be derived from experimental data. However, no considerable improvement of the $\sigma_{ji}'^2$ against those variances that give the Q and s_Q values of Table 1 can be expected, since these values are already close to their expectation values.

Apparently, it must be checked whether there are undetected systematic errors, or if the large goodness-of-fit parameter signals properties of the sample that are not reproduced by the model. For this purpose integrated intensities with different crystals of different sizes will be measured. This can reveal errors caused by absorption, extinction and multiple scattering but not, for example, those of thermal diffuse scattering. For the two data sets of the monoclinic phase given in Table 1 the largest deviations $(F-F_{\text{calc}})/\sigma_F$ confirm each other and the refined parameters agree within the e.s.d.'s, indicating the measurements to be reproducible. If the intensities rejected by the χ^2 test are taken into account the monopole population parameter is changed drastically and most of the multipole population parameters are changed significantly. This result shows that the electron density derived with these data alone may allow chemical bonding in VO_2 to be discussed with qualitative arguments; for a quantitative interpretation it seems to be unsuitable.

The consistency of the two data sets of the monoclinic phase was also checked by means of the quantities Q, σ_Q and s_Q from (11)–(13) and independently by normal probability plots as proposed by Abrahams & Keve (1971) and as described in *International Tables for X-ray Crystallography* (1974). The results are summarized in Table 2 and in Figs. 3(*a*), (*b*). In the outer regions of the plots *each* value, in the middle regions each *fifth* and in the inner regions each *tenth* value is plotted. The scale factor $k=k_2$ for the second data set was determined by minimizing s [($A2$) of Gonschorek (1982*b*)] with respect to $g_2=k_2^{-1}$; $g_1=k_1^{-1}$ was kept constant. The general condition $s=\min$



(*a*) (*b*)

Fig. 3. (*a*) Normal probability plot before the χ^2 test was applied to the intensities of the two data sets: no intensities rejected. (*b*) Normal probability plot after the χ^2 test was applied individually to both data sets: some intensities with the largest χ^2 values rejected.

is, if there are only two data sets, identical with the more special condition (2) given by Abrahams & Keve (1971); the sum over i in ($A2$) can be transformed into $(\delta m_h)^2$. Note that minimization of s in ($A2$) by simply putting $\partial s/\partial g_i = 0$, $i = 1, \ldots, n$, often does not converge, *cf.* Sparks (1970). The arrays in Fig. 3 were approximated by straight lines $\delta m_{\text{exper}} = a + b\delta m_{\text{expect}}$ and the values of a and b are given in Table 2 together with their e.s.d.'s.

The values of Q, σ_Q and s_Q in Table 2 and the plot given in Fig. 3(a) for the data before the χ^2 test show clearly one or both data sets to deviate from a normal distribution in accordance with the large Q values 1·14 and 1·13 given in Table 1. The Q values 0·96 and 0·98 for the two data sets after the χ^2 test may indicate slightly overestimated variances leading to a slope of the normal probability plot less than unity, which is in accordance with the numbers in Table 2 and with the plot in Fig. 3(b).

This work was made possible by a grant from the Deutsche Forschungsgemeinschaft and by the support of Professor Th. Hahn. The program for the least-squares refinement was written by E. Nowack and the drawings were prepared by R. Becker. The criticism of a referee of *Acta Cryst.* on the first version of this paper encouraged the author to develop the relations (9) to (13) and (21).

References

ABRAHAMS, S. C. (1964). *Acta Cryst.* **17**, 1327-1328.
ABRAHAMS, S. C. (1969). *Acta Cryst.* **A25**, 165-173.
ABRAHAMS, S. C. (1974). *Acta Cryst.* **B30**, 261-268.
ABRAHAMS, S. C., HAMILTON, W. C. & MATHIESON, A. McL. (1970). *Acta Cryst.* **A26**, 1-18.
ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* **A27**, 157-165.
ALTE DA VEIGA, L. M., ANDRADE, L. R. & GONSCHOREK, W. (1982). *Z. Kristallogr.* **160**, 171-178.
BLESSING, R. H., COPPENS, P. & BECKER, P. (1974). *J. Appl. Cryst.* **7**, 488-492.
DUNITZ, J. D. (1979). *X-ray Analysis and the Structure of Organic Molecules.* Ithaca: Cornell Univ. Press.
FLACK, H. D. (1984). *Methods and Applications in Crystallographic Computing*, edited by T. ASHIDA & S. HALL, pp. 53-67. Oxford Univ. Press.
FLACK, H. D., VINCENT, M. G. & VINCENT, J. A. (1980). *Acta Cryst.* **A36**, 495-496.
FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* **A34**, 517-525.
GONSCHOREK, W. (1981). *Port. Phys.* **12**, 49-60.
GONSCHOREK, W. (1982*a*). *Z. Kristallogr.* **158**, 197-204.
GONSCHOREK, W. (1982*b*). *Z. Kristallogr.* **160**, 187-203.
GONSCHOREK, W. & FELD, R. (1982). *Z. Kristallogr.* **161**, 1-5.
HAHN, TH. (1960). *Z. Kristallogr.* **113**, 403-429.
HAMILTON, W. C. & ABRAHAMS, S. C. (1970). *Acta Cryst.* **A26**, 18-24.
HAMILTON, W. C., ROLLETT, J. S. & SPARKS, R. A. (1965). *Acta Cryst.* **18**, 129-130.
HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* **A29**, 510-513.
IBERS, J. A. & HAMILTON, W. C. (1964). *Acta Cryst.* **17**, 781-782.
International Tables for X-ray Crystallography (1974). Vol. IV, edited by J. A. IBERS & W. C. HAMILTON, pp. 293-310. Birmingham: Kynoch Press.
JEFFERY, J. W. (1964). *Acta Cryst.* **17**, 1329.
JEFFERY, J. W. & ROSE, K. M. (1964). *Acta Cryst.* **17**, 343-350.
LEHMANN, M. S. (1975). *J. Appl. Cryst.* **8**, 619-622.
MCCANDLISH, L. E., STOUT, G. H. & ANDREWS, L. C. (1975). *Acta Cryst.* **A31**, 245-249.
NIELSEN, K. (1977). *Acta Cryst.* **A33**, 1009-1010.
REES, B. (1976). *Acta Cryst.* **A32**, 483-488.
REES, B. (1977*a*). *Isr. J. Chem.* **16**, 180-186.
REES, B. (1977*b*). *Isr. J. Chem.* **16**, 154-158.
REES, B. (1978). *Acta Cryst.* **A34**, 254-256.
ROLLETT, J. S. (1984). *Methods and Applications in Crystallographic Computing*, edited by T. ASHIDA & S. HALL, pp. 163-175. Oxford Univ. Press.
SCHULZ, H. (1971). *Acta Cryst.* **A27**, 540-544.
SHOEMAKER, D. P. (1968). *Acta Cryst.* **A24**, 136-142.
SPARKS, R. A. (1970). *Crystallographic Computing*, edited by F. R. AHMED, S. R. HALL & C. P. HUBER, pp. 282-284. Copenhagen: Munksgaard.
WAERDEN, B. L. VAN DER (1971). *Mathematische Statistik.* Berlin: Springer.
WILSON, A. J. C. (1976). *Acta Cryst.* **A32**, 781-783.
WILSON, A. J. C. (1979). *Acta Cryst.* **A35**, 122-130.

Acta Cryst. (1985). A41, 195-203

Optimal Strategy of Surveying Reciprocal Space in Protein Crystallography

BY S. V. NIKONOV AND YU. N. CHIRGADZE

Institute of Protein Research, USSR Academy of Sciences, 142292 *Pushchino, Moscow Region, USSR*

(*Received* 25 *May* 1984; *accepted* 6 *November* 1984)

Abstract

A general approach to the search for an optimal strategy of surveying reciprocal space is considered. It consists in determining a set of photographs that give the maximum number of structure factors in the minimum exposure time. Such a set of photographs satisfies the required strategy if the following initial conditions are fulfilled: the given completeness of the data set, the resolution limit, the crystal setting and the technical conditions of the method used for the data-set collection. A general algorithm is applied to the screened precession and screenless rotation-oscillation methods used in protein crystallography.

0108-7673/85/020195-09$01.50