

Table 4. Test result on the effect of $F(000)$

$B = 50; \Delta f_{\text{true}} = -1000; F(000)_{\text{true}} = 367.$

$F(000)$	0	200	367	400	500
S	0.460	0.878	0.970	0.975	0.985
Δf_{est}	-1000	-1000	-1000	-1000	-1000

be seen that $F(000)$ has a significant influence on the value of S , but has little effect on Δf_{est} . Hence we could omit the term $F(000)$ in the calculation of S .

Influence of the presence of heavy atoms

In principle, the Sayre equation is not valid for structures simultaneously containing light and heavy atoms. However, the results shown above on the test structure containing different kinds of atoms like C, N, Cl and Cu demonstrate that the Sayre equation could give satisfactory results. To see the effect of heavy atoms, the copper of the test structure was replaced by platinum. From Table 5, we can see that in the case of platinum perchlorophthalocyanine. Δf_{est} is -940 \AA for $\Delta f = -1000 \text{ \AA}$ giving the resultant image shown in Fig. 2(d), which is still acceptable. A better result was obtained by using the equation of Woolfson (1958) instead of Sayre's. Here Δf_{est} is -970 \AA and an image like Fig. 2(c) was obtained.

Concluding remarks

The procedure proposed in this paper has been shown to be successful in processing theoretical images

Table 5. Test result on the effect of a heavy atom

$B = 50; F(000) = 300; \Delta f_{\text{true}} = -1000 \text{ \AA}.$

Sample	Pt(Clphthalocyanine)		Cu(Clphthalocyanine)
	Sayre equation	Woolfson's equation	Sayre equation
S	0.903	0.959	0.952
Δf_{est}	-940	-970	-1000

without preliminary structural information. In addition, it has been shown that the procedure is not sensitive to errors in the temperature factor and $F(000)$ or to the presence of heavy atoms. The next step in the investigation is to apply the method to experimental EM's. Another important task still to be begun is the extension of the method to include the dynamical diffraction effect.

References

- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286-290.
 FAN HAI-FU, ZHONG ZI-YANG, ZHENG CHAO-DE & LI FANG-HUA (1985). *Acta Cryst.* **A41**, 163-165.
 FIJES, P. L. (1977). *Acta Cryst.* **A33**, 109-113.
 KLUG, A. (1978/79). *Chem. Scr.* **14**, 245-256.
 LI FANG-HUA & FAN HAI-FU (1979). *Acta Phys. Sin.* **28**, 276-278.
 SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.
 UNWIN, P. N. T. & HENDERSON, R. (1975). *J. Mol. Biol.* **94**, 425-440.
 UYEDA, N. & ISHIZUKA, K. (1974). Eighth Int. Congr. Electron Microscopy, Vol. 1, pp. 322-323.
 WOOLFSON, M. M. (1958). *Acta Cryst.* **11**, 277-283.

Acta Cryst. (1986). **A42**, 356-362

Positional Parameters Obtained With Anharmonic Temperature Factors

By C. SCHERINGER

Institut für Mineralogie der Universität Marburg, D-3550 Marburg/Lahn, Federal Republic of Germany

(Received 4 February 1985; accepted 2 April 1986)

Abstract

Comparisons of distances between the positions obtained from harmonic and anharmonic refinements lead to the conclusion that the positional parameters may have different physical meanings for the two cases. The mean positions are obtained if the temperature factor (t.f.) has no first-order terms in the reciprocal-lattice vector \mathbf{Q} (harmonic t.f., Fourier transforms of the Gram-Charlier series); if there are

first-order terms in \mathbf{Q} , other positions will be obtained whose meaning needs to be established. The advantages associated with the mean positions are described, and the disadvantages associated with other positions are illustrated with an example from the literature. A procedure is described in which the physical meaning (if there is any) of a non-mean position can be established and the mean position calculated. The problem of parameter bias is analyzed and numerical results are discussed for three

structures. Provided the anharmonic model produces a significantly better fit to the experimental data, it is concluded that better estimates of the mean positions are obtained with anharmonic than with harmonic t.f.'s.

1. Introduction

While comparing positional parameters and bond lengths obtained with anharmonic refinements, we found that different types of temperature factors (abbreviated t.f.'s) result in different types of positional parameters. With harmonic t.f.'s no problem arises: the mean positions are always obtained. With anharmonic t.f.'s other positions can be obtained, like the most frequent positions from an isolated-atom-potential (IAP) approach (Dawson, Hurley & Maslen, 1967; Willis, 1969; Willis & Pryor, 1975; Scheringer, 1985). With anharmonic t.f.'s it may happen that the physical meaning of the obtained positional parameters is lost (see the example in § 4). We shall show that the choice of type of positional parameters does not matter for the calculation of structure factors but that one must determine what type of parameters one has defined. This is particularly important if one wants to compare positional parameters obtained from different refinements. In this paper we shall describe a procedure which makes it possible to determine the meaning of the positional parameters unequivocally, provided there is a meaning. For the comparison of bond lengths as distances between positions it is necessary to use positional parameters that have the same meaning. In §§ 2 and 3 we shall show that the parameters describing the mean positions offer several advantages and are best suited for the comparison of bond lengths.

Another problem arising from the use of anharmonic t.f.'s is as follows: even if the same type of position is determined throughout, e.g. the mean position, different numerical values of the positional parameters are often obtained with harmonic and anharmonic t.f.'s. The question arises which t.f. will yield the 'best' values of the positional parameters. Johnson (1969, 1970) remarks that anharmonic t.f.'s do not give 'usable estimates for the atomic positional parameters' whereas harmonic t.f.'s give 'unbiased estimates'. Johnson (1969, 1970) does not establish his opinion by means of statistical theory or by quoting relevant literature, but he points out that there are correlations among the tensor components of odd order which also affect the positional parameters. This, however, does not answer our basic question, and in § 5 we shall analyze the problem of the best estimate of positional parameters by means of the method of least squares.

2. The meaning of the positional parameters

First, we consider why the positional parameters can have different meanings. From the general derivation

of the (also anharmonic) t.f. given elsewhere (Scheringer, 1986a), it can be seen that the positional parameter \mathbf{x} in the structure factor has no other immediate meaning than the designation of the origin of the vibrational coordinates \mathbf{u} of an atom in the unit cell; i.e. $\mathbf{u} = \mathbf{0}$ at the position \mathbf{x} . The vibrational coordinates \mathbf{u} are the variables of the probability density function (abbreviated p.d.f.) whose Fourier transform is the t.f. The origin $\mathbf{u} = \mathbf{0}$ in the p.d.f. can be freely chosen. A change in the location $\mathbf{u} = \mathbf{0}$ in the p.d.f. causes a corresponding change in the positional parameters \mathbf{x} , and, hence, in the meaning of these parameters. If only the origin $\mathbf{u} = \mathbf{0}$ is shifted and the shape and position of the p.d.f. in the unit cell is preserved, the same structure factors will be calculated. The analytical formulation concerning such a shift of origin is as follows: Choose two origins for the p.d.f., shown as I ($\mathbf{u} = \mathbf{0}$) and II ($\mathbf{u}' = \mathbf{0}$) in Fig. 1. The corresponding positions in the unit cell are \mathbf{x} and \mathbf{x}' , the p.d.f.'s $p(\mathbf{u})$ and $p'(\mathbf{u}')$ and the t.f.'s $T(\mathbf{Q})$ and $T'(\mathbf{Q})$ respectively. Omitting the scattering factors, we obtain as contributions to the structure factor

$$T(\mathbf{Q}) \exp(i\mathbf{Q} \cdot \mathbf{x}) \quad \text{for origin I,} \quad (1a)$$

$$T'(\mathbf{Q}) \exp(i\mathbf{Q} \cdot \mathbf{x}') \quad \text{for origin II.} \quad (1b)$$

We shall show that (1a) and (1b) give the same contribution. With $\mathbf{x}' = \mathbf{x} + \Delta$ (see Fig. 1) we have $\exp(i\mathbf{Q} \cdot \mathbf{x}') = \exp(i\mathbf{Q} \cdot \mathbf{x}) \exp(i\mathbf{Q} \cdot \Delta)$. If we view the p.d.f. from position II in Fig. 1, it appears to be shifted by an amount $-\Delta$ compared with the view from position I. Hence, the Fourier transforms of the two p.d.f.'s $p(\mathbf{u})$ and $p'(\mathbf{u}')$ differ only by a phase factor; i.e. $T'(\mathbf{Q}) = T(\mathbf{Q}) \exp[i\mathbf{Q} \cdot (-\Delta)]$. The two factors in $T'(\mathbf{Q}) \exp(i\mathbf{Q} \cdot \mathbf{x}')$ containing Δ and $-\Delta$ cancel and the equality of (1a) and (1b) is proven.

Since the origin $\mathbf{u} = \mathbf{0}$ in the p.d.f. can be freely chosen, the meaning of the positional parameters \mathbf{x} is defined only by the physical meaning of the location $\mathbf{u} = \mathbf{0}$ in the p.d.f. In actual refinements, the location $\mathbf{u} = \mathbf{0}$ is fixed in the p.d.f. by the form of the t.f. being used. Several forms of the t.f. have been used in the past and it is expedient to sort them into two classes. Class 1: $\langle \mathbf{u} \rangle = \mathbf{0}$, $\mathbf{x} = \text{mean position}$; class 2: $\langle \mathbf{u} \rangle \neq \mathbf{0}$, $\mathbf{x} = \text{some other position}$. Members of class 1 are the conventional harmonic t.f., the Fourier transform of the Gram-Charlier series [*International Tables for X-ray Crystallography*, 1974, p. 317, equation (9)] and the cumulant expansion without first-order terms of

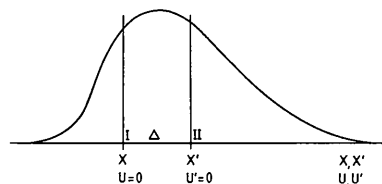


Fig. 1. Two different origins, I and II, of the vibrational coordinates chosen for the same p.d.f.

\mathbf{Q} [Johnson, 1969, equation (4); *International Tables for X-ray Crystallography*, 1974, p. 317, equation (10), omitting the first cumulant term with K^j]. Members of class 2 are IAP t.f.'s without site-symmetry restrictions (Mair & Barnea, 1975; Whiteley, Moss & Barnea, 1978; Scheringer, 1985) and some t.f.'s used for describing molecular librations (Willis & Pryor, 1975, equations 6.29–6.31). A further t.f. of class 2, which is composed of three different factors, will be discussed in § 4.

The two classes are also distinguished by the presence or absence of first-order terms in the reciprocal-lattice vector \mathbf{Q} in the t.f. T.f.'s without first-order terms in \mathbf{Q} are always members of class 1; t.f.'s with such terms are members of class 2 (there is one exception which we shall discuss in § 4). A 'first-order term' must have \mathbf{Q} in the first power but may represent different functions of \mathbf{Q} . Our statements above follow from a theorem on the calculation of moments of a p.d.f. from its Fourier transform (Kendall & Stuart, 1969, pp. 60–61). The first moment, *i.e.* the mean, of a p.d.f. is given by

$$\langle u_j \rangle = \partial T(\mathbf{Q}) / i \partial Q_j |_{\mathbf{Q}=\mathbf{0}}, \quad j = 1, 2, 3. \quad (2)$$

Evaluation of (2) proves our statements. Only first-order terms in \mathbf{Q} yield components $\langle u_j \rangle \neq 0$. With an inclusion or omission of first-order terms in the t.f. the origin $\mathbf{u} = \mathbf{0}$ in the p.d.f. is always shifted, but the p.d.f. is fully preserved if and only if the terms are of the type $\exp(i\mathbf{Q} \cdot \Delta)$ (see above). Hence, by including or omitting first-order terms of \mathbf{Q} in the t.f., the shape of the p.d.f. is usually slightly changed. Similarly, since first-order terms in \mathbf{u} in the p.d.f. give rise to first-order terms in \mathbf{Q} in the t.f. which usually differ from $\exp(i\mathbf{Q} \cdot \Delta)$, inclusion or omission of first-order terms in \mathbf{u} in the p.d.f. not only causes an origin shift but also changes the shape of the p.d.f. slightly.

With t.f.'s of class 2 there is always the problem of establishing the meaning of the positional parameters, *i.e.* the meaning of the location $\mathbf{u} = \mathbf{0}$ in the p.d.f. Besides the mean, the maximum of the p.d.f. obviously represents a location of physical significance. However, many locations in a p.d.f. have no particular meaning. Hence, it is possible that one has unknowingly defined a location $\mathbf{u} = \mathbf{0}$ in the p.d.f. which has no physical meaning and has calculated bond distances which are correspondingly meaningless (see § 4).

3. Advantages associated with t.f.'s of class 1

We recommend the use of t.f.'s of class 1, *i.e.* $\langle \mathbf{u} \rangle = \mathbf{0}$, \mathbf{x} = mean position, because they offer several advantages which we sum up as follows:

(1) Uncertainties in the meaning of the positional parameters cannot arise.

(2) Bond distances calculated from conventional harmonic and anharmonic refinements can be compared directly.

(3) Formulae for bond-length corrections are only known for bond lengths which were defined as distances between the mean positions of two atoms. The corrections can be applied directly to the distances calculated from the positional parameters.

(4) T.f.'s of class 1 (usually) have no first-order terms in \mathbf{Q} and thus assume their simplest form (shortest programming, shortest computing times).

(5) With $\langle \mathbf{u} \rangle = \mathbf{0}$, the central moments of the p.d.f. can be calculated directly from the t.f. (Kendall & Stuart, 1969, pp. 60–61).

(6) The positional parameters not only denote the positions of the means of the individual atomic p.d.f.'s but also the positions of the means of the joint p.d.f. which describes the vibrations of all atoms in the crystal. This follows from the fact that the moments of a (joint) p.d.f. are defined by moments of the respective marginal p.d.f.'s (*International Tables for X-ray Crystallography*, 1959, ch. 2.6). It has been shown elsewhere that the p.d.f. of a given atom is the three-dimensional marginal p.d.f. of the crystal p.d.f. (Scheringer, 1986a). There is no other type of positional parameter that represents a respective positional property of the crystal p.d.f. (or of the crystal potential; see § 4).

4. Problems arising with t.f.'s of class 2

Our analysis in § 2 has shown that, with t.f.'s of class 2 (*i.e.* $\langle \mathbf{u} \rangle \neq \mathbf{0}$), the meaning of the positional parameters is not automatically given and must be established when unconventional t.f.'s are used. Furthermore, for comparison with results from harmonic refinements, the mean positions must be calculated explicitly. In this section we shall first discuss the general attack on these two problems. With IAP t.f.'s it is commonly assumed that the minimum of the potential has been determined. We shall show that this assumption is not generally valid. Finally, we shall discuss an example from the literature in which the meaning of the positional parameters has not been established and so has led to an inadequate interpretation of the calculated bond distance.

Determination of the meaning of the positional parameters and calculation of the mean position

If the p.d.f. is given in analytical form, the location $\mathbf{u} = \mathbf{0}$ in the p.d.f. can be calculated and the meaning of this location (if there is any) can be established either by analytical means or simply with the aid of a drawing. If the analytical form of the p.d.f. is not known, the p.d.f., and hence the location of its origin $\mathbf{u} = \mathbf{0}$, can be calculated by means of numerical Fourier inversion of the t.f. The position $\mathbf{u} = \mathbf{0}$ can,

however, be determined more easily relative to the mean $\langle \mathbf{u} \rangle$ by means of (2). This may already give an indication as to whether or not the position $\mathbf{u} = \mathbf{0}$ denotes a point of physical significance.

Having determined $\langle \mathbf{u} \rangle$ from the t.f., one can immediately calculate the mean position \mathbf{x}_m in the unit cell. Let \mathbf{x} in Fig. 1 be the position of $\mathbf{u} = \mathbf{0}$ in the cell, \mathbf{x}' the mean position \mathbf{x}_m , and $\Delta = \langle \mathbf{u} \rangle$ follows. Hence

$$\mathbf{x}_m = \mathbf{x} + \langle \mathbf{u} \rangle. \quad (3)$$

However, it is better to determine the mean positions directly in the refinement. There are two procedures, which differ in detail. The first represents the exception for class 2 which we have mentioned in § 2. In addition to the first-order term in \mathbf{Q} present in the t.f., one introduces a second first-order term, the factor $\exp(i\mathbf{Q} \cdot \langle -\mathbf{u} \rangle)$. Then both terms together yield $\langle \mathbf{u} - \mathbf{u} \rangle = \mathbf{0}$. In each cycle of the refinement, $\langle -\mathbf{u} \rangle$ is calculated anew from the given first-order term in \mathbf{Q} and the current values of the parameters. With this procedure, the position and shape of the p.d.f. are fully maintained in the unit cell. The second procedure is simpler to apply. In the given t.f. all first-order terms in \mathbf{Q} are omitted. Since the first-order terms usually are not exponential factors, the shape of the p.d.f. is slightly changed. Hence, the two procedures are not fully equivalent. Obviously, they become more equal as $\langle \mathbf{u} \rangle$ becomes smaller. Usually one will not be able to decide which form of the p.d.f. is (physically) better.

For the example of CdSe (Whiteley, Moss & Barnea, hereafter WMB), we have assessed the differences in the numerical values of the parameters which arise with the two procedures. With the first procedure, the thermal parameters obtained by WMB are not altered at all. To determine the effect on the thermal parameters with the second procedure, we should perform the refinement using WMB's t.f. without first-order terms in \mathbf{Q} . We cannot do this since we do not have the necessary data, but we can assess the changes of the thermal parameters as follows. We multiply WMB's t.f. with the factor $\exp(i\mathbf{Q} \cdot \langle -\mathbf{u} \rangle)$ and so ensure that the (calculated) structure factors of the final least-squares solution are preserved. Now we expand the factor $\exp(i\mathbf{Q} \cdot \langle -\mathbf{u} \rangle)$ in a series and so obtain the changes of the thermal parameters of different orders. The linear first-order terms completely cancel and thus our second procedure is followed. The harmonic parameters B_{33} are modified by the terms $2\pi \langle -u_3 \rangle^2 / c^2$. Using $c = 7.01 \text{ \AA}$, $\langle u_3 \rangle(\text{Cd}) = 0.006 \text{ \AA}$ and $\langle u_3 \rangle(\text{Se}) = -0.003 \text{ \AA}$ (Barnea, private communication), we obtain the changes $\Delta B_{33}(\text{Cd}) = 0.0028 \text{ \AA}^2$ and $\Delta B_{33}(\text{Se}) = 0.0007 \text{ \AA}^2$ (to a good approximation). These are 0.16 and 0.07% of the values of B_{33} . The changes in the third-order terms are calculated from $(2\pi)^3 \langle -u_3 \rangle^3 / (3! c^3)$ and amount to 0.10 and 0.08% of the third-order terms

for Cd and Se respectively. The effect of the higher terms in the expansion of $\exp(i\mathbf{Q} \cdot \langle -\mathbf{u} \rangle)$ can be neglected. The differences in the thermal parameters, as obtained with the two procedures, are clearly smaller than the errors in the parameters. Thus, with the quality of present-day intensity data, it does not matter which procedure is used. Omission of the first-order terms in \mathbf{Q} in the t.f. is simpler, at any rate.

Positions obtained with the IAP t.f.

IAP t.f.'s are commonly defined so that the p.d.f. does not contain first-order terms in \mathbf{u} (Mair & Barnea, 1975; WMB; Scheringer, 1985). Hence, from the p.d.f. given, we can only deduce that the origin $\mathbf{u} = \mathbf{0}$ is located at the maximum of the p.d.f. Any assertion beyond this statement cannot be derived from the p.d.f. itself and thus is an interpretation performed by the investigator. Since the IAP t.f. is derived by means of an effective potential and classical statistics, it has become common practice to identify the maximum of the p.d.f. with the minimum of the potential. However, such an identification cannot rigorously be established because the expansion of the IAP p.d.f. diverges in every case and does not reproduce the Boltzmann function (Scheringer, 1985). Furthermore, Ibers (1959) has shown that the maxima of the p.d.f. do not generally correspond to minima of a potential. In particular, Ibers's (1959) equation (6) shows that the distances between the maxima of the p.d.f. are not equal to the distances between the minima on the potential-energy hypersurface of the molecule.

Moreover, the position of the $(3nN - 6)$ -dimensional potential minimum of the nN atoms in the crystal cannot be derived from the n IAP t.f.'s (except for the case where it is fixed by site symmetry). In general, the maxima of the n p.d.f.'s in a cell do not correspond to the $(3nN - 6)$ -dimensional maximum of the joint p.d.f. of the crystal, and even if they do, one does not know how the $(3nN - 6)$ -dimensional potential minimum is related to the $(3nN - 6)$ -dimensional maximum of the joint p.d.f. Thus, it is impossible to deduce the positions of crystal-potential minima from IAP t.f.'s. Note, on the other hand, the advantage (6) obtained for the mean positions with t.f.'s of class 1.

The meaning of the atomic position has not been established

As an illustration we shall now discuss the calculation of the O-H bond length in potassium oxalate hydrate, as performed by Eriksson, Hermansson, Lindgren & Thomas (1982), hereafter EHLT. The vibrations of the H atoms were described as being partly anharmonic, *i.e.* in the librations around the oxygen atoms and in the bond-stretching modes. EHLT used six different models for structure-factor

calculation: the conventional harmonic model (model I) and five anharmonic models with various modifications. The model of central interest is model II for which the p.d.f. is composed of three factors, *i.e.*

$$T(\mathbf{Q}) = T(\mathbf{Q})_{\text{harm}} T(\mathbf{Q})_{\text{lib}} T(\mathbf{Q})_{\text{stretch}} \quad (4)$$

Both $T(\mathbf{Q})_{\text{lib}}$ and $T(\mathbf{Q})_{\text{stretch}}$ are composed of several factors which were obtained from a normal coordinate analysis [see EHLT, equation (4)]. O–H bond distances were calculated for all six models (EHLT, Table 1). Model I gives 0.963 Å, model II 0.967 Å. EHLT assess the utility of each model by stating that the anharmonic model II facilitates ‘... a more sophisticated analysis of the diffraction-obtained geometrical and thermal parameters’ and that the conventional harmonic model is ‘... shown to give considerable systematic errors in the geometrical parameters for the water molecule’.

In our opinion, EHLT failed to establish the meaning of the positional parameters of the H atom, which is associated with their model II. We shall show that the mean positions for models I and II coincide within the limits of experimental errors, and that the O–H bond distance of 0.967 Å for model II has no meaning at all.

With the three factors given in (4), the p.d.f. is the convolution of the three single p.d.f.’s p_{harm} , p_{lib} and p_{stretch} . Because of symmetry we obtain $\langle u_1 \rangle = \langle u_2 \rangle = 0$ [see EHLT (Fig. 2, axis 3 in the direction O–H)]. $\langle u_3 \rangle$ is unknown. We obtain it from (4) by means of (2) according to

$$\langle u_3 \rangle = \langle u_3 \rangle_{\text{lib}} + \langle u_3 \rangle_{\text{stretch}} \quad (5)$$

because $\langle u_3 \rangle_{\text{harm}} = 0$ and all p.d.f.’s are normalized, *i.e.* $T(\mathbf{Q} = \mathbf{0}) = 1$. To evaluate (5), we must use t.f.’s of the single internal modes, $i = 1, \dots, n$, and determine the means $\langle u_3 \rangle_i$ by use of (2). Then $\langle u_3 \rangle = \sum \langle u_3 \rangle_i$. From the product of the n_1 librational modes we find, from (2) and EHLT [equation (5)],

$$\langle u_3 \rangle_{\text{lib}} = \sum_{i=1}^{n_1} \left(-\frac{1}{2} a L_{11} \right)_i \quad (6)$$

We do not know the values of L_{11} for the individual internal modes, but we can estimate $\langle u_3 \rangle_{\text{lib}}$ by virtue of the fact that $\langle u_3 \rangle_{\text{lib}}$ is the reverse bond-length correction. From the bond lengths of models I and VI (EHLT, Table 1) we obtain $\langle u_3 \rangle_{\text{lib}} = 0.963 - 0.989 = -0.026$ Å. From Fig. 3 of EHLT, $\langle u_3 \rangle_{\text{stretch}} = 0.0107$ Å for each of the two stretching modes (Eriksson, 1984, private communication). Hence $\langle u_3 \rangle = -0.026 + 2 \times 0.0107 = -0.0046$ Å. With these values, we find the mean position of the H atom for model II to be located on the O–H bond at a distance of $0.967 - 0.005 = 0.962$ Å from the O atom. Within the limits of error, this value coincides with the bond length of 0.963 Å obtained with the harmonic model.

To determine the physical meaning of EHLT’s bond distance of 0.967 Å (model II), one may evalu-

ate the p.d.f. of the t.f. (4) and find the meaning of the position $u_3 = 0$ in the p.d.f. Since the p.d.f. is not known analytically, a numerical Fourier inversion of the t.f. for points along u_3 could help. However, the following consideration may be sufficient. Since the two models for bond stretching and librations (models III and IV) are basically different, the sum $\langle u_3 \rangle_{\text{lib}} + \langle u_3 \rangle_{\text{stretch}}$ will assume any unforeseeable value, depending on the number and type of internal modes that are included in the calculation. Hence, the origin $u_3 = 0$ will assume any unforeseeable position relative to the mean $\langle u_3 \rangle$ and will lie anywhere in the p.d.f. Since generally it cannot be expected that such a position denotes a point of physical significance, the obtained O–H distance of 0.967 Å is correspondingly meaningless (unless the contrary is proven). Thus EHLT’s statement that the conventional harmonic refinement is ‘... shown to give considerable systematic errors in the geometrical parameters for the water molecule’ has no physical basis. It is better to avoid such situations by placing the origin $\mathbf{u} = \mathbf{0}$ at the mean of the p.d.f. In EHLT’s investigation, this can be done by omitting the first-order term $[1 - (i/2)L_{11}aQ_3]^{-1/2}$ in equation (5) of EHLT, and placing the origin of p_{stretch} in Fig. 3 of EHLT at its mean.

5. Parameter bias

Refinements have shown that different numerical values for the mean positions are obtained with harmonic and anharmonic t.f.’s. Hence the question is: which t.f. gives rise to the best estimate of the mean position? Johnson (1969, 1970) states that harmonic t.f.’s will do so since, in his opinion, only these give rise to unbiased estimates. Common sense suggests that with a better physical model, *i.e.* with anharmonic t.f.’s, better estimates can be obtained. This view also seems to be the basis of EHLT’s investigation. In the following it is sufficient to consider only the mean positions since these can always be calculated from any other position.

‘Bias’ in the estimated parameters is a concept in mathematical statistics. Let $\mathbf{E}(\mathbf{p})$ be the expected values of the parameters \mathbf{p} and \mathbf{p}_{true} the true (unknown) values; then the bias \mathbf{b} is defined by

$$\mathbf{b} = \mathbf{E}(\mathbf{p}) - \mathbf{p}_{\text{true}} \quad (7)$$

Linnik (1961, pp. 127–128) has shown that in linear least squares the estimated parameters are unbiased if the observed data are unbiased (with respect to their error distribution). The ‘model’, *i.e.* the relation between parameters and data, is assumed to be known and correct. Wilson (1973), by means of a linear Taylor expansion of the calculated structure factors, obtained the same result for one parameter in non-linear least squares (see also Wilson, 1976, 1979). It is not difficult to extend Wilson’s (1973) treatment to several parameters, but we shall not do so here.

Wilson (1973) has formally also included effects of an incomplete or incorrect model. Such effects give rise to 'biasing terms'. Extending Wilson's (1973) results to the case of several parameters, we obtain

$$\mathbf{p} = \mathbf{p}_{\text{true}} + \mathbf{f}(\boldsymbol{\sigma}) + \mathbf{m}(\boldsymbol{\delta}), \quad (8)$$

where $\boldsymbol{\sigma}$ denotes the errors in the observed structure factors and $\boldsymbol{\delta}$ the differences between calculated and true structure factors (calculated from an incomplete model but with the true parameters). The statistical bias is now obtained by calculating the expected values from the error distribution of the observed data, *i.e.* by $E[\mathbf{f}(\boldsymbol{\sigma})]$. The term $\mathbf{m}(\boldsymbol{\delta})$ does not contribute to statistical bias since $\mathbf{m}(\boldsymbol{\delta})$ is obtained from a theoretical model. If we assume no bias in the data, *i.e.* $E[\mathbf{f}(\boldsymbol{\sigma})] = \mathbf{0}$, we obtain from (8)

$$E(\mathbf{p}) = \mathbf{p}_{\text{true}} + \mathbf{m}(\boldsymbol{\delta}). \quad (9)$$

In (9) $\mathbf{m}(\boldsymbol{\delta})$ looks formally like statistical bias although it is not. With 'bias due to defects in the model' Wilson (1984, private communication) wanted to describe errors in the parameters obtained when the model is incomplete or incorrect and here Wilson did not consider the statistical meaning of the word 'bias'. In our opinion, terms like $\mathbf{m}(\boldsymbol{\delta})$ are correctly placed in (8), since they denote a systematic deviation between \mathbf{p} and \mathbf{p}_{true} , whereas in equations like (9) they might be mistaken for statistical bias.

We conclude that by means of statistics we cannot confirm Johnson's (1969, 1970) statement that harmonic t.f.'s give unbiased estimates of the positional parameters and anharmonic t.f.'s do not. Correlations among odd-order tensors may affect the obtained positional parameters, but this does not imply that these are 'biased estimates'. Rather, it can be argued that the positional parameters become erroneous because higher odd-order tensors have been omitted, although they may be required to complete the model and thus enable a better description of the physical reality.

Examples

We shall consider numerical examples from three different investigations. They have in common that the anharmonic contributions originate from librational motion. The numerical results all agree in that the bond distances obtained from *anharmonic* refinements are *shorter* than those obtained from the harmonic refinement.

Our first example is the O-H bond distance in EHLT's investigation. The same type of position of the H atom is defined with models V and VI. With the anharmonic model V an O-H distance of 0.988 Å is obtained; with the nearly harmonic model VI (no third-order term in Q) a distance of 0.989 Å is obtained (see EHLT, Table 1).

In our second example we consider the results on urea (neutron diffraction data) reported by Pryor &

Sanger (1970). They compare the bond lengths obtained with two different models: (1) a conventional refinement and bond lengths corrected for the effects of molecular libration, and (2) a conventional refinement including anharmonic effects of curvilinear motion (third-order tensors) and bond-length correction (first-order tensors) (see Pryor & Sanger, 1970, Table 8). All four bond lengths obtained with the inclusion of anharmonic components were shorter than those obtained with the conventional refinement; the differences were 0.002, 0.003, 0.010 and 0.002 Å for the C-O, C-N, N-H(1) and N-H(2) bond lengths, respectively.

Finally, we present some new results obtained with new refinements of the structure of thiopyridone (Ohms, Guth, Kutoglu & Scheringer, 1982), $R(F) = 0.028$ for 873 neutron diffraction data (Scheringer, 1986*b*). We used a Gram-Charlier series as p.d.f. with third-order terms in \mathbf{u} (components C^{ijk}). There are only three anharmonic parameters, L_{11} , L_{22} , L_{33} (referring to the three principal axes of libration), from which all the anharmonic components C^{ijk} were calculated. With the three parameters L_{ii} added to the conventional harmonic model the improvement in fitting the observed structure factors was significant [level of significance $\alpha < 0.005$, $L_{ii}/\sigma(L_{ii}) = 50, 60, 31$ for $i = 1, 2, 3$]. With anharmonic t.f.'s, the bond lengths are shorter by 0.0004–0.0052 Å (mean 0.0023 Å) than those obtained with harmonic t.f.'s.

Since in all three investigations the bonds were shortened when anharmonic components of librational origin were included, we are led to conclude that this bond shortening is a systematic effect. It arises from the better (anharmonic) model that was used to describe librational motion. Since one cannot prove that harmonic t.f.'s give unbiased estimates of the positional parameters, we conclude that anharmonic t.f.'s give better estimates if the following two conditions are fulfilled: (1) the motions of the atoms are really anharmonic, and (2) the anharmonic parameters give rise to a significantly better agreement between observed and calculated structure factors. This view may be put on a more solid basis if, for some molecules, the bond lengths derived from the most accurate structure determinations are found to correspond to bond lengths obtained by means of spectroscopic or gas-electron-diffraction methods (Landolt-Börnstein, 1976).

I thank Dr A. Eriksson for supplying the numerical value of the mean of the bond-stretching p.d.f., and Professor A. J. C. Wilson for correspondence and a discussion on the problem of parameter bias.

References

- DAWSON, B., HURLEY, A. C. & MASLEN, V. W. (1967). *Proc. R. Soc. London Ser. A*, **298**, 289–306.

- ERIKSSON, A., HERMANSSON, K., LINDGREN, J. & THOMAS, J. O. (1982). *Acta Cryst.* **A38**, 138-142.
- International Tables for X-ray Crystallography* (1959). Vol. II. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)
- International Tables for X-ray Crystallography* (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)
- IBERS, J. A. (1959). *Acta Cryst.* **12**, 251-252.
- JOHNSON, C. K. (1969). *Acta Cryst.* **A25**, 187-194.
- JOHNSON, C. K. (1970). *Thermal Neutron Diffraction*, pp. 132-160. Oxford Univ. Press.
- KENDALL, M. G. & STUART, A. (1969). *The Advanced Theory of Statistics*, Vol. I. London: Griffin.
- LANDOLT-BÖRNSTEIN (1976). New Series, Vol. 7. *Structure Data of Free Polyatomic Molecules*. Berlin, Heidelberg, New York: Springer.
- LINNIK, J. W. (1961). *Die Methode der kleinsten Quadrate in moderner Darstellung*. Berlin: VEB Verlag der Wissenschaften.
- MAIR, S. L. & BARNEA, Z. (1975). *Acta Cryst.* **A31**, 201-207.
- OHMS, U., GUTH, H., KUTOGLU, A. & SCHERINGER, C. (1982). *Acta Cryst.* **B38**, 831-834.
- PRYOR, A. W. & SANGER, P. L. (1970). *Acta Cryst.* **A26**, 543-558.
- SCHERINGER, C. (1985). *Acta Cryst.* **A41**, 73-79.
- SCHERINGER, C. (1986a). *Acta Cryst.* Submitted.
- SCHERINGER, C. (1986b). In preparation.
- WHITELEY, B., MOSS, G. & BARNEA, Z. (1978). *Acta Cryst.* **A34**, 130-136.
- WILLIS, B. T. M. (1969). *Acta Cryst.* **A25**, 277-300.
- WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal Vibrations in Crystallography*. Cambridge Univ. Press.
- WILSON, A. J. C. (1973). *Acta Cryst.* **B29**, 1488-1490.
- WILSON, A. J. C. (1976). *Acta Cryst.* **A32**, 994-996.
- WILSON, A. J. C. (1979). *Acta Cryst.* **A35**, 122-130.

Acta Cryst. (1986). **A42**, 362-368

Unique Solutions for the SIR and SAS Phase Problems and the Use of Partial Structural Information in Phase Refinement

BY DAVID A. LANGS

Medical Foundation of Buffalo, Inc., 73 High St, Buffalo, NY 14203, USA

(Received 12 October 1985; accepted 3 April 1986)

Abstract

Algebraic formulae are presented which permit a unique phased solution for diffraction data measured from a single isomorphous pair of crystals. Trial calculations performed on an SIR (single isomorphous replacement) data set from an 84-atom structure demonstrate that complete phasing can be achieved from a single chirally positioned replacement atom representing less than one percent of the total scattering power of the derivative structure. Similar phase refinements employing error-free SIR data for 2Zn pig insulin are less remarkable, and converge to an average phase error of 50°. The phase convergence of the formulae can be markedly improved if estimates of the cosine invariants from the SIR data are available [Hauptman (1982). *Acta Cryst.* **A38**, 289-294; Fortier, Moore & Fraser (1985). *Acta Cryst.* **A41**, 571-577]. The precision of these cosine estimates was found not to be critical; modular estimates of +1 or -1 were sufficient to allow the SIR phase refinement of the insulin structure to converge to an average phase error of 6°, which compares favorably with the value of 3° produced if the cosine invariants were known precisely. The derived formulae are also shown to be applicable to single-crystal analyses which utilize one-wavelength anomalous dispersion or partial structural fragments to initiate phasing. Test examples indicate that tangent-formula recycling procedures based on the derived formulae

compare favorably with the traditional tangent-formula methods to exploit partial structure information.

Introduction

Recent developments in the improvement of translation-function methods (Langs, 1985) have introduced formulae which enable one to determine unambiguously the sine component of translation-function coefficients, given only the real component of these inherently complex-valued quantities. These procedures may be shown to be useful in other crystallographic situations in which it would be advantageous to retrieve the imaginary component of a complex variable for which only the real component is known. Several important applications may be found in structure-determination methods which use either partial structural models, anomalous scattering or isomorphous replacement to initiate phasing.

Crystallographic single isomorphous replacement (SIR) methods possess a known twofold ambiguity in the determination of non-centrosymmetric phases based on the determined positions of the replacement atoms. The resolution of this ambiguity by multiple isomorphous replacement relationships (Green, Ingram & Perutz, 1954; Harker, 1956) is dependent on the ability to prepare additional isomorphous derivatives of the native structure with replacement atoms possessing significantly large scattering powers