Acta Cryst. (1991). A47, 336-340

Exact Random-Walk Models in Crystallographic Statistics. VII. An All-Space-Group Study of the Effects of Atomic Heterogeneity on the P.D.F.'s of $|E|$

BY SAVELY RABINOVICH, URI SHMUELI, ZAFRA STEIN AND RONEN SHASHUA

School of Chemistry, Tel Aviv University, 69 978 *Tel Aviv, Israel*

AND GEORGE H. WEISS

Division of Computer Research and Technology, Laboratory of Physical Sciences, National Institutes of Health, Bethesda, Maryland 20982, *USA*

(*Received* 8 *April* 1990; *accepted* 22 *November* 1990)

Abstract

Exact expressions have been found for the probability density functions (p.d.f.'s) of the magnitude of the normalized structure factor for all the two-dimensional and most three-dimensional space groups [Part VI: Rabinovich, Shmueli, Stein, Shashua & Weiss (1991). *Acta Cryst.* A47, 328-335]. The results of that investigation are used in the present article to examine some effects of atomic heterogeneity, in the various space-group symmetries, on the p.d.f.'s. Some typical comparisons are made between p.d.f.'s based on the central limit theorem and p.d.f.'s computed from exact formulae. In addition, the exact results are compared to histograms of simulated values of $|E|$. It is found that the p.d.f.'s for some space groups are influenced rather strongly by the presence of outstandingly heavy scatterers, but they are quite insensitive to the presence of such scatterers in other space groups. The often made general statement 'The presence of outstandingly heavy scatterers may invalidate the indications of Wilson's statistics' is made more precise here, insofar as it depends on the particular space group.

Introduction

The probability density functions (p.d.f.'s) and other statistics of the magnitude $|E|$ of the normalized structure factor have served during the past four decades as a conveniently computable means for the resolution of space-group ambiguities. Expressions for these statistics are incorporated into most structure-determination packages but these formulations are based on the Wilson (1949) statistics, which do not account for effects of atomic heterogeneity. The indications of these easily computable ideal statistics are good guides for the resolution of space-group ambiguities, for structures consisting of not too dissimilar atoms - irrespective of the space group. Howells, Phillips & Rogers (1950) first pointed out that the presence of outstandingly heavy atoms may cause misleading indications of Wilson (1949) statistics. This observation was followed by many investigations into the problem of deriving generalized statistics, which allow for the atomic composition as well as for space-group symmetry. Much of the early work on intensity statistics is summarized in the monograph by Srinivasan & Parthasarathy (1976). Later work concentrated on explicit introduction of space-group symmetry into the available generalized intensity statistics (Wilson, 1978; Shmueli, 1979; Shmueli & Wilson, 1981, 1983; Shmueli, 1982; Shmueli & Kaldor, 1981, 1983). In spite of considerable success with these generalizations, the resulting density functions are available as complicated truncated expansions whose properties and degree of convergence are difficult to estimate. Hence they have a limited utility in practice. An important turning point in this line of research was the recognition by Shmueli, Weiss, Kiefer & Wilson (1984) of the fact that the p.d.f. of $|E|$ can be expanded in a Fourier or a Fourier-Bessel series. The resulting exact expressions for the Fourier and Fourier-Bessel p.d.f.'s turned out to have excellent convergence properties and can be readily computed to any accuracy. This technique was used to solve the problem of composition dependence for the space groups $P1$ and $P\bar{1}$ (Shmueli *et al.*, 1984), as well as for other space groups of low symmetry (Shmueli & Weiss, 1987). The Fourier method was also applied to the problem of calculating the p.d.f. when one or more non-crystallographic centres of symmetry are present in the space group $P\bar{1}$ (Shmueli, Weiss & Kiefer, 1985; Shmueli & Weiss, 1985; Shmueli, Weiss & Wilson, 1989) and to the possibility of distribution of the scatterers among variable special positions (*i.e.* lines and planes) in most space groups of the monoclinic and orthorhombic systems (Shmueli & Weiss, 1988). Effects of dispersion in $P\bar{1}$, studied earlier in the central-limit-theorem approximation by Wilson (1980), and generalized by Shmueli & Wilson (1983), have recently been re-examined in the Fourier approach by Shmueli, Rabinovich & Weiss (1990). It

0108-7673/91/040336-05$03.00

follows that an exactly and readily computable p.d.f. for intensity statistics in the space group $P\bar{1}$ can now be constructed, taking into account the atomic composition, partial or complete hypercentrosymmetry, in addition to anomalous scattering. Furthermore, this p.d.f. is also computable to any accuracy.

The desirability of extending these results to higher (or arbitrary) space-group symmetries has been evident for some time. While the extensions of the central limit theorem result (*e.g.* Shmueli & Wilson, 1981; Shmueli & Kaldor, 1981) call for the calculation of symmetry-dependent moments of the trigonometric structure factor, the functional form of the p.d.f. - albeit highly complicated - remaining unchanged, the calculation of the coefficients of Fourier p.d.f.'s becomes more complicated analytically and numerically, as the symmetry increases. This problem was recently examined in some detail and characteristic functions for Fourier and Fourier-Bessel p.d.f.'s were obtained for the 17 plane groups and the first 206 space groups (Rabinovich, Shmueli, Stein, Shashua & Weiss, 1991; hereafter denoted by RSSSW). We wish, in the present paper, to explain how the theoretical results of the preceeding paper, Part VI (RSSSW), apply to actual cases as well as to summarize the conclusions that emerge from an extensive numerical test of these results. Our summary will be illustrated by some typical exact p.d.f.'s with regard to the presence of an outstandingly heavy atom in the asymmetric unit. We recall that by the attribute 'exact' we mean 'computable to any accuracy'.

Calculation of Fourier p.d.f.'s of $|E|$: an example

In order to illustrate the use of the results of RSSSW, we outline the calculation of the exact Fourier representation of the p.d.f. of $|E|$ for the centrosymmetric space group $I4_1/a$, for the set of reflections with odd l. In the absence of significant anomalous dispersion the general expression for the relevant p.d.f. is

$$p(|E|) = \alpha\left\{1 + 2\sum_{u=1}^{\infty} C(\pi\alpha u)\cos(\pi u\alpha|E|)\right\}, \quad (1)$$

where α is the reciprocal of the maximum value of $|E|_{\max}$ and C is the symmetry-dependent Fourier coefficient. This coefficient is the product of the atomic contributions from the asymmetric unit

$$C(\pi\alpha u) = \prod_{j=1}^{N/16} C_j(\pi\alpha u) = \prod_{j=1}^{N/16} L_j(4\pi\alpha u, \pi/4), \quad (2)$$

where N is the number of atoms in the unit cell and $L_j(4\pi\alpha u, \pi/4)$ is the atomic contribution to the Fourier coefficient, for the space group $I4_1/a$ and the subset of reflections with $l = 2n+1$. This contribution is given in Table 2 of RSSSW as the atomic characteristic function $L_j(4\omega_j, \pi/4)$; the Fourier coefficients are derived from the table by replacing ω_1, ω_2 and ω with $\pi\alpha u$, $\pi\alpha v$ and $\pi\alpha(u^2+v^2)^{1/2}$, respectively. It is assumed in the latter derivation that all the atoms are confined to general positions, their contributions to the structure factor are independent and there is no non-crystallographic symmetry. According to RSSSW, the atomic contribution to the Fourier coefficient for the space group and reflection parity considered is given by

$$L_j(4\pi\alpha u, \pi/4) = J_0^4(4\pi\alpha u n_j) + 2\sum_{k=1}^{\infty}(-1)^k J_k^4(4\pi\alpha u n_j), \quad (3)$$

where $J_k(x)$ is the Bessel function of the first kind, of order k (*e.g.* Abramowitz & Stegun, 1972) and n_j is the normalized scattering factor. Only a few terms of the series in (3) are required in order to achieve good convergence. Standard routines for the computation of Bessel functions of any order are readily available from popular mainframe-maintained mathematical libraries (*e.g.* IMSL) and may also be devised for microcomputers (Press, Flannery, Teukolsky & Vetterling, 1986).

In higher tetragonal and all hexagonal space groups, one must make use of numerical integration to calculate values of the atomic contributions to the Fourier coefficients. Such integrals are conveniently computed with the aid of adaptive algorithms [*e.g.* the Romberg method as described by Davis & Rabinowitz (1967)].

All the p.d.f.'s corresponding to the characteristic functions that are given in Table 1 of RSSSW were computed with the aid of a Bessel-function package from the local mainframe library and subroutine *DCADRE* from the mainframe-maintained IMSL library, which employs the Romberg adaptive integration algorithm.

Summary of the calculations

Our calculations are based on an assumed asymmetric unit containing fourteen C atoms and one U atom. For this composition, we first simulated the distribution of $|E|$ (*cf.* Shmueli *et al.*, 1984) using the expression

$$E(\mathbf{h}) = \sum_{j=1}^{N/g} n_j[\xi_j(\mathbf{h}) + i\eta_j(\mathbf{h})], \quad (4)$$

where g is the order of the point group times the multiplicity of the Bravais lattice and $\xi_j + i\eta_j$ is the relevant (possibly) complex trigonometric structure factor. The expressions for ξ and η were taken from *International Tables for X-ray Crystallography* (1965) and from the revised version of the structure factor tables, to appear in Volume B of the new edition of *International Tables for Crystallography* (1991). The number of simulated structure factors, from which the frequency histogram of $|E|$ was constructed, was

kept at 100 000 to ensure a meaningful comparison with the corresponding theoretical p.d.f.'s. These are the exact p.d.f. based on the results of RSSSW and the corresponding (centric or acentric) p.d.f. based on Wilson (1949) statistics. The theoretical p.d.f.'s and the histogram were then put on the same scale and discrepancy factors, an R factor and χ^2, were evaluated as described by Shmueli *et al.* (1984). The above procedure was repeated for each space group, or class of space groups, for which the moments of the trigonometric structure factor have the same value (Shmueli & Kaldor, 1981, 1983). For some non-symmorphic space groups these moments and also the characteristic functions may assume several values or expressions depending on the parity of the reflection indices (Shmueli & Kaldor, 1981, 1983; RSSSW).

We found excellent agreement in the comparison between all the simulated histograms and the corresponding exact p.d.f.'s, the R factors falling in the 0·01–0·02 range. Apart from confirming again that the Fourier method works very well, these comparisons constitute a further check on the correctness of the calculations. There may, however, be large discrepancies between the exact p.d.f.'s, based on the assumed space-group symmetry and composition of the asymmetric unit, and the ideal (centric or acentric) p.d.f.'s based on the central limit theorem (Wilson, 1949). The quantity closest to a real experimental distribution is represented by the histogram of frequencies.

The greatest discrepancy between the ideal and exact p.d.f.'s is found in the case of the two triclinic space groups, which were the first to be investigated and the easiest to treat (Shmueli *et al.*, 1984). We illustrate in Fig. 1 the results obtained for space group $P\bar{1}$. This is a *unique* example of a centric p.d.f. which, because of a very conspicuous peak at an intermediate $|E|$ value, can be mistaken for an ideal acentric one.

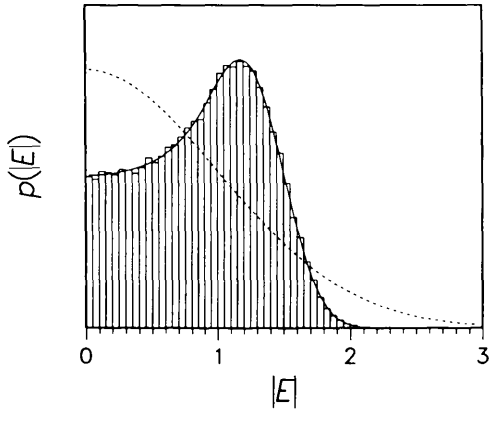

Fig. 1. Comparison of p.d.f.'s of $|E|$ for the space group $P\bar{1}$. The exact (solid line), ideal (dashed line) and simulated (histogram) densities are shown.

A similar but weaker tendency, which is therefore less likely to lead to space-group ambiguities, is found in the exact p.d.f.'s for the centrosymmetric space groups *Fddd* with $h+k=4n$, $k+l=4n$ and $l+h=4n+2$, $I4_1/a$ with $l=2n+1$ and $P\bar{3}$.

Our calculations and their graphical representations also show that exact p.d.f.'s of a number of non-centrosymmetric space groups tend, in a variety of ways, to the ideal centric p.d.f. and are thus relevant to the problem of space-group ambiguities. These space groups or classes of space groups are *Pmm*2 (*Cmm*2), *P*4 ($P4_1$ with $l=2n$, $I4_1$ with $l=2n+1$), *P*4*mm* ($I4_1md$ with $2k+l=2n$), $P\bar{6}$, $P6_3cm$ with $l=2n+1$, *P*6*cc* with $l=2n$ and $P\bar{6}2c$ (both parities of l). The statistics of the space groups *Pmm*2 (this includes all the space groups isomorphous to the point group *mm*2) and *P*6*cc* are illustrated in Fig. 2. An important qualification must be made: the p.d.f.'s in Fig. 2, as well as the others mentioned in the same

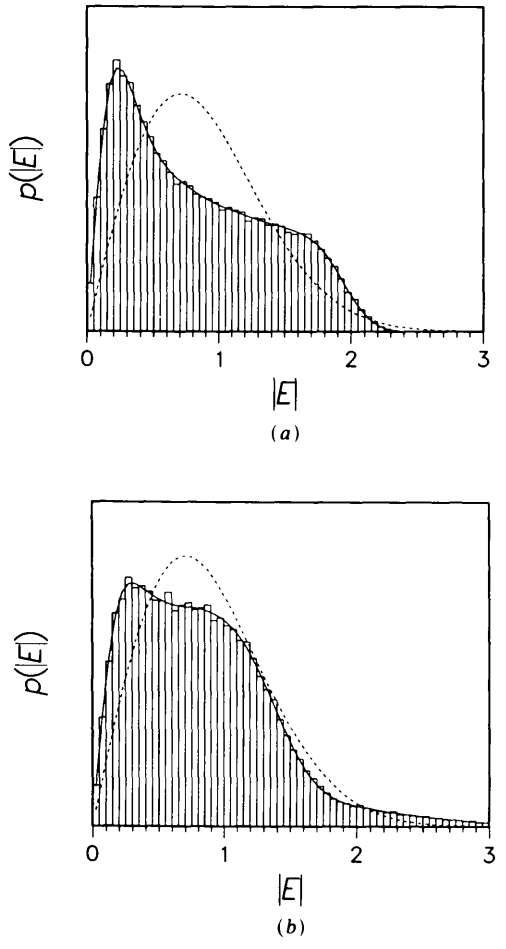

Fig. 2. Comparison of p.d.f.'s of $|E|$ for the space groups *Pmm*2 and *P*6*cc*. The exact (solid line), ideal (dashed line) and simulated (histogram) densities are shown. (*a*) *Pmm*2, (*b*) *P*6*cc*, $l=2n$.

connection, start off from zero but have the peak displaced to a small value of $|E|$. Since the treatment of the intensities of very weak reflections is usually rather uncertain, the distribution of very small $|E|$'s may often be distorted or underpopulated irrespective of whether the crystal is centrosymmetric or not. The low end of the distribution is therefore less reliable for the purpose of comparison.

The p.d.f.'s of a number of centrosymmetric space groups resemble somewhat a bicentric one (Shmueli, Weiss & Kiefer, 1985). An indication of such a tendency was already noted by Shmueli & Wilson (1981), who examined the fourth moment of $|E|$ for a series of space groups and noticed that for some space groups such moments were significantly greater than 3 but always smaller than 4·5 - the fourth moment for an ideal bicentric distribution. The space groups or space-group classes belonging to this category are *Pmmm* (*Cmmm*, *Fddd* with $h+k=4n$ *etc.*), $P4/m$, $I4_1/a$ with $l=2n$, $P4/mmm$, $I4_1/amd$ with $l=2n$, $P6_3/m$ with $l=2n+1$, $P6/mcc$ (both parities of l) and $P6_3/mcm$ with $l=2n+1$. The p.d.f.'s for the cubic space groups $Pn\bar{3}n$ with $h=2n$, $k=2n$, $l=2n+1$, $Fm\bar{3}c$ with $h+k+l=2n+1$ and $Ia\bar{3}d$ with $h=2n$, $k=2n$, $l=2n$, $h+k+l=4n+2$ also show a similar although less emphasized tendency to hypersymmetry. The statistics for $P6/mcc$ with $l=2n+1$, shown in Fig. 3, illustrate this behaviour.

The exact p.d.f.'s of several space groups appear to be almost insensitive to atomic heterogeneity. This is reflected in a good to excellent agreement of the exact and ideal p.d.f.'s even for the highly heterogeneous $C_{14}U$ asymmetric unit. The space group or space-group classes belonging to this category are all those isomorphous to the point groups 222 and $\bar{4}$, $I4_1/amd$ with $l=2n+1$, space groups related to the point group 32, $P6_222$ with $l=6n, 6n+1$ and $6n+2$, and space groups isomorphous to the

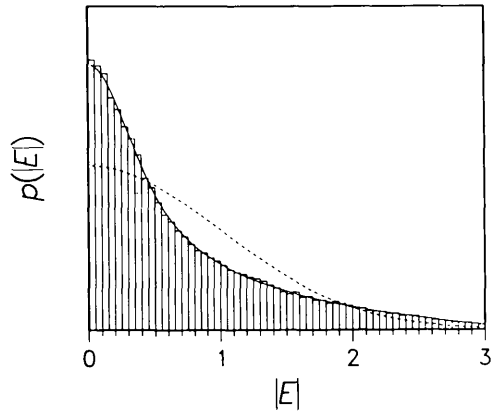

Fig. 3. Comparison of p.d.f.'s of $|E|$ for the space group $P6/mcc$, $l=2n+1$. The exact (solid line), ideal (dashed line) and simulated (histogram) densities are shown.

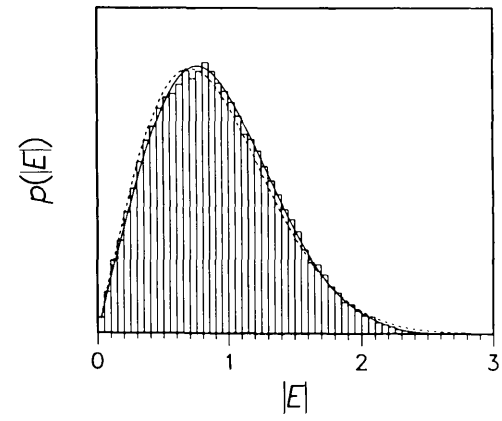

Fig. 4. Comparison of p.d.f.'s of $|E|$ for the space group $P312$. The exact (solid line), ideal (dashed line) and simulated (histogram) densities are shown.

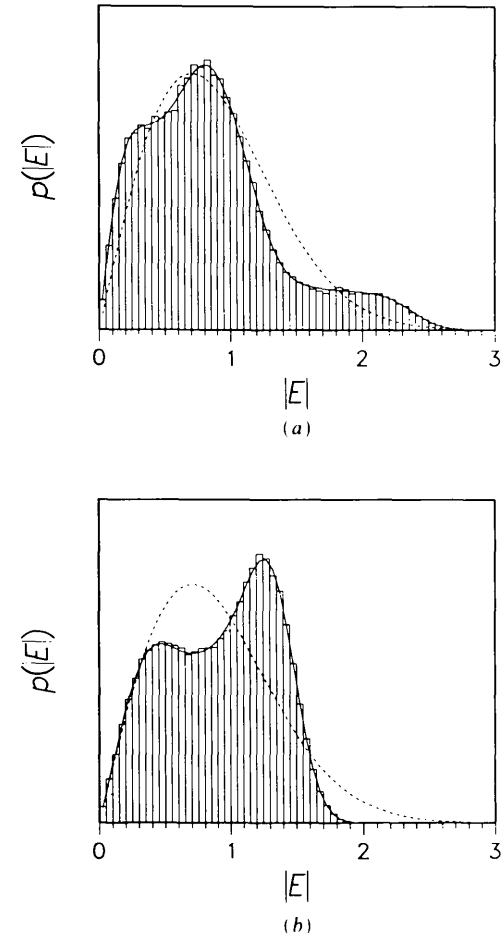

Fig. 5. Comparison of some p.d.f.'s of $|E|$ for the space group $P6_1$. The exact (solid line), ideal (dashed line) and simulated (histogram) densities are shown. (*a*) $P6_1$, $l=6n+1$, (*b*) $P6_1$, $l=6n+2$.

cubic point groups 23 and $m\bar{3}$. An example of such insensitivity is illustrated in Fig. 4 by the statistics for the trigonal $P312$.

It should also be pointed out that within some non-symmorphic space groups the exact p.d.f.'s may differ widely, depending on the parity of the reflection indices. The statistics for two out of the four different reflection subsets for the space group $P6_1$ are displayed in Fig. 5.

A remark about the statistics of the cubic space groups nos. 207–230, including the various subsets, is in order. Exact p.d.f.'s were not formulated in a final form for these space groups since the expressions appeared rather unwieldy and indicated excessive computing effort that might be called for in their evaluation. Moreover, comparisons of histograms of $|E|$ for a $C_{14}U$ asymmetric unit with the appropriate ideal p.d.f.'s showed that they (the histograms) are either insensitive to atomic heterogeneity or display a weak tendency to hypersymmetry. It might perhaps be interesting to examine the feasibility of constructing, in such instances, reliable approximate p.d.f.'s by the Gram–Charlier correction-factor approach (*e.g.* Shmueli & Wilson, 1981; Shmueli, 1982) since the departures from ideal behaviour are here rather small.

This research was supported in part by grant no. 88-00210 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. All the computations related to this paper were carried out at the Tel Aviv University Computation Center on a Cyber 180-990 computer.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.
DAVIS, P. J. & RABINOWITZ, P. (1967). *Numerical Integration.* London: Blaisdell.
HOWELLS, R. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210–214.
International Tables for Crystallography (1991). Vol. B, edited by U. SHMUELI, Ch. 1.4. In the press.
International Tables for X-ray Crystallography (1965). Vol. I, edited by N. F. M. HENRY & K. LONSDALE. Birmingham: Kynoch Press.
PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes.* Cambridge Univ. Press.
RABINOVICH, S., SHMUELI, U., STEIN, Z., SHASHUA, R. & WEISS, G. H. (1991). *Acta Cryst.* A**47**, 328–335.
SHMUELI, U. (1979). *Acta Cryst.* A**35**, 282–286.
SHMUELI, U. (1982). *Acta Cryst.* A**38**, 362–371.
SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* A**37**, 76–80.
SHMUELI, U. & KALDOR, U. (1983). *Acta Cryst.* A**39**, 619–621.
SHMUELI, U., RABINOVICH, S. & WEISS, G. H. (1990). *Acta Cryst.* A**46**, 241–246.
SHMUELI, U. & WEISS, G. H. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 53–66. Guilderland: Adenine Press.
SHMUELI, U. & WEISS, G. H. (1987). *Acta Cryst.* A**43**, 93–98.
SHMUELI, U. & WEISS, G. H. (1988). *Acta Cryst.* A**44**, 413–417.
SHMUELI, U., WEISS, G. H. & KIEFER, J. E. (1985). *Acta Cryst.* A**41**, 55–59.
SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* A**40**, 651–660.
SHMUELI, U., WEISS, G. H. & WILSON, A. J. C. (1989). *Acta Cryst.* A**45**, 213–217.
SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 342–353.
SHMUELI, U. & WILSON, A. J. C. (1983). *Acta Cryst.* A**39**, 225–233.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
WILSON, A. J. C. (1978). *Acta Cryst.* A**34**, 986–994.
WILSON, A. J. C. (1980). *Acta Cryst.* A**36**, 945–946.

Acta Cryst. (1991). A**47**, 340–345

On Integrating the Techniques of Direct Methods with Anomalous Dispersion. II. Statistical Properties of the Two-Phase Structure Invariants

BY D. Y. GUO, ROBERT H. BLESSING AND HERBERT A. HAUPTMAN

Medical Foundation of Buffalo, 73 High Street, Buffalo, New York 14203, *USA*

(*Received* 10 *July* 1990; *accepted* 15 *January* 1991)

Abstract

Results of a statistical study of probabilistic estimates of two-phase structure invariants (TPSI) for Friedel pairs in the case of single-wavelength anomalous scattering are reported. Numerical analysis of the TPSI sign, magnitude and error distributions shows that the concise formula for TPSI by probability theory [Hauptman (1982). *Acta Cryst.* A**38**, 632–641; Giacovazzo (1983). *Acta Cryst.* A**39**, 585–592] has desirable statistical properties. Computational results for the known structures of cocaine methiodide (*N*-methylcocaine iodide) and of cytochrome c_{550} and its $PtCl_4^{2-}$ derivative show that when $|E|$ values are large most of the signs of the TPSI are correctly determined – for $|E| > 1{\cdot}0$, 90% or more of the TPSI signs

0108-7673/91/040340-06$03.00