Acta Cryst. (1983). A**39**, 882–892

Least-Squares Analyses of Diffuse Scattering from Substitutionally Disordered Molecular Crystals: Application to 2,3-Dichloro-6,7-dimethylanthracene

BY JOEL EPSTEIN AND T. R. WELBERRY*

Research School of Chemistry, Australian National University, Canberra, 2600, *Australia*

(*Received* 5 *April* 1983; *accepted* 28 *June* 1983)

Abstract

Diffuse intensities of X-rays scattered from substitutionally disordered molecular crystals have been measured from Weissenberg photographs and analysed using linear least-squares methods to determine the short-range correlations between molecular sites. The measured diffuse scattering was represented as a linear combination of calculated random and correlation distributions. This procedure is completely general and makes no assumptions about correlations between non-nearest neighbours. Of the 74 correlation coefficients describing the short-range order in 2,3-dichloro-6,7-dimethylanthracene, 35 were found to be significant, 13 at the 3σ level. The largest coefficients within each correlation neighbourhood were +0·46 (7), +0·37 (4) and −0·24 (5) involving six of the nearest-neighbour sites.

Introduction

In previous papers we have described our interest in disordered molecular crystals, and a number of studies have been reported (Welberry & Jones, 1980; Jones & Welberry, 1980; Welberry, Jones & Epstein, 1982; Epstein, Welberry & Jones, 1982). A method of recording and measuring diffuse scattering data has been developed (Welberry, 1983) and a combination of optical diffraction analogue experiments (Welberry & Jones, 1980) together with direct calculations (Epstein, Welberry & Jones, 1982) has been used to interpret the results. To date, interpretation of the diffuse scattering patterns has been by visual comparison with either optical transforms of a computer-generated model or contour maps of calculated intensity distributions. Using these methods, only semi-quantitative estimates (at best) of short-range intermolecular correlation coefficients could be obtained and the number of distinct correlations that could be assigned was limited. In this paper we describe the development of a least-squares procedure which allows quantitative determination of such parameters, using numerical measurements of diffuse scattering from X-ray films, and calculated diffuse patterns which assume only a prior knowledge of the 'average' crystal structure obtained by conventional Bragg diffraction. In this work, no assumption is made of the way in which correlations between more distant sites depend on nearest-neighbour correlations. Such assumptions were a feature of the previous studies by us and by other workers (Flack, 1970; Glazer, 1970).

The molecular crystals being studied are derivatives of anthracene or benzene in which disorder arises because of the similar packing volumes of halogen substituents (Cl or Br) and methyl substituents. At a given substituent site the halogen or methyl groups may be interchanged. In the examples studied to date, each molecular site may be occupied by the molecule in one of two possible orientations, *A* or *B*. These two orientations differ essentially only in the disposition of the methyl and halogen substituents, the aromatic nucleus of the molecule remaining fixed. Such disorder is 'locked-in' at growth since the energy barriers to subsequent reorientation are relatively high. For the purposes of the present work we have chosen the example of 2,3-dichloro-6,7-dimethylanthracene (Welberry, Jones & Puza, 1983). For this compound each molecular site has four sites of substituent disorder. In previous work we have found significant short-range ordering when disordered sites on neighbouring molecules are in close contact. We might therefore expect that the correlation structure for this compound would be more complex than our earlier examples which had only two disordered sites per molecule. In this paper we use this example primarily to illustrate the methods described; a more detailed account of the disorder properties of 2,3-dichloro-6,7-dimethylanthracene will be published elsewhere.

Diffuse scattering from a disordered molecular crystal

If thermal diffuse scattering is neglected, the intensities of X-rays scattered by a substitutionally disordered molecular crystal consist of sharp peaks at the

* To whom correspondence should be addressed.

0108-7673/83/060882-11$01.50

reciprocal-lattice positions, corresponding to Bragg scattering from an ordered crystal of 'average' molecules, and broad diffuse scattering elsewhere in reciprocal space, resulting from the substitutional disorder.

From the analysis of the Bragg intensities, the site occupancies of the components of the crystal (*i.e.* the one-body distribution function of the crystal) may be determined, as well as the structural and vibrational parameters of the individual components. From the diffuse intensities the short-range correlations between the sites (*i.e.* the two-body distribution function) may be determined. The diffuse scattering of X-rays from a two-component substitutionally disordered molecular crystal may be represented by (Epstein, Welberry & Jones, 1982)

$$I_{\rm diff}(\mathbf{s}) = Nm_A m_B \sum_{n}^{\rm cell} \sum_{m} C_{nm}^{AB} \Delta F_n(\mathbf{s}) \Delta F_m^*(\mathbf{s}), \quad (1)$$

where the magnitude of the scattering vector $\mathbf{s}$ is $4\pi \sin\theta/\lambda$; N is the number of unit cells; m_A is the concentration of molecules of type A; C_{nm}^{AB} is a correlation coefficient, related to the conditional probability that site n of the crystal is occupied by a molecule of type A *given that* site m is occupied by a molecule of type B; and $\Delta F_n(\mathbf{s})$ is the difference between the vibrationally averaged form factors of a type-A molecule and a type-B molecule at site n.

$$\Delta F_n(\mathbf{s}) = \sum_{\alpha} f_{n\alpha}^{(A)}(\mathbf{s}) T_{n\alpha}^{(A)}(\mathbf{s}) \exp\left(i\mathbf{s}.\mathring{\mathbf{R}}_{n\alpha}^{(A)}\right) - \sum_{\beta} f_{n\beta}^{(B)}(\mathbf{s}) T_{n\beta}^{(B)}(\mathbf{s}) \exp\left(i\mathbf{s}.\mathring{\mathbf{R}}_{n\beta}^{(B)}\right). \quad (2)$$

In (2), $f_{n\alpha}^{(A)}(\mathbf{s})$ is the scattering factor for pseudoatom α of the molecule of type A at site n; $T_{n\alpha}^{(A)}(\mathbf{s})$ is the usual temperature factor for this pseudoatom; and $\mathring{\mathbf{R}}_{n\alpha}^{(A)}$ is the vector from some origin to the equilibrium position of the corresponding nucleus.

The sum over n of (1) includes all molecular sites within the unit cell. For each site n, the sum over m includes all molecular sites within the correlation neighbourhood of this site (Epstein, Welberry & Jones, 1982).

$I_{\rm diff}(\mathbf{s})$ of (1) consists of two types of contributions: terms with $m = n$ ($C_{nm}^{AB} = 1$) make up the random distribution, corresponding to the diffuse X-ray scattering from a crystal in which the arrangement of A-type and B-type molecules is random; and terms for which $m \neq n$, corresponding to correlation distributions. For each member m of the correlation neighbourhood of site n, the pair of sites m and n, together with symmetry-related pairs of sites, give rise to a correlation distribution. The contribution from this distribution to $I_{\rm diff}(\mathbf{s})$ of (1) is weighted by the correlation coefficient C_{nm}^{AB}. The correlation coefficients between symmetry-related pairs of molecular sites are assumed to be equal.

The random and correlation distributions provide the basis functions for the analysis of the diffuse scattering. When no correlation exists between molecular sites (all correlation coefficients are zero) the diffuse scattering is that described by the random distribution only. Each distinct correlation distribution modulates this intensity distribution. Within each interference fringe of the random distribution, a particular correlation distribution re-distributes intensity, removing intensity from regions in which the correlation distribution is negative and adding intensity to regions in which the correlation distribution is positive. These modulations are characteristic of the *inter*molecular vectors between two sites and provide a means of determining the extent of correlation between the two sites. Examples of random and correlation distributions for 9-bromo-10-methylanthracene have been given previously (Epstein, Welberry & Jones, 1982). Further examples for 2,3-dichloro-6,7-dimethylanthracene are given later in this paper.

Measurement of diffuse scattering

Previous work in which the diffuse X-ray patterns were visually compared to optical or calculated patterns utilized the X-ray data in the form of an image on a photographic film. This was obtained in the manner described by Welberry (1983) from long-exposure Weissenberg photographs with the use of an Optronics P1700 photomation scanner/writer. Before the final image was printed using this device an empirical background correction was applied to remove extraneous intensity due to air scattering, fluorescence, Compton scattering, scattering from glass fibre and glue, and general film fog. In addition, the resultant optical density readings attributed to the disorder scattering were scaled by an enhancement factor to yield a suitably intense photographic image. For the present numerical work essentially the same procedure was adopted, although since no advantage is gained by using an enhancement factor, the intensity values were merely put on a convenient scale.

A number of aspects of the data collection are worth a special note. Firstly, it is apparent from the diffuse scattering pictures shown in Figs. 2(*a*), 3(*a*), 4(*a*) and 5(*a*) that in addition to the required diffuse scattering a number of other features are present in the reciprocal-lattice section. The Bragg peaks themselves must be avoided together with various streaks emanating from them (usually due to strain or powder caused by sample shaping). Similarly, Bragg peaks often have regions of strong thermal diffuse scattering around them due to acoustic phonons, and there may also be more general regions of thermal diffuse scattering which must also be avoided. In the present work we do not attempt to account for thermal diffuse scattering but avoid the necessity to consider it by using only data

obtained from regions in reciprocal space where the thermal scattering is assumed to be minimal. In all, for each reciprocal-lattice section, the intensity was measured at about 5000 reciprocal points. From these data 2000 were selected using the above criteria for use in the analysis.

Secondly, it was found that the empirical background correction applied to the data for previous semi-quantitative photographic work was not adequate for the present purposes although it was considered a useful first approximation. An additional correction was made to the intensities according to the following stratagem. We rely on the fact that, for the disorder scattering in which we are interested, the presence of correlations merely modifies the random distribution and will not affect regions where this distribution is zero. The random distribution may be calculated from a knowledge of the average structure and so the regions in reciprocal space where disorder diffuse scattering should be zero may be calculated at the outset. Preliminary tests showed that while the observed intensities invariably went through a minimum close to the calculated zeros of the random distribution the observed intensity values were consistently non-zero. This intensity must be additional scattering not due to substitutional disorder. Since it was a slowly varying function of reciprocal-space position it was treated as an additional background correction.

Another consideration was one of resolution. The use of conventional Weissenberg photographs introduces a resolution function which was discussed by Welberry (1983). Since this effect is a minimum at low angles we used only data for which $\theta < 25°$ (note in previous studies we have used data to $\theta = 50°$ for Cu $K\alpha$ radiation). (It should be noted also that, in general, thermal diffuse scattering is more pronounced at high angles.)

Least-squares analysis

The random and correlation distributions may be calculated using the positional and vibrational parameters determined from the analysis of the Bragg intensities. These distributions, calculated at the reciprocal-space points at which data are measured, form the set of basis functions $\{D_l(\mathbf{s})\}$, in terms of which the calculated diffuse intensities are described.

$$I^{(c)}_{\rm diff}(\mathbf{s}_i) = \sum_l x_l D_l(\mathbf{s}_i). \tag{3}$$

In (3), x_l is a *distribution coefficient*, related to the correlation coefficient by

$$x_l = k\, C^{AB}_{nm}, \tag{4}$$

where k is the product of the constant terms of (1) and a scale factor required to relate the calculated diffuse intensities to the relative intensities obtained from experiment.

The distribution coefficients may be treated as variable parameters to be determined by minimizing

$$\varepsilon = \sum_i w_i [I^{(o)}_{\rm diff}(\mathbf{s}_i) - I^{(c)}_{\rm diff}(\mathbf{s}_i)]^2, \tag{5}$$

where the weighting factor w_i is the inverse of the estimated variance of $I^{(o)}_{\rm diff}(\mathbf{s}_i)$.

Since $C^{AB}_{nn} = 1$ the coefficient x_r for the random distribution may be used to determine the scale factor of (4). The correlation coefficients may then be determined directly from the ratios (x_l/x_r).

When diffuse intensities are recorded using the Weissenberg camera, data from different sections of reciprocal space are measured separately. These data may be combined, using the intensities measured at common reciprocal-space points, to determined relative scaling factors between different sections. Alternatively, each section of reciprocal space may be analysed separately. For the analysis of 2,3-dichloro-6,7-dimethylanthracene, the latter procedure was adopted. Some difficulties arise with this procedure and these will be discussed below.

In general, not all of the calculated distributions for a particular section of reciprocal space are independent. For example, in the $h0l$ section, the correlation distribution corresponding to the pair of sites n and m is identical to that for the pair of sites n and m' when sites m and m' are related by a cell translation along the b axis. For these pairs of sites, however, C^{AB}_{nm} will not in general be equal to $C^{AB}_{nm'}$. The distribution coefficients corresponding to the independent distributions in a particular section of reciprocal space will therefore generally be related to a linear combination of several correlation coefficients, *i.e.*

$$x_l = k(C^{AB}_{nm} + \alpha C^{AB}_{nm'} + \beta C^{AB}_{nm''} + \ldots). \tag{6}$$

When analysing data from each reciprocal-space section separately, the distribution coefficients may still be treated as linear parameters, to be determined by minimizing the least-squares residual for that section. The correlation coefficients, however, are no longer determined directly from the distribution coefficients. An additional step is required to solve the equations relating the distribution coefficients to the correlation coefficients.

Using the set of distribution coefficients obtained from the analysis of a particular section of reciprocal space, the diffuse intensities calculated on a fine grid over the entire section may be significantly negative in regions where no data can be measured, such as near the origin of reciprocal space. To ensure a physically meaningful least-squares solution, the calculated diffuse intensities should then be constrained to be non-negative at selected points in reciprocal space, *i.e.* at $\mathbf{s}_j$,

$$\sum_l x_l D_l(\mathbf{s}_j) \geq 0. \tag{7}$$

Minimizing ε of (5) subject to the linear inequality constraints of (7) is formally a problem in quadratic programming. The solution may be obtained using an active set strategy (Gill & Murray, 1974). A vector $\mathbf{X}_0 = (x_1^{(0)}, x_2^{(0)}, \ldots, x_l^{(0)}, \ldots)$ is first found which satisfies the constraints of (7) (*i.e.* $\mathbf{X}_0$ is a feasible vector). For the analysis of the diffuse scattering, a convenient feasible vector is $\mathbf{X}_0 = (x_r, 0, 0, \ldots, 0)$, where x_r is the coefficient of the random distribution obtained from the unconstrained minimization of ε. Since the random distribution is non-negative everywhere, the constraints of (7) are necessarily satisfied. The vector $\mathbf{X}_0$ is then updated by proceeding towards the unconstrained minimum until one of the constraints is no longer satisfied. This constraint is made active (*i.e.* the *inequality* constraint replaced by an *equality* constraint) and a new minimum subject to this constraint found. This procedure is repeated, successively activating constraints as required. When no further constraints are required to be activated, constraints which are no longer necessary are deleted from the active set and the procedure again repeated. The final solution vector is obtained at the minimum of ε with fewest constraints active.

For each section of reciprocal space the set of independent distribution coefficients which satisfy the appropriate non-negativity constraints may be obtained. Estimates of the standard deviations of the coefficients and their correlations may be obtained from the variance–covariance matrix corresponding to the final set of active constraints. Because of the non-uniqueness of correlation distributions in a particular reciprocal-space section, the correlation coefficients cannot, in general, be determined from the ratios of distribution coefficients (x_l/x_r) as discussed earlier. To each distribution coefficient there corresponds an equation such as (6). More equations will, in general, be obtained than there are unknown correlation coefficients. These equations may then be solved approximately by least squares.

The unknowns in equations such as (6) are the scale factors (one for each reciprocal-space section) and the correlation coefficients. To solve the equations approximately, non-linear least-squares techniques may be used. Alternatively (several) equations may be selected and the corresponding estimates of the distribution coefficients used to eliminate the scale factors. In this way reliable approximate values for the unknown correlation coefficients may be obtained using *linear* least squares. The equation for the random distribution coefficient for a particular section is a convenient choice for eliminating the scale factor for that section. This equation may be, for example,

$$x_r = k(1 + C_4 + C_5). \tag{8}$$

The use of (8) to eliminate the scale factor of (6) results in the new equation

$$x'_l = C^{AB}_{nm} + \alpha C^{AB}_{nm'} + \beta C^{AB}_{nm''} + \ldots - x'_l C_4 - x'_l C_5, \tag{9}$$

where the *reduced distribution coefficient* x'_l is defined by

$$x'_l = x_l/x_r. \tag{10}$$

Note that the reduced distribution coefficient x'_l appears on both sides of (9). To solve these equations for the correlation coefficients, the reduced distribution coefficients on the right-hand side of equations such as (9) are replaced by the corresponding estimates, $\hat{x}'_l$, obtained from the constrained least-squares analyses of the measured diffuse intensities, *i.e.*

$$x'_l = C^{AB}_{nm} + \alpha C^{AB}_{nm'} + \beta C^{AB}_{nm''} + \ldots - \hat{x}'_l C_4 - \hat{x}'_l C_5. \tag{11}$$

The variance of the coefficient $\hat{x}'_l$ appearing on the right-hand side of (11) is neglected. The correlation coefficients are then linear parameters in equations such as (11) and may be determined by minimizing

$$\varepsilon = \sum_l \sigma_l^{-2}(\hat{x}'_l - x'_l)^2, \tag{12}$$

where σ_l^2 is the estimated variance of $\hat{x}'_l$. Estimates of the standard deviations of the correlation coefficients and their correlations may be obtained from the resulting least-squares variance–covariance matrix. Note that covariances between estimates of the reduced distribution coefficients $\hat{x}'_l$ are neglected in this procedure.

Example analysis: 2,3-dichloro-6,7-dimethylanthracene

Observed diffuse scattering

Crystals of 2,3-dichloro-6,7-dimethylanthracene were grown by vacuum sublimation. The average structure had space group $P2_1/c$ with four molecules per unit cell (Welberry, Jones & Puza, 1983). Within the resolution of the measured Bragg intensities, the anthracene frame was ordered, but the chlorine and methyl substituents were disordered. At each site in the crystal, the molecule may assume one of two orientations; in one (labelled orientation *A*) the Cl substituents are attached to the carbons at positions 2 and 3, while in the other (labelled orientation *B*) the Cl substituents are attached to the carbons at positions 6 and 7. From the analysis of the Bragg intensities the partial occupancy for molecules in orientation *A*, m_A, was determined to be 0·60, with $m_B = 1 - m_A = 0{\cdot}40$.

The observed intensity distributions for $\sin\theta/\lambda < 0{\cdot}28$ Å^{-1} are shown for the $0kl$, $h0l$, $h1l$ and hhl sections in Figs. 2(*a*), 3(*a*), 4(*a*) and 5(*a*), respectively. These figures represent data which were first recorded on Weissenberg photographs, then undistorted and enhanced, as described by Welberry (1983). Measurements of diffuse scattering were restricted to regions away from suspected thermal diffuse scattering (indicated on the figures) and away from Bragg peaks. In

the $0kl$ section, for example, measurements were restricted to the region $0 < k < 1{\cdot}5$ and $0 < l < 7{\cdot}0$ as shown in Fig. 2(*a*).

Calculation of random distributions

At each molecular site, the pseudoatoms of the anthracene frame are assumed to be identical in orientation *A* and orientation *B*. Therefore only the methyl and chlorine substituents contribute to the disorder diffuse scattering. From the analysis of the Bragg intensities the equilibrium position of the methyl carbon of a type-*A* molecule could not be resolved from the position of the corresponding chlorine of a type-*B* molecule. For the calculations of the random distributions, idealized positions, rather than these averaged positions were assumed. The idealized positions were calculated from previously determined Cl–C(aromatic) and CH_3–C(aromatic) internuclear distances as discussed by Jones & Welberry (1980) for disordered crystals of 9-bromo-10-methylanthracene. The idealized positions for the sublimation-grown crystals of 2,3-dichloro-6,7-dimethylanthracene are given in Table 1.

Though anisotropic vibrational parameters for non-hydrogen nuclei were refined in the analysis of the Bragg intensities, equivalent isotropic vibrational parameters (approximately 0·05 Å^2 at each site) were used for the calculations of diffuse scattering.

From these equivalent positions and vibrational parameters, neglecting the scattering from the hydrogen pseudoatoms of the methyl substituents, $\Delta F_n(\mathbf{s})$ of (2) was calculated and the random distributions computed. Analytic approximations for the chlorine and methyl-carbon scattering factors were taken from Table 2.2*B* in *International Tables for X-ray Crystallography* (1974).

The calculated random distributions for the $0kl$, $h0l$, $h1l$ and hhl sections are shown in Figs. 2(*c*), 3(*c*), 4(*c*) and 5(*c*), respectively. Neglecting the effects of thermal diffuse scattering, differences between these distributions and the corresponding observed diffuse intensities indicate the effects of correlations between the molecular sites of the disordered crystal.

Correlation neighbourhoods

The sites at x,y,z, $\bar{x},\bar{y},\bar{z}$, $\bar{x}$, $y+\frac{1}{2}$, $\bar{z}+\frac{1}{2}$ and $x,\bar{y}+\frac{1}{2}$, $z+\frac{1}{2}$ were labelled by the designation codes 55501, 55502, 55503 and 55504, respectively (Epstein, Welberry & Jones, 1982). The correlation neighbourhood for site 55501 included 120 sites. The orientations of *A*-type molecules in five of the sites of the correlation neighbourhood, 65503, 55503, 65504, 55504 and 66602 are shown in Fig. 1. The diffuse scattering in each section of reciprocal space reveals details about the corresponding projection of the correlation neighbourhood. For the $0kl$, $h0l$ and hhl sections, Figs. 2(*d*), 3(*d*) and 5(*d*) show the appropriate projections of the neighbourhood of 55501. All molecules are shown in the *A*-type orientation. The extent of the neighbourhood was determined by including surrounding molecular sites up to two cell translations along axes **a**,**b** and **c** from the site at 55501. The neighbourhoods for the other three sites in the unit cell are related by symmetry to the neighbourhood for 55501.

Table 1. *Idealized positions (fractional coordinates) for the chlorine and methyl-carbon substituents of* 2,3-*dichloro*-6,7-*dimethylanthracene*

	Molecule type *A*				Molecule type *B*		
	x	*y*	*z*		*x*	*y*	*z*
Cl(1)	0·4162	0·2332	0·4680	C(15)	0·4184	0·2394	0·4598
Cl(2)	0·8446	0·4144	0·4431	C(16)	0·8157	0·4074	0·4366
C(15)	−0·2628	0·2408	0·0935	Cl(1)	−0·2932	0·2347	0·0871
C(16)	0·1358	0·4101	0·0702	Cl(2)	0·1374	0·4171	0·0620

The 120 sites in the correlation neighbourhood of 55501 represent the maximum possible number of distinct correlation distributions used for analysis. Because of the assumption that correlations between symmetry-related pairs of sites are equal, each of these distributions has the space-group symmetry. However, two distinct neighbours may give rise to correlation distributions which are identical in all reciprocal space. This occurs when the internuclear vector from 55501 to site *m* of the neighbourhood is related by symmetry to the corresponding vector to some other site, *m'*. The correlation coefficients between 55501 and these two sites are also assumed to be equal although this is not required from consideration of the symmetry of the observed diffuse scattering. For example, the coefficient between 55501 and 65503 was assumed to be equal to that between 55501 and 64503. In this way 74 distinct coefficients were used to describe the correlations to the 120 sites of each neighbourhood.

Calculation of correlation distributions

The correlation distributions were calculated as described for the random distribution. Examples of correlation distributions for the four sections are shown

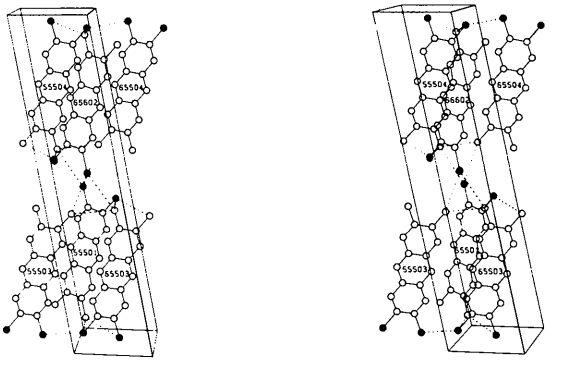

Fig. 1. Stereoview of several molecules in the correlation neighbourhood of site 55501. All molecules are shown in the *A*-type orientation. Chlorine substituents are indicated by filled circles. Cl/Me contacts less than 4 Å are shown by dashed lines.

in Figs. 2(*e*), 2(*f*), 3(*e*), 3(*f*), 4(*d*), 5(*e*) and 5(*f*). For the 0*kl* section, the correlation distribution for site 65503 (Fig. 2*f*) removes intensity from the central regions of the fringes of the random distribution and adds intensity to the outer regions. The superposition of this correlation distribution and the random distribution would produce an intensity distribution which agrees qualitatively with some of the features of the observed intensity distribution, suggesting that the correlation between sites 55501 and 65503 is significant. For the *h*0*l* section, the correlation distribution for site 65504 (Fig. 3*e*) re-distributes the intensity of the random distribution in qualitative agreement with features of the observed distribution, indicating that the correlation between 55501 and 65504 is also expected to be significant.

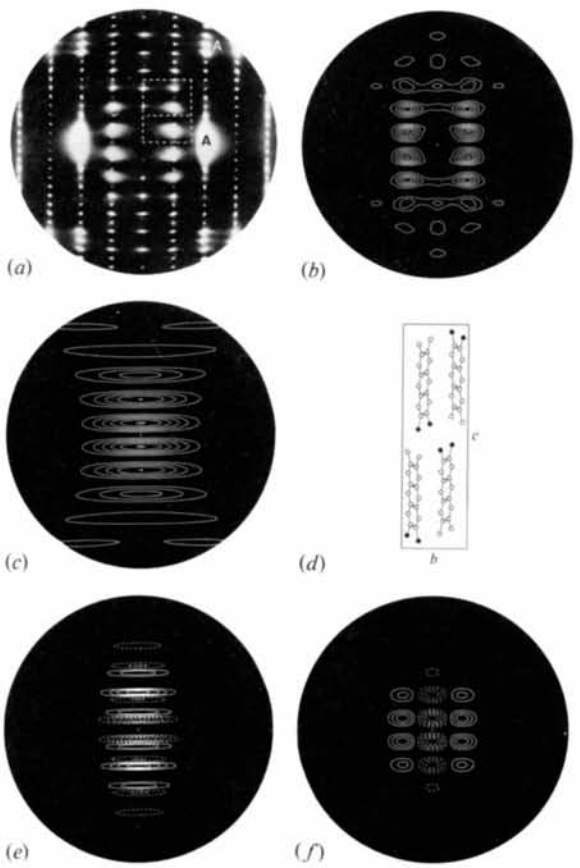

Fig. 2. Diffuse scattering for the 0*kl* section. (*a*) Observed diffuse scattering. Regions of thermal diffuse scattering are labelled by the letter *A*. Analysis of data measured within the region bounded by dashed lines was carried out using least squares as discussed in the text. (*b*) Calculated diffuse scattering after least-squares analysis. Contour level ~85 intensity units. (*c*) Calculated random distribution. Contour level ~50. (*d*) *a*-axis projection of an ordered representation of the structure. Chlorines are indicated by filled circles. (*e*) Calculated correlation distribution for site 65504 (within correlation neighbourhood of 55501). Dashed contours are negative, contour level ~50. (*f*) Correlation distribution for site 65503. Contours as in (*e*).

As discussed earlier, not all of the 75 distributions (the random distribution and 74 correlation distributions) are independent for each section of reciprocal space. For the 0*kl*, *h*0*l* and *h*1*l* sections, determining which distributions are identical is straightforward. For example, the correlation distributions for 66602, 76602, 56602 and 86602 are identical in the 0*kl* section. The distribution coefficient for the 66602 distribution therefore represents the sum of four distinct correlation coefficients. For the *hhl* section, however, the relationship between distributions is less obvious. From (1) and (2), the various distributions may be represented in terms of the sum of cosines involving internuclear vectors. Let **R** be the internuclear vector between two particular pseudoatoms, one belonging to the molecule at site 55501, and the other belonging to the molecule at site 86602. The correlation distribution for 86602 then involves the terms

$$\cos(\mathbf{s}.\mathbf{R}) + \cos(\mathbf{s}.\mathbf{R}'), \qquad (13)$$

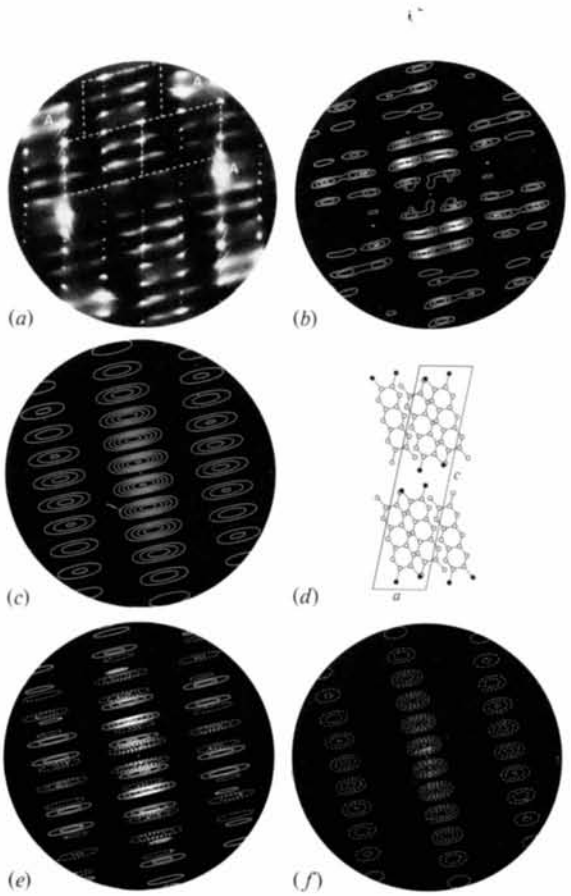

Fig. 3. Diffuse scattering for the *h*0*l* section. (*a*) Observed diffuse scattering. Features marked as in Fig. 2(*a*). (*b*) Calculated diffuse scattering after least-squares analysis. Contour level ~55. (*c*) Random distribution (*d*) *b*-axis projection of an ordered representation of the structure. (*e*) Correlation distribution for site 65504. (*f*) Correlation distribution for site 65503. (*c*)–(*f*) as described in Fig. 2.

where $\mathbf{R}'$ is related to $\mathbf{R}$ by the twofold screw axis of space group $P2_1/c$. The correlation distributions for 66602, 77602 and 75602 then involve the following terms

$$66602: \cos[\mathbf{s}.(\mathbf{R} + 2\mathbf{a})] + \cos[\mathbf{s}.(\mathbf{R}' - 2\mathbf{a})] \quad (14)$$

$$77602: \cos[\mathbf{s}.(\mathbf{R} + \mathbf{a} - \mathbf{b})] + \cos[\mathbf{s}.(\mathbf{R}' - \mathbf{a} - \mathbf{b})] \quad (15)$$

$$75602: \cos[\mathbf{s}.(\mathbf{R} + \mathbf{a} + \mathbf{b})] + \cos[\mathbf{s}.(\mathbf{R}' - \mathbf{a} + \mathbf{b})]. \quad (16)$$

From consideration of the terms in (13)–(16) for the *hhl* section

$$D_{75602}(\mathbf{s}) = D_{86602}(\mathbf{s}) + D_{66602}(\mathbf{s}) - D_{77602}(\mathbf{s}). \quad (17)$$

Similarly,

$$D_{64602}(\mathbf{s}) = D_{86602}(\mathbf{s}) + D_{55602}(\mathbf{s}) - D_{77602}(\mathbf{s}). \quad (18)$$

From (17) and (18)

$$\begin{aligned} &C_{66602}D_{66602}(\mathbf{s}) + C_{64602}D_{64602}(\mathbf{s}) + C_{75602}D_{75602}(\mathbf{s}) \\ &+ C_{77602}D_{77602}(\mathbf{s}) + C_{55602}D_{55602}(\mathbf{s}) + C_{86602}D_{86602}(\mathbf{s}) \\ &= (C_{66602} + C_{75602})D_{66602}(\mathbf{s}) \\ &\quad + (C_{55602} + C_{64602})D_{55602}(\mathbf{s}) \\ &\quad + (C_{86602} + C_{64602} + C_{75602})D_{86602}(\mathbf{s}) \\ &\quad + (C_{77602} - C_{64602} - C_{75602})D_{77602}(\mathbf{s}). \end{aligned} \quad (19)$$

The distribution coefficients for the four distributions of the right-hand side of (19) are therefore related to the six distinct correlation coefficients of (19) by the equations

$$x_{66602} = k(C_{66602} + C_{75602}) \quad (20)$$

$$x_{55602} = k(C_{55602} + C_{64602}) \quad (21)$$

$$x_{86602} = k(C_{86602} + C_{64602} + C_{75602}) \quad (22)$$

$$x_{77602} = k(C_{77602} - C_{64602} - C_{75602}). \quad (23)$$

Equations (20)–(23) are four of the 56 equations relating the distribution coefficients of the *hhl* section to the 74 unknown correlation coefficients. A further 26 equations were obtained from the 0*kl* section and 33 equations from each of the *h*0*l* and *h*1*l* sections. The number of equations corresponds to the number of independent distributions for each section.

Least-squares results

After initial least-squares analyses, the diffuse intensities calculated near the origin of reciprocal space, where no data could be measured, were negative. [Note that the intensity calculated at the origin of reciprocal space must necessarily be zero since $\Delta F_n(\mathbf{0})$

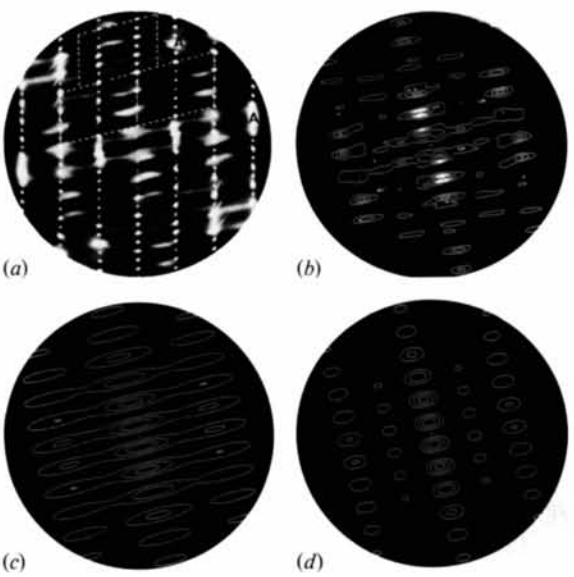

Fig. 4. Diffuse scattering for the *h*1*l* section. (*a*) Observed diffuse scattering. Features marked as in Fig. 2(*a*). (*b*) Calculated diffuse scattering after least-squares analysis. Contour level ~110. (*c*) Random distribution. (*d*) Correlation distribution for site 65503. Contours for (*c*) and (*d*) as in Figs. 2(*c*) and (*e*), respectively.

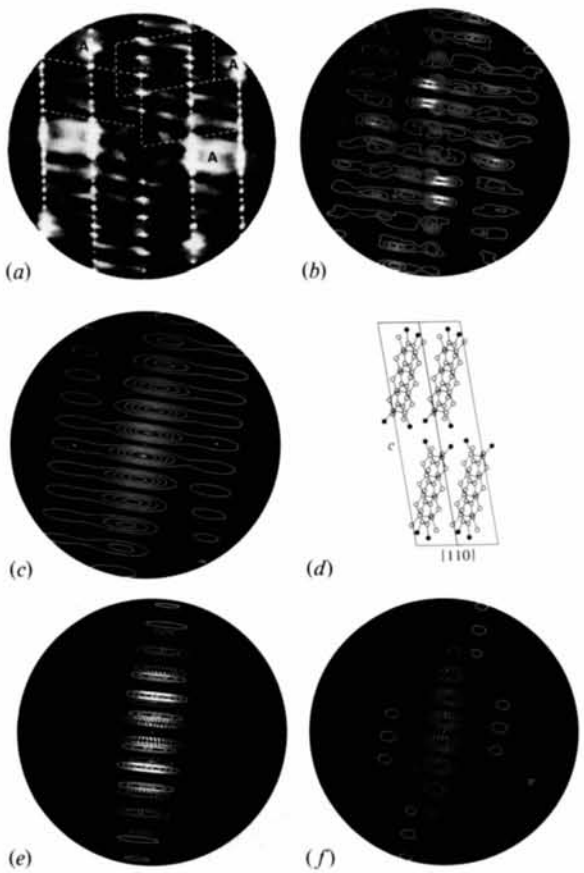

Fig. 5. Diffuse scattering for the *hhl* section. (*a*) Observed diffuse scattering. Features marked as in Fig. 2(*a*). (*b*) Calculated diffuse scattering after least-squares analysis. Contour level ~60. (*c*) Random distribution. (*d*) Projection of an ordered representation of the structure along [1$\bar{1}$0]. (*e*) Correlation distribution for site 65504. (*f*) Correlation distribution for site 65503. (*c*)–(*f*) as described in Fig. 2.

vanishes when type-*A* and type-*B* molecules are iso-electronic.] At selected points near the origin, the calculated diffuse scattering was therefore constrained to be non-negative.

During the unconstrained analyses, data which differed from the calculated diffuse intensities by more than three standard deviations (after background subtraction) were rejected. Rejection of data was justified since diffuse streaks, which were apparently not due to substitutional disorder (*e.g.* powder streaks), were evident on the undistorted photographs.

Details of the least-squares analyses for the four sections of reciprocal space are given in Table 2. As an example of the results obtained, the reduced distribution coefficients $\bar{x}'_l$ and their standard deviations obtained for the 0*kl* section are given in Table 3. For this section, the largest distribution coefficients were those for the random and 65503 distributions, as expected from qualitative consideration of Figs. 2(*a*), (*c*) and (*f*). Note that although the value of a particular reduced distribution coefficient may not differ significantly from zero, the related correlation coefficients may not necessarily be negligibly small because of possible cancellation between the correlation coefficients [see for example (20)–(23)].

Table 2. *Details of the least-squares analyses for* 2,3-*dichloro*-6,7-*dimethylanthracene*

	Reciprocal-space section			
	0*kl*	*h*0*l*	*h*1*l*	*hhl*
Number of data	2115	2014	2014	3171
Number rejected*	45	14	78	70
Range of sin $(\theta)/\lambda$ (Å^{-1})	≤0·16	≤0·26	≤0·27	≤0·29
Number of distribution coefficients	26	33	33	56
Active constraints at solution	7	2	0	2
Weighted *R* factor†	0·187	0·354	0·230	0·340

* Rejection criterion: $\Delta I > 3\sigma_I$.
† Defined by $R = \{\sum_l w_l |I^{(o)}_{diff}(\mathbf{s}_l) - I^{(c)}_{diff}(\mathbf{s}_l)|^2 / \sum_l w_l |I^{(o)}_{diff}(\mathbf{s}_l)|^2\}^{1/2}$.

Table 3. *Reduced distribution coefficients from the constrained least-squares analysis of the* 0*kl* *section for* 2,3-*dichloro*-6,7-*dimethylanthracene*

Standard deviations are given in parentheses and refer to the last significant figures.

Distribution	Coefficient	Distribution	Coefficient
Random	1·000	57502	−0·028 (12)
56501	0·456 (8)	35402	0·017 (10)
57501	0·054 (7)	36402	−0·042 (11)
75601	0·032 (6)	95702	−0·023 (10)
76601	0·061 (16)	96702	0·034 (12)
36401	−0·072 (16)	65503	1·363 (9)
66602	0·188 (12)	66503	0·109 (9)
67602	0·039 (12)	45403	0·012 (11)
65602	−0·063 (12)	85603	−0·031 (13)
74602	−0·009 (11)	55504	0·363 (12)
55502	0·168 (13)	56504	−0·037 (14)
56502	0·056 (12)	54504	0·039 (13)
54502	0·045 (10)	85604	−0·013 (12)

Contour plots of the calculated intensity distributions for the 0*kl*, *h*0*l*, *h*1*l* and *hhl* sections are shown in Figs. 2(*b*)–5(*b*), respectively. Neglecting thermal diffuse scattering, the agreement with the corresponding measured diffuse intensities of Figs. 2(*a*)–5(*a*) is considered satisfactory.

From the distribution coefficients obtained for the four sections, equations such as (11) involving reduced distribution coefficients were formed, as discussed previously. Since the equations obtained from the *h*0*l* and *h*1*l* sections were identical in form, estimates of corresponding reduced distribution coefficients were averaged to produce one set of equations. A final set of 112 equations was obtained involving the 74 correlation coefficients. This number of equations proved to be insufficient and three correlation coefficients which were expected to be small, C_{57501}, C_{74501} and C_{85602}, were set equal to zero.

With this further approximation, a solution of the least-squares equations resulting from minimizing ε of (12) was obtained with a weighted *R* factor of 0·103. The values obtained for the 71 correlation coefficients and their standard deviations are given in Table 4.

Thirty-five of the coefficients were significantly different from zero, with the first 25 entries of Table 4 significant at the 2σ level and the first 13 entries significant at the 3σ level. The largest correlations were between site 55501 and the nearest-neighbour sites 65503 and 64503 [correlation coefficient 0·46 (7)], and between site 55501 and the nearest-neighbour sites 65504 and 45404 [correlation coefficient 0·37 (4)]. The first 13 entries of Table 4 correspond to the correlation coefficients between site 55501 and 24 sites of the correlation neighbourhood. Equivalent sites in Table 4 are bracketed (two sites are equivalent if the two sets of internuclear vectors to 55501 are symmetry related). Elements of the least-squares correlation matrix with magnitudes greater than 0·6 are given in Table 5. The maximum value was 0·97 between C_{47501} and C_{43501}.

Discussion

Five approximations have been made during the course of the present work. These are:

(i) The diffuse scattering from displacement disorder (of either static or dynamic origin) has been neglected.

(ii) Random and correlation distributions were calculated from non-hydrogen disordered pseudoatoms only. The scattering from disordered hydrogen pseudoatoms has been neglected.

(iii) The diffuse intensities measured at the zeros of the random distribution were assumed to be scattering other than from substitutional disorder and were subtracted as background scattering.

Table 4. *Correlation coefficients for the neighbours of site* 55501 *for* 2,3-*dichloro*-6,7-*dimethylanthracene*

Standard deviations are given in parentheses and refer to the last significant figure. Equivalent sites are bracketed.

Site	Coefficient	Site	Coefficient	Site	Coefficient
65503 } 64503 }	0·46 (7)	45403 } 44403 }	0·04 (2)	55602	0·05 (7)
65504 } 45404 }	0·37 (4)	86602	0·13 (8)	57502	0·01 (9)
55504 } 55404 }	−0·24 (5)	55502	0·12 (9)	46502	0·04 (7)
76602	0·23 (5)	75602	0·07 (5)	45502	0·03 (9)
56501 } 54501 }	0·20 (4)	74504 } 34404 }	0·06 (4)	47502	−0·03 (9)
56502	0·20 (7)	47501 } 63501 }	−0·06 (4)	43502	−0·02 (6)
54504 } 54404 }	−0·17 (5)	46401 } 64601 }	−0·05 (3)	65502	0·06 (7)
56504 } 56404 }	−0·14 (4)	85603 } 84603 }	0·05 (3)	36502	0·00 (8)
64504 } 44404 }	0·13 (4)	54502	0·04 (3)	35502	−0·06 (6)
46501 } 64501 }	0·12 (4)	76504 } 36404 }	0·04 (3)	95702	−0·01 (2)
66504 } 46404 }	0·10 (3)	96702	0·03 (2)	36402	−0·02 (2)
35401 } 75601 }	0·07 (2)	45501 } 65501 }	−0·02 (5)	35402	0·02 (2)
45401 } 65601 }	−0·06 (2)	35501 } 75501 }	−0·02 (4)	66503 } 63503 }	0·04 (4)
66502	−0·18 (8)	34501 } 76501 }	−0·05 (5)	75503 } 74503 }	0·00 (4)
67602	−0·17 (7)	36401 } 74601 }	0·01 (3)	46503 } 43503 }	−0·02 (3)
66602	−0·16 (7)	44401 } 66601 }	−0·01 (3)	45503 } 44503 }	0·04 (4)
55503 } 54503 }	0·15 (7)	85601 } 25401 }	0·01 (2)	85503 } 84503 }	−0·01 (2)
56503 } 53503 }	0·10 (4)	86601 } 24401 }	0·01 (2)	35503 } 34503 }	−0·01 (3)
43501 } 67501 }	0·08 (4)	84601 } 26401 }	0·01 (2)	65603 } 64603 }	−0·02 (2)
76503 } 73503 }	−0·07 (3)	77602	0·04 (8)	35403 } 34403 }	0·02 (2)
75603 } 74603 }	−0·06 (3)	65602	−0·01 (7)	75504 } 35404 }	0·00 (5)
44501 } 66501 }	−0·06 (3)	64602	−0·02 (6)	45504 } 65404 }	0·03 (5)
55403 } 54403 }	−0·05 (2)	74602	0·02 (7)	85604 } 25304 }	0·01 (1)
34401 } 76601 }	0·04 (2)	56602	−0·04 (8)		

Table 5. *Correlations* (*magnitude* > 0·6) *between coefficients of Table* 4

C_{47501}, C_{43501}	−0·97	C_{57502}, C_{36502}	+0·74	C_{56602}, C_{86602}	−0·68
C_{57502}, C_{47502}	−0·93	C_{67602}, C_{64602}	−0·71	C_{67602}, C_{86602}	+0·67
C_{67602}, C_{77602}	−0·91	C_{66503}, C_{46503}	−0·70	C_{45501}, C_{44501}	+0·66
C_{64602}, C_{74602}	0·90	C_{56503}, C_{76503}	−0·70	C_{56502}, C_{36502}	+0·66
C_{56501}, C_{45501}	−0·83	C_{64602}, C_{77602}	+0·70	C_{85603}, C_{75603}	−0·66
C_{66502}, C_{36502}	−0·80	C_{77602}, C_{56602}	+0·70	C_{45501}, C_{75501}	−0·64
C_{46502}, C_{36502}	−0·79	C_{57502}, C_{46502}	−0·70	C_{55502}, C_{46502}	−0·64
C_{56502}, C_{46502}	−0·78	C_{46502}, C_{47502}	+0·70	C_{55502}, C_{45502}	−0·64
C_{56502}, C_{66502}	−0·76	C_{43501}, C_{76501}	−0·69	C_{64602}, C_{56602}	+0·63
C_{47502}, C_{36502}	−0·76	C_{67602}, C_{74602}	+0·69	C_{77602}, C_{86602}	−0·63
C_{66503}, C_{56503}	−0·74	C_{47501}, C_{76501}	+0·68	C_{56502}, C_{45502}	−0·63
C_{67602}, C_{56602}	−0·74	C_{77602}, C_{74602}	−0·68	C_{64504}, C_{54504}	−0·61
				C_{64602}, C_{55602}	−0·61
				C_{74602}, C_{55602}	−0·61
				C_{56602}, C_{55602}	−0·60

(iv) The variances of reduced distribution coefficients appearing on the right-hand side of equations such as (11) have been neglected.

(v) Covariances between reduced distribution coefficients have also been neglected when solving the equations for the correlation coefficients.

By careful selection of data, thermal diffuse scattering in the immediate vicinity of the Bragg peaks may be ignored without consequence. Elsewhere in reciprocal space, the diffuse scattering from displacement disorder (dynamic or static) is of greater concern, particularly when this scattering occurs in the same regions of reciprocal space as that from substitutional disorder. The analysis of the diffuse scattering has been confined to those intensities measured at low angles, where the contribution from displacement disorder is least. The neglect of thermal diffuse scattering at more general positions in reciprocal space may be critically examined by carrying out quantitative analyses of the diffuse scattering measured at several different temperatures. As mentioned in the *Introduction*, the degree of substitutional disorder is 'frozen in' at crystal growth and remains constant as the crystal temperature is varied. The sets of correlation coefficients obtained from analysis at different temperatures should therefore not differ significantly. Preliminary studies at ~300 and ~180 K of disordered crystals of one isomer of dibromodimethyldiethylbenzene showed that the pattern of diffuse scattering at general positions in reciprocal space was qualitatively the same at the two temperatures. At the lower temperature, the intensity of diffuse scattering at high angles was increased, presumably due to the smaller amplitudes of vibration. These studies are proceeding and will be reported elsewhere.

Approximation (ii) is justified for the analyses of 2,3-dichloro-6,7-dimethylanthracene since the dominant contributions to the calculated distributions are from Cl–Cl and C–Cl pairs of pseudoatoms. The scattering from Cl–H, C–H and H–H pairs may be neglected by comparison.

The approximation (iii) that the diffuse scattering measured at the zeros of the random distribution may be considered as background scattering was regarded as the most sensible way to proceed. The random distribution was calculated from structural parameters obtained from analysis of the Bragg intensities. The fringe pattern of the random distribution depends on the magnitude and direction of the internuclear vectors between disordered pseudoatoms at the same molecular site. If the positions of the disordered pseudoatoms were incorrect, or the disordered hydrogen pseudoatoms could not be neglected [see approximation (ii)], then the zeros of the random distribution would not be in 1:1 correspondence with the minima of the measured intensities. The intensity measured at the zeros of the random intensity was therefore attributed to scattering

from air or from the crystal mount, or to Compton scattering, or to fluorescence from the halogen substituents. During preliminary least-squares analyses, this additional scattering was well described by a term which was constant over the reciprocal-space section. However, treating this scattering as local background correction was considered preferable.

The equations relating the distribution coefficients to the correlation coefficients were solved approximately by first using the equation for the random distribution coefficient to eliminate the scale factor for each reciprocal-space section. This procedure is equivalent to introducing a set of constraints such that the calculated random distribution coefficients are set equal to their estimated values. In this way the equations relating the reduced distribution coefficients to the correlation coefficients are linear. Approximations (iv) and (v) then enabled these equations to be solved approximately by linear least squares. None of the constraints or (iv) and (v) need be introduced if the scale factors and correlation coefficients are considered simultaneously as non-linear parameters. The procedures required to solve the non-linear equations are more complicated and convergence to the required solution is expected to be slow. The procedures used in the present work are justified by their simplicity. More complex techniques are not expected to alter significantly the correlation coefficients reported in Table 4. Combining the data obtained for each reciprocal-space section may be more preferable than treating each section separately. Non-linear techniques as well as approximations (iv) and (v) could then be avoided. The correlation coefficients could be determined directly from the ratios (x_l/x_r).

In studies of 9-bromo-10-methylanthracene reported previously (Epstein, Welberry & Jones, 1982; Welberry, Jones & Epstein, 1982), nearest neighbours were assigned primary correlation coefficients and the coefficients for other sites in the neighbourhood were assumed to be products of these. One feature of the present work is that no assumptions are made about the relationship of non-nearest-neighbour coefficients to the primary coefficients. The coefficients of Table 4 may be compared with the corresponding product-rule estimates, assuming the most direct correlation pathway. Product-rule estimates for the first fifteen non-nearest-neighbour correlations of Table 4 are given in Table 6. For next-nearest neighbours, the product-rule estimates are generally in good agreement with the values of Table 4. Estimates for the correlations between more distant neighbours are generally smaller in magnitude than in Table 4.

Table 6. *Product-rule estimates of non-nearest-neighbour site correlation coefficients of Table* 4

Site	Product	Estimate	From Table 4
56501	$C_{65503} \times C_{65503}$	0·21 (5)	0·20 (4)
54504	$C_{55404} \times C_{65503} \times C_{65503}$	−0·05 (1)	−0·17 (5)
56504	$C_{55404} \times C_{65503} \times C_{65503}$	−0·05 (1)	−0·14 (4)
64504	$C_{65503} \times C_{65503} \times C_{65504}$	0·08 (2)	0·13 (4)
46501	$C_{55503} \times C_{65503}$	0·07 (3)	0·12 (4)
66504	$C_{65503} \times C_{65503} \times C_{65504}$	0·08 (2)	0·10 (3)
75601	$C_{65504} \times C_{65504}$	0·14 (2)	0·07 (2)
65601	$C_{65504} \times C_{55504}$	−0·09 (2)	−0·06 (2)
66502	$C_{65503} \times C_{55504}$	−0·11 (3)	−0·18 (8)
67602	$C_{66602} \times C_{65503} \times C_{65503}$	−0·03 (1)	−0·17 (7)
56503	$C_{55503} \times C_{65503} \times C_{65503}$	0·03 (2)	0·10 (4)
67501	$C_{55503} \times C_{65503}$	0·01 (1)	0·08 (4)
76503	$C_{55503} \times C_{65503} \times C_{65503}$	0·03 (2)	−0·07 (3)
75603	$C_{65504} \times C_{66602}$	−0·06 (3)	−0·06 (3)
44501	$C_{65503} \times C_{55503}$	0·07 (3)	−0·06 (3)

Summary and concluding remarks

A procedure has been presented for the quantitative determination of correlation coefficients between sites of a substitutionally disordered molecular crystal from the analysis of the diffuse intensities of X-rays scattered by the crystal. Neglecting thermal diffuse scattering, data measured from Weissenberg photographs have been analysed by linear least squares using calculated random and correlation distributions. No assumptions were made about the correlations between non-nearest-neighbour sites. For disordered crystals of 2,3-dichloro-6,7-dimethylanthracene grown by sublimation, correlations for sixty neighbours of each molecular site were found to be significantly different from zero.

The procedures reported here represent a means by which the analysis of diffuse X-ray scattering from disordered crystals may be carried out in a routine fashion. Some of the assumptions made are somewhat crude, but enable the procedures to be kept simple. Further work is required to appraise critically the accuracy of the results obtained.

We are indebted to Dr M. R. Osborne, Research School of Social Sciences (Statistics), Australian National University, who developed for us an efficient algorithm for the solution of the quadratic programming problem. We are also grateful to him for many helpful discussions on numerical procedures during the course of this work.

References

EPSTEIN, J., WELBERRY, T. R. & JONES, R. D. G. (1982). *Acta Cryst.* A**38**, 611–618.

FLACK, H. D. (1970). *Philos. Trans. R. Soc. London Ser. A*, **266**, 575–591.

GILL, P. E. & MURRAY, W. (1974). *Numerical Methods for Constrained Optimization*. London: Academic Press.

GLAZER, A. M. (1970). *Philos. Trans. R. Soc. London Ser. A*, **266**, 635–639.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press.

JONES, R. D. G. & WELBERRY, T. R. (1980). *Acta Cryst.* B**36**, 852–857.

WELBERRY, T. R. (1983). *J. Appl. Cryst.* **16**, 192–197.

WELBERRY, T. R. & JONES, R. D. G. (1980). *J. Appl. Cryst.* **13**, 244–251.

WELBERRY, T. R., JONES, R. D. G. & EPSTEIN, J. (1982). *Acta Cryst.* B**38**, 1518–1525.

WELBERRY, T. R., JONES, R. D. G. & PUZA, M. (1983). *Acta Cryst.* C**39**, 1123–1127.

Acta Cryst. (1983). A**39**, 892–896

Statistical Geometry. II. Numerical Solution *via* the Single Pixel Equation

BY STEPHEN W. WILKINS

CSIRO, Division of Chemical Physics, PO Box 160, *Clayton, Victoria, Australia* 3168

(*Received* 13 *September* 1982; *accepted* 1 *July* 1983)

Abstract

A simple single pixel equation (SPE) is presented which, when solved self-consistently for each pixel, can yield *exact* solutions to the statistical inversion problem for diffraction data outlined in paper I of this series [Wilkins, Varghese & Lehmann (1983). *Acta Cryst.* A**39**, 47–60]. The SPE approach was used to obtain the results presented in I and is shown here to have both practical and heuristic advantages in that it: (i) provides a very transparent approach to the task of solving the fundamental equations of the statistical geometric problem, (ii) can greatly improve the rate of convergence and (iii) readily allows the convexity of the constraint contributions to be monitored and, if desired, controlled. For the important case of 'phase refinement' *via* constraint (1) of I and the assumption of: (i) a complete data set of E_k up to the resolution limit of the data and (ii) uniform errors (*i.e.* $\sigma_{k,1} = \sigma$), it is shown that the *maximum-entropy structure* (*MES*) *can be fully refined via the SPE in only one Fourier transform cycle*, and so should be extremely efficient for biological macromolecules.

1. Introduction

In the first paper in this series (Wilkins, Varghese & Lehmann, 1983, hereafter termed I), we laid the foundations for an information-theory-based approach, termed statistical geometry (SG), to the crystallographic inversion problem and presented (see also Gull & Daniell, 1978) a set of N coupled non-linear equations (eqs. I.14*a*) for the discrete distribution, $\mathbf{p}$, of scattering density in the unit cell. If the SG method of structure determination and refinement is to be made a practical tool applicable to biological macromolecules, then highly efficient methods of solving these equations must be developed (*cf.*, *e.g.*, Gull & Daniell, 1978; Collins, 1982). This paper and the following one in the series (Wilkins, 1983) are both directed to that end.

In the present paper we develop and discuss a new approach to the exact numerical solution of these non-linear equations which proceeds *via* a single pixel approximation (SPA) and in many ways resembles the mean-field-type of approximations often encountered in quantum mechanics and statistical mechanics. Because of its simplicity, such an approach offers both heuristic and practical advantages (*e.g.*, improved convergence properties) although it need not necessarily lead to a solution. For a very important special case, namely that of 'phase refinement' using constraint (1) of I alone and some other typically reasonable assumptions (see § 4) it is shown that the structure can be *fully refined* (within the SG framework) *via the SPE in only one Fourier transform cycle*, and so offers an extremely efficient approach to structure refinement even for biological macromolecules.

2. The single pixel approximation (SPA)

Following the notation and definitions introduced in I and starting from equations (I.14) we may Taylor expand the exponent there to first order in p_j about an *arbitrary* trial structure, $\mathbf{p}^t = \mathbf{p}^{(0)}$, and write

$$p_j = \exp\left\{-\lambda_0 - \boldsymbol{\lambda}.\left[\mathbf{f}_j^{1,(0)} + \sum_{j'}(p_{j'} - p_{j'}^{(0)})\mathbf{f}_{j'j}^{2,(0)}\right]\right\} \quad \text{for } j = 1, \ldots, N, \tag{1}$$

where λ_0 is the Lagrange multiplier associated with the normalization constraint (termed the structural freedom in I) and $\boldsymbol{\lambda}$ the Lagrange multiplier vector

0108-7673/83/060892-05$01.50