Acta Cryst. (1998). A**54**, 438–446

The *Ab Initio* Solution of Structures from Powder Diffraction Data: the Use of Maximum Entropy and Likelihood to Determine the Relative Amplitudes of Overlapped Reflections Using the Pseudophase Concept

WEI DONG AND CHRISTOPHER J. GILMORE*

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland. E-mail: chris@chem.gla.ac.uk

(*Received 6 August 1997; accepted 6 January 1998*)

Abstract

The maximum-entropy (ME) method coupled with likelihood evaluation is used to solve crystal structures *ab initio* from powder diffraction data. The conventional ME approach is extended to include the estimation of individual reflection intensities that are overlapped as well as their phases. A method based on Duncan's extension of t and F tests is used in which protection against overestimating the associated significance levels is provided for situations where multiple t tests are performed and where misleading estimates of significance would normally result. When the likelihood estimates prove sensitive to the partitioning of intensities, considerable accuracy in deconvoluting them is possible and the final electron-density maps are of high quality even from data collected using laboratory sources.

1. Introduction

Solving crystal structures *ab initio* from powder diffraction data is difficult because of the familiar problem in which the three-dimensional diffraction pattern is collapsed into one dimension, thus giving rise to peak overlap in the data set. The overlap problem serves to limit the applicability of powder diffraction as a routine tool for determining and refining crystal structures from powders. The synchrotron has made a profound difference in many cases but the intrinsic overlap arising from line broadening and the existence of exact peak overlap in high-symmetry space groups still impose very severe limitations on what problems can and cannot be addressed. Recent methodologies that have been proposed to approach this problem include a Bayesian formalism to deconvolute the overlapped data (Sivia & David, 1994); the fast iterative Patterson squaring (FIPS) method of Estermann *et al.* (Estermann *et al.*, 1992; Estermann & Gramlich, 1993); the use of annealing methods coupled with model fitting by Andreev *et al.* (1996); an extension to this idea using genetic algorithms as an optimization method (Shankland *et al.*, 1997); model building using Monte Carlo methods coupled with fitting the full diffraction profile by Tremayne *et al.* (1997) and, finally, the use of maximum entropy and anomalous scattering by Burger *et al.* (1997). There is a good review of the current status of powder diffraction as a tool for solving and refining structures by Harris & Tremayne (1996), a general review of maximum-entropy methods of solving crystal structures by Gilmore (1996) and surveys of the ME method in this context by Shankland (1994) and Gilmore (1993), all of which will prove useful in the context of this paper.

Despite these undoubtedly important advances, the overlap problem is not, in general, solved and we present here a method based on entropy maximization coupled with likelihood evaluation to determine not only the phases of reflections under a given overlap but also their relative intensities. It is an extension of the original theory of Bricogne (1991) as applied to powders (Gilmore *et al.*, 1991; Shankland *et al.*, 1993), which has proved successful with both organic and inorganic materials. (See, for example, Tremayne, Lightfoot, Glidewell *et al.*, 1992; Tremayne, Lightfoot, Harris *et al.*, 1992.) The method has some superficial similarities with procedures used in *SIRPOW*92 where amplitude partition is also employed (Altomare *et al.*, 1994) but we use likelihood estimates as a figure of merit and analyse the results in a different way.

2. The overlap problem defined in terms of hyperphases and pseudophases

Let a given overlap set contain m_a acentric and m_c centric reflections. The net intensity, I, is

$$I = \sum_{i=1}^{m_a} p_i(A_i^2 + B_i^2) + \sum_{j=m_a+1}^{m} p_j C_j^2, \qquad (1)$$

where p_j is the multiplicity of reflection j, A and B are the real and imaginary parts of an acentric structure factor and C is the structure-factor component for centric reflections. Multiplicities are readily derived from point-group symmetry so that the problem we face with an overlap is the determination of the A, B and C coefficients in equation (1). It is profitable to rephrase this problem in terms of hyperphases. Following Bricogne (1991), we define the overlap as a vector **F** in

an n-dimensional space $\mathbb{R}^n$ as

$$\mathbf{F} = \begin{bmatrix} p_1^{1/2}A_1 \\ p_1^{1/2}B_1 \\ p_2^{1/2}A_2 \\ p_2^{1/2}B_2 \\ \vdots \\ p_{m_a}^{1/2}A_{m_a} \\ p_{m_a}^{1/2}B_{m_a} \\ p_{m_a+1}^{1/2}C_{m_a+1} \\ \vdots \\ p_m^{1/2}C_m \end{bmatrix}, \tag{2}$$

where $n = 2m_a + m_c$ is the number of degrees of freedom, and $m = m_a + m_c$ is the number of moduli under the overlap. $\mathbf{F}$ can be visualized as a hypersphere in $\mathbb{R}^n$ and we wish to parameterize it to resolve the overlap intensities. Bricogne (1991) provides the tools for doing this using an extension of earlier ideas from Stubbs & Diamond (1975) and Holmes *et al.* (1975). A total of $m-1$ splitting angles, $\psi_{j,j=1,m-1}$, $0 \le \psi \le \pi/2$, are defined called *pseudophases*, which define a point on the positive unit hypersphere. As an example, consider an overlap of net intensity $I^{1/2}$ comprising two acentric and one centric reflections. The radius of the hypersphere is $R = ||\mathbf{F}|| = I^{1/2}$. Define R_1, R_2 and R_3 as the individual intensities of these reflections, then

$$\begin{aligned} R' &= R\sin\psi, \quad 0 \le \psi_1 \le \pi/2 \\ R_1 &= R'\cos\psi_2, \quad 0 \le \psi_2 \le \pi/2 \\ R_2 &= R'\sin\psi_2 \\ R_3 &= R\cos\psi_1. \end{aligned} \tag{3}$$

The angles ψ_1 and ψ_2 are the two pseudophases; once they are known, the individual intensities of the reflections are readily determined and now it is only necessary to derive the conventional acentric and centric phases to fully characterize the overlap. This is the process of determining *hyperphases*. The phasing process is now defined exclusively in terms of angles.

Alternatively, a binary tree can be constructed to express this procedure (Bricogne, 1991):

$$\begin{aligned} n(\nu l) + n(\nu r) &= n(\nu) \\ R(\nu l) &= R(\nu)\cos\psi(\nu) \\ R(\nu r) &= R(\nu)\sin\psi(\nu). \end{aligned} \tag{4}$$

Here, $R(\nu)$ is the radius of the hypersphere at level ν, $R(\nu l)$ and $R(\nu r)$ the radius of the sphere for left and right side branches and n the dimension of the sphere. Among the $n-1$ hyperphases ψ, $m-1$ are the splitting angles or pseudophases and the radius of $\mathbf{F}$ is partitioned among m reflections by them. The remaining m_a hyperphases are the conventional phases for the m_a acentric reflections.

3. Duncan's procedure for multiple significance tests

In order to determine the necessary pseudophases, we are going to carry out a sequence of multiple t tests to decide on which ratios of intensities lying under an overlap are the most likely in a statistical sense. However, a blind application of such methods leads to a serious overestimate of the associated level of significance and hence introduces damaging systematic errors, so we need a procedure to protect us from this.

Consider an overlap in which there are three reflections of intensity I_1, I_2 and I_3, respectively, which we wish to determine. To do this, various partitions of the net intensity are tried, and each one has associated with it some figure of merit (in this case the log-likelihood gain, LLG) and a measure of its variance. We can carry out t tests on various combinations of LLGs and their associated standard deviations. There will be (Duncan, 1955):

(i) 3! = 6 decisions of the form LLG_1 is significantly less than LLG_2, LLG_2 is significantly less than LLG_3 *etc.* In the usual shorthand, we write these six choices as (1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2) and (3,2,1).

(ii) $2\times3!/2$ = 6 decisions of the form LLG_1 is significantly less than LLG_2 and LLG_3 but LLG_2 and LLG_3 are not significantly different from each other. This is written (1,2,3), and the other possible choices are (2,1,3), and (3,1,2) with an obvious extension to rankings of the type (1,3,2), (1,2,3) and (2,3,1). The latter, for example, signifying that LLG_2 and LLG_3 do not differ significantly from each other but are significantly less than LLG_1.

(3) $1\times3!$ = 6 decisions of the form LLG_1 is significantly less than LLG_3 but LLG_1 and LLG_2 do not differ significantly and LLG_2 and LLG_3 do not differ significantly either. This is written as (1,2,3). The remaining decisions are written (1,3,2), (2,1,3), (2,3,1), (3,1,2) and (3,2,1).

(iv) $\overline{1}\times3!/3!$ = 1 decisions of the form (1,2,3), *i.e.* there are no significant differences among the intensities.

This is a total of 19 possible tests and, whereas it is computationally trivial, a situation arises that if several t tests are performed at, say, a 5% significance level, the probability that one of these gives an erroneous indication is greater than 0.05. (See, for example, Cochran & Cox, 1957, pp. 75–76; Duncan, 1955). If the t tests are independent, this probability of error is surprisingly large: 0.23 for 5 tests, 0.40 for 10 tests and 0.64 for 20 tests. It is clearly important to protect against this source of error and Duncan (1955) has provided such a procedure in which the t-test tables in the form of studentized ranges (Pearson & Hartley, 1966, Table 29)

are modified to include variable protection levels that are adjusted against incorrect indications of significance. This is the procedure used here.

The LLGs are ranked in order and an initial analysis of variance carried out. A standard error of the mean LLG, $S_{\rm error}$, is computed along with the number of degrees of freedom, n. Duncan's tables are then used: the number of degrees of freedom and the significance level define the necessary entries. A studentized range statistic is computed:

$$q_r = (T_i - T_j)/(nS_{\rm error})^{1/2}, \qquad (5)$$

where T_i and T_j are two means to be compared. A critical value for each degree of freedom and the range r can be found from the tables prepared by Duncan. The hypothesis of equality of the two means will be rejected if q_r is greater than the specified level. These are the *significant studentized ranges*. Each of these is multiplied by the standard error to form what Duncan calls the *shortest significant ranges*. We now test each difference in the order: largest minus smallest, largest minus second smallest, ..., second minus smallest, second minus second smallest, ..., finishing with second smallest minus smallest. Each difference is significant only if it exceeds the corresponding shortest significance range. There is an exception to this rule: no difference is significant if the two LLGs concerned are both contained in a subset (which can be the complete set) of the mean LLGs which has a non-significant range.

We will now use this procedure in an ME–likelihood environment.

4. The determination of pseudophases using the maximum-entropy–likelihood method and Duncan's procedure

The method we have evolved is an extension of the ideas we have used to study problems in single-crystal and powder phasing (Bricogne & Gilmore, 1990; Gilmore *et al.*, 1990, 1991, 1993) and works as follows:

(i) The data are normalized using the *MITHRIL* program (Gilmore, 1984, 1988) to give unitary structure factors $|U_{\mathbf{h}}|^{\rm obs}$. The overlapped reflections are included.

(ii) An origin is defined in the conventional manner using non-overlapped reflections. If this is not possible, origin definition can be left incomplete or overlaps can be used but this procedure was not necessary for the structures described here. The enantiomorph is also defined if both necessary and feasible at this stage. These phased basis-set reflections comprise the root node of a maximum-entropy–likelihood phasing tree (which should not be confused with the tree described in §2 for parameterizing overlaps) and subjected to constrained entropy maximization in which both the amplitudes and phases act as the constraints (Bricogne, 1984; Bricogne & Gilmore, 1990; Gilmore *et al.*, 1990).

(iii) A set of reflections, both overlapped and non-overlapped, is selected *via* the algorithm of optimum second-neighbourhood extension (Gilmore *et al.*, 1990; Bricogne, 1993, 1997*a*). This works in the same way as single-crystal data sets except that, for overlaps, all the reflections in a given set must be considered simultaneously and if one reflection is selected from an overlap set for the basis set then all the reflections under the same overlap are included. Pseudophases arising from overlapped intensities are constrained to lie between 0 and $\pi/2$ in steps of $\pi/10$ (18°), *i.e.* working as 18, 36, 54 and 72°. This seems to give the finest grid to which this formalism is sensitive. 0 and 90° were not used since these correspond to a situation where one intensity is zero and this introduces computational difficulties as well as being unlikely when the overlap is large. There are possibly better ways of doing this using spherical error-correcting codes as a source of experimental design (Bricogne, 1991, 1993, 1997*b*; MacWilliams & Sloane, 1977), but we intend to test these at a later time. Conventional phases are permuted in the usual way. Each choice of conventional and pseudophase defines a node on the second level of the phasing tree and each node is subjected to constrained entropy maximization. At convergence, each node has associated with it a log-likelihood gain (LLG) computed using the generalized form of likelihood when powder overlaps are present. It is these LLGs that are used in tests of significance. This formalism is fully described by Bricogne (1991) and Gilmore *et al.* (1991) but it may be useful to summarize the relevant mathematics here.

Let a given overlap comprise $R_{i,\,i=1,\,m}$ observed intensities, and $r_{i,\,i=1,\,m}$ the calculated intensities produced as a result of extrapolation from a maximum-entropy optimization. (A single non-overlapped reflection is treated in a similar way by putting $m = 1$.) For each overlap or single reflection, define

$$R = \sum_{i=1}^{m} p_i R_i^2 \qquad (6)$$

and

$$r = \sum_{i=1}^{m} p_i r_i^2. \qquad (7)$$

Let

$$z = Rr/\Sigma, \qquad (8)$$

where Σ is a refinable parameter related to the unit-cell contents *via* $\Sigma \simeq 1/N$, where there are N non-H atoms, assumed equal, in the unit cell. Define

$$X_n(z) = \exp(-z)\,{}_0F_1(-; n/2; z^2/4), \qquad (9)$$

where ${}_0F_1$ is a confluent hypergeometric function. The

log likelihood, LH, is written

$$\mathrm{LH} = \sum_{\text{all extrapolates}} [-n/2 \log \Sigma - (R-r)^2/2\Sigma + \log X_n(z)]. \quad (10)$$

For the null hypothesis, LH_0, we set $r = 0$ and hence $z = 0$ so that

$$\mathrm{LH}_0 = \sum_{\text{all extrapolates}} [-n/2 \log \Sigma - R^2/2\Sigma]. \quad (11)$$

The log-likelihood gain (LLG) is then

$$\mathrm{LLG} = \mathrm{LH} - \mathrm{LH}_0. \quad (12)$$

Clearly, the higher the value of LLG the more closely the process of entropy maximization has predicted the pattern of observed intensities and the more likely it is that our phase choices are correct.

(iv) The likelihood estimates are analysed. When pseudophase permutation is involved, the LLG analysis is divided into two stages: pseudophases and conventional phases. Since a four-point sampling method is used for the splitting angles instead of the coarse binary sampling for conventional phases, an F test is first invoked to test if pseudophase permutations have a statistical effect on the LLG at a given significance level. When significant differences do exist, the LLG means for all permuted pseudophase values are compared using Duncan's procedure at either the 1 or the 5% level.

(v) If no significant differences are found, then the LLGs are not sensitive to the relevant pseudophases. If this is the case, then the overlaps can be deconvoluted by multiplicity-weighted equipartitioning (pseudophase = $\pi/4$). This is surprisingly frequent and may possibly arise from our use of likelihood in its simplest diagonal approximation (Bricogne & Gilmore, 1990). Other, more sophisticated, likelihood formalisms (Bricogne, 1993, 1997*a*) may prove more sensitive but these have not yet been studied by us.

(vi) If the significance test indicates suitable pseudophases, then the conventional phase angles are extracted by the Student t test (Shankland *et al.*, 1993; Gilmore *et al.*, 1997).

(vii) Those nodes that pass (vi) and have the highest LLGs are used to compute Sim-weighted centroid maps (Bricogne & Gilmore, 1990; Bricogne, 1991; Gilmore *et al.*, 1991) in which the overlapped reflections are included. The map coefficients use the observed U magnitudes, phase angles resulting from the ME extrapolation process and Sim-type weights w_i (Sim, 1959) computed *via*

$$w_i = \frac{R^2{}_0F_1(-; n/2+1; z^2/4)}{n\Sigma_0F_1(-; n/2; z^2/4)}. \quad (13)$$

For small molecules, this is usually sufficient to reveal at least part of the structure, which can then be completed by conventional Fourier and refinement techniques. If it is not, then another cycle of tree building and entropy maximization can be carried out from step (iii).

5. Applying pseudophase permutation to the structure determination of TTPD, $C_2S_4O_2$

A powder diffraction data set from TTPD (1,3,4,6-tetrathiapentalene-2,5-dione) was selected to test the pseudophase permutation approach. It crystallizes in space group $P2_1/a$ with a = 8.3049, b = 10.8945, c = 3.9366 Å, β = 104°, with two C, two S and one O atoms in each asymmetric unit. Since it is severely overlapped (110 out of 176 reflections) and most of the reflections with $l = \pm 1$ are in the overlap sets, in order to get access to these reflections, pseudophases have to be used (Lightfoot *et al.*, 1993). In fact, this structure has already been solved using the ME method by permuting the phases of equipartitioned overlap reflections. The intensity data come from a laboratory source.

A three-level phasing tree was generated. All the reflections that were permuted at each level of the phasing tree and their calculated and true phases are listed in Table 1. For the first level, the origin was defined with three linearly independent reflections in the conventional way. In the second level, four non-overlapped reflections and a single overlap set containing two reflections (211 and 31$\bar{1}$) were permuted generating a total of $2^6 \times 4 = 256$ nodes.

After entropy maximization, all the 256 nodes were analysed for magnitude partition (Table 2). The null hypothesis of magnitude partition having no significant effect on LLG was rejected by an F test. At the 1% significant level, application of Duncan's procedure indicated that all the LLG means for the four pseudophase values were significantly different and selected a value of 18° for the pseudophase. The calculated pseudophase for this overlap set is 25°. The 64 nodes using this pseudophase for the overlap set were passed for conventional phase analysis using the Student t test. Node 38 with the highest LLG exhibited a good corresponding centroid map. The phases calculated from the structure are also listed in Table 1; it can be seen that all the phases of the six permuted reflections were correct for node 38.

A third phasing level was then constructed using node 38 as the parent. Five non-overlapped and one overlapped set of reflections were permuted and 128 nodes created for each of the four possible pseudophase values. From the pseudophase analysis of all the 512 nodes thus generated (Table 3), the magnitude partition showed a significant effect on the LLG at the 1% confidence level. Duncan's analysis showed that all the four permuted pseudophase values were different from each other; they were sorted by LLG means and that of

Table 1. *Permuted reflections and their calculated phases for TTPD*

Node 38 is the likelihood preferred node for level 2 of the phasing tree; node 76 is the preferred node for level 3 and node 171 has been chosen as an example of what happens to phase predictions when an incorrect pseudophase is selected. All phase angles are in degrees.

	h	*k*	*l*	Pseudophase choices	True pseudo-phase	Predicted pseudophase	True phase	Predicted phases: node 38	Predicted phases: node 76	Predicted phases: node 171
Origin										
	2	0	−1							
	1	4	−2							
	2	7	0							
2nd level										
	6	0	−1				180	180	180	180
	6	0	0				180	180	180	180
	3	2	0				360	360	360	360
	1	4	0				180	180	180	180
	Overlap			18/36/54/72	25	18				
	2	1	1				180	180	180	180
	3	1	−1				360	360	360	360
3rd level										
	6	2	−1				180		180	360
	5	1	0				360		360	180
	5	3	1				180		180	180
	4	1	0				360		360	360
	0	0	3				180		180	180
	Overlap			18/36/54/72	80	72				
	0	1	1				180		180	360
	1	1	−1				180		180	360

Table 2. *Pseudophase analysis for TTPD at level 2 of the phasing tree*

All tests are carried out at the 1% significance level. For a definition of q_r, see equation (5).

F test	Sum of squares for pseudophase choices	2.14	Degrees of freedom	3	
	Sum of squares for error	0.98	Degrees of freedom	189	
	F ratio	138.00			
	Probability that all hyperphases are same	0.0			
Means compared	Pseudophase	72	54	36	18
	LLG means	0.2861	0.3781	0.4716	0.5259
	q_r				
	72	–	10.234	20.624	26.670
	54	–	–	10.390	16.436
	36	–	–	–	6.046
	18	–	–	–	–
	r		2	3	4
	Critical value at 1% significance level		3.70	3.85	3.97
	No. of degrees of freedom: 189				

72° had the highest mean LLG. This was consistent with the calculated value of 80°. The LLGs of the 128 nodes having a pseudophase of 72° were analysed for conventional phases. Node 76 was the preferred node and the corresponding centroid map correctly located all the non-H atoms. Comparison with the calculated phases showed that all the permuted reflections in this level had been correctly phased (see Table 3). Fig. 1 shows the LLG preferred centroid map.

For comparison, another 128 nodes with a pseudophase of 36°, corresponding to the lowest mean LLG, were also analysed and four different nodes selected; none of them had an interpretable centroid map. Two out of five were non-overlapped and all the overlap reflections were wrongly phased.

6. Other tests of the method

6.1. *SAPO-40*

SAPO-40 [$(Si, Al, P)_{64}O_{128} \cdot 4TPAOH$] is a large-pore molecular sieve. Its structure has been reported by Dumont *et al.* (1993) and McCusker & Baerlocher (1995). The data come from a synchrotron source and are 82% overlapped. We used the space group *Pmmn* with cell parameters a = 21.9410, b = 13.6912, c =

Table 3. *Pseudophase analysis for TTPD at level 3 of the phasing tree*

All tests are carried out at the 1% significance level.

F test	Sum of square for pseudophases	1018.68	Degrees of freedom		3
	Sum of squares for error	1080.37	Degrees of freedom		381
	F ratio	119.75			
	Probability that all hyperphases are same	0.0			
Means compared	Pseudophase	36	18	54	72
	LLG means	−0.1118	1.1712	2.8595	3.4717
	q_r				
	36	–	8.620	19.963	24.076
	18	–	–	11.343	15.456
	54	–	–	–	4.113
	72	–	–	–	–
	r		2	3	4
	Critical value at 1% significance level		3.69	3.84	3.95
	No. of degrees of freedom: 381				

7.1244 Å. As McCusker & Baerlocher (1995) have shown, this is acceptable until atom-type designation is imposed for P and Al; this causes a doubling of the *c* axis and the space group becomes *Pccn*. However, for a starting point for Rietveld refinement, the smaller cell is sufficient.

In general, we found that, unlike TTPD, the LLGs were not sensitive to overlap partition but some phasing sequences were successful in this regard and the following is typical:

(i) An origin was defined in the usual way generating the root node.

(ii) Eight non-overlapped reflections were given permuted phases generating a 256 node second level. (The space group is centrosymmetric.) The correct node, *i.e.* the one with zero phase error, was selected at this point. This node was in the top eight as ranked by LLG analysis and we could have generated a more complex tree here using all these eight nodes as parents for a third level, which indeed one would do if working on an unsolved structure. However, since we are interested in establishing the feasibility of magnitude partition here, and to contain an otherwise very long calculation and for purposes of clarity, only the correct node was used as the parent for the next level.

(iii) 14 reflections were given permuted phases and the strongest overlap involving the 240 and 601 reflections was given true and pseudophase permutation. Since this would generate $2^{16} \times 4 = 65\,536$ nodes in a full factorial design, a Nordstrom–Robinson code was used as a source of phase permutation (Bricogne, 1993, 1997*a*; Gilmore & Bricogne, 1997). This procedure will be discussed in detail elsewhere (Gilmore *et al.*, 1998). The use of the Nordstrom–Robinson code plus pseudophase permutation generated 1024 nodes.

(iv) The Duncan procedure operating at a 5% confidence level indicated a pseudophase angle of 72°; a value of 54° was not significantly different although it had a lower average LLG. The mean of these indications is 63°. The true value established from the refined structure is 61°. This is an excellent agreement. Analysis at the 1% level instead of 5% indicated no significant differences. The best map shown in projection down the *z* axis with the atomic coordinates indicated as crosses is shown in Fig. 2. Its quality is high with most of the density in correct places and the pore clearly visible.

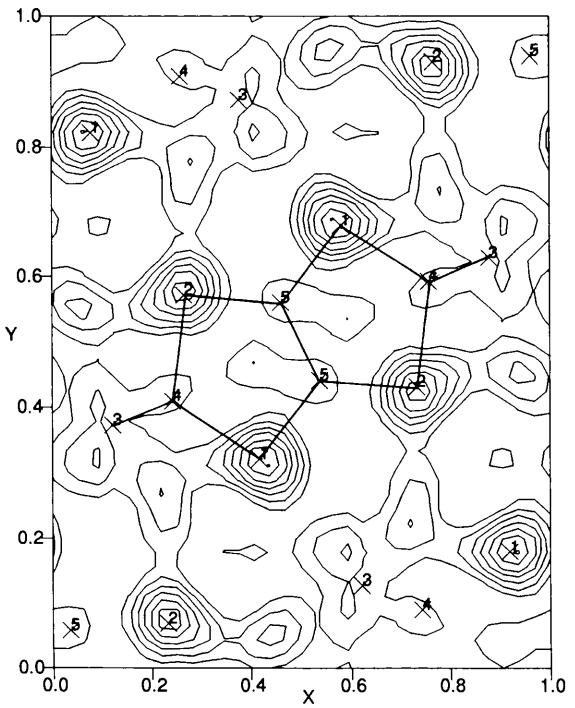

Fig. 1. The centroid map in projection down the *c* axis for node 76 for TTPD. Atoms labelled 1 and 2 are S, 3 is O and the remainder are C atoms. All atoms are labelled with crosses; coordinates come from the refined crystal structure.

Smaller phasing trees than this were unsuccessful in predicting correct pseudophases.

6.2. *Formylurea, $C_2H_4N_2O_2$*

Formylurea is the first equal-atom organic structure to be solved from powder diffraction data (Lightfoot *et al.*, 1992). It crystallizes in the noncentrosymmetric space group $Pn2_1a$ with cell parameters a = 16.817, b = 6.062, c = 3.669 Å and Z = 4. It is more difficult in some respects than the previous examples even though the structure is very small. A three-level tree was constructed as follows:

(i) An origin was defined in the usual way generating the root node. The enantiomorph was left undefined. The following procedure defines it *de facto*.

(ii) Four acentric reflections were given permuted phases generating a 256 node second level. No pseudophase permutation was employed at this stage. A node with a U-weighted mean phase error of 24° was in the top 8 as ranked by LLG analysis and was selected at this point. As in the case of SAPO-40, we could have generated a more complex tree here using all the eight LLG preferred nodes as parents for a third level but for the same reasons only the correct node was used as the parent for level 3 of the tree.

(iii) One acentric reflection was given permuted phases and an overlap involving the 541, 10,3,0 and 432 acentric reflections was given true and pseudophase permutation. This generated $4^4 \times 4 = 1024$ nodes, all of which were subjected to entropy maximization.

(iv) The Duncan procedure operating at a 5% confidence level indicated pseudophase angles of 72 and 18° for the two angles. The true values are 79 and 13°, respectively. This is a very encouraging result indeed.

6.3. *$KAlP_2O_7$*

The powder data for this structure were extracted from the JCPDS database (McMurdie *et al.*, 1986). The crystal structure was originally solved from single-crystal data (Ng & Calvo, 1973). There are 84 unique non-overlapped reflections and 53 sets of overlaps totalling 133 reflections. The space group is $P2_1/a$ with a = 8.046, b = 9.657, c = 7.331 Å and β = 106.93°; there are one K, one Al, two P and seven O atoms in the asymmetric unit. This structure was originally used as a test of the maximum-entropy method in powder diffraction (Gilmore *et al.*, 1991). A three-level phasing tree was constructed as follows:

(i) An origin was defined using three reflections to generate the root node.

(ii) Eight reflections (the space group is centrosymmetric) were given permuted phases generating a 256 node second level. A node with a U-weighted mean phase error of 0° was in the top eight as ranked by LLG analysis and was selected, by the same arguments as used above, for generating level 3 of the tree.

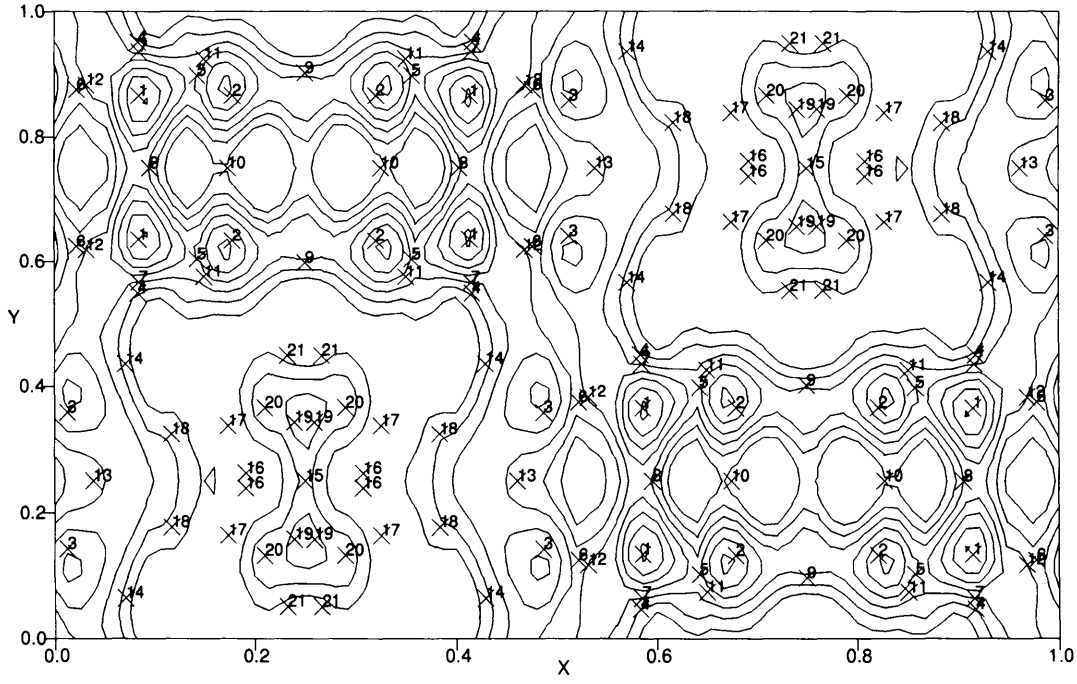

Fig. 2. The centroid map for the best node for SAPO-40 projected down the c axis. The pore is clearly visible. The crosses numbered 15–21 are the TPA^+ (tetrapropylammonium hydroxide) ions in the 12-ring channel.

(iii) Five reflections were given permuted phases, and an overlap involving the 402 and $52\bar{2}$ reflections was given true and pseudophase permutation. This generated 128 nodes, which were subjected to entropy maximization.

(iv) The Duncan procedure operating at a 5% confidence level indicated pseudophase angles of either 54 or 72°. The mean of these indications is 63°. The true value established from the refined structure is 49°. This is a good agreement. The best centroid map shown in projection down the z axis with the atomic coordinates indicated as crosses is shown in Fig. 3. The positions of all but one O atom are correctly indicated.

7. Summary and conclusions

The use of phasing trees, entropy maximization and likelihood evaluation is able to predict correctly the amplitudes of reflections in powder overlaps when Duncan's method of multiple significance testing is used. The accuracy with which this can be carried out is sometimes quite remarkable. Since we are only using the simplest of likelihood functions in which a Rice distribution is employed in an equal-variance diagonal form, the method is probably capable of further refinement by using more elaborate likelihood functions (Bricogne, 1991, 1997*a*). A current difficulty is that the LLGs are frequently not sensitive to the exact partitioning of the overlap intensities but more sophisticated likelihood functions may address this problem also. In its current form, the Duncan procedure is routinely available as a command in the latest version of the *MICE* computer program and it is now routine within this software to include and analyse overlap partitions in the calculations. There is a computational problem, however, in that large phasing trees are a necessary consequence of pseudophase permutation and there is an additional difficulty that with such complexity it can become harder to recognize the true solution(s). The former problem is becoming less important. On our local network of 15 processors on a Unix subnet, it is a simple matter to explore over 10 000 nodes in 24 h and the problem of finding the correct solution can be surmounted by an automatic recycling scheme within a maximum-entropy environment. This has yet to be fully programmed but it does seem that there is now an additional formalism available to process overlapped data when solving structures from powder diffraction data.

We thank Lynne McCusker and Christian Baerlocher for the SAPO-40 data, Phil Lightfoot, Maryjane Tremayne and Peter Bruce for the remaining data sets used here, and the EPSRC for financial support.

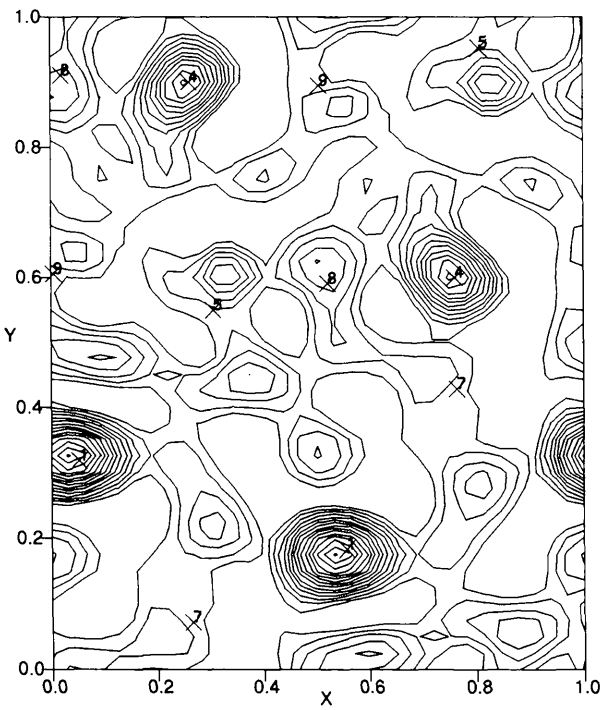

Fig. 3. A centroid map slice at around $z = 0.27$ for the LLG preferred node for $KAlP_2O_7$. The cross labelled 1 is a K^+ ion, 4 is a P atom and the remainder are O atoms.

References

Altomare, A., Cascarano, G., Giacovazzo, C., Guagliardi, A., Burla, M. C., Polidori, G. & Camalli, M. (1994). *J. Appl. Cryst.* **27**, 435–436.

Andreev, Y. G., Lightfoot, P. & Bruce, P. G. (1996). *J. Chem. Soc. Chem. Commun.* pp. 2169–2170.

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.

Bricogne, G. (1991). *Acta Cryst.* A**47**, 803–829.

Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.

Bricogne, G. (1997*a*). *Methods Enzymol.* **276**, 361–423.

Bricogne, G. (1997*b*). *Methods Enzymol.* **276**, 424–448.

Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* A**46**, 284–297.

Burger, K., Prandl, W. & Doyle, S. (1997). *Z. Kristallogr.* **212**, 493–505.

Cochran, W. G. & Cox, G. M. (1957). *Experimental Designs.* New York: John Wiley.

Dumont, N., Gabelica, Z., Derouane, E. G. & McCusker, L. (1993). *Microporus Mater.* **1**, 149–160.

Duncan, D. B. (1955). *Biometrics*, **11**, 1–42.

Estermann, M. A. & Gramlich, V. (1993). *J. Appl. Cryst.* **26**, 396–404.

Estermann, M. A., McCusker, L. B. & Baerlocher, C. (1992). *J. Appl. Cryst.* **25**, 539–543.

Gilmore, C. J. (1984). *J. Appl. Cryst.* **17**, 42–46.

Gilmore, C. J. (1988). *J. Appl. Cryst.* **21**, 571–572.

Gilmore, C. J. (1993). *Crystallographic Computing 6: a Window on Modern Crystallography*, edited by H. D. Flack, L. Párkányi & K. Simon, pp. 26–46. Oxford University Press.

Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.
Gilmore, C. J. & Bricogne, G. (1997). *Methods Enzymol.* **277**, 65–78.
Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *Acta Cryst.* A**46**, 297–308.
Gilmore, C. J., Bricogne, G. & Dong, W. (1998). Submitted.
Gilmore, C. J., Henderson, K. & Bricogne, G. (1991). *Acta Cryst.* A**47**, 830–841.
Gilmore, C. J., Marks, L. D., Grozea, D., Collazo, C., Landree, E. & Twesten, R. (1997). *Surf. Sci.* **381**, 77–91.
Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Proc. R. Soc. London Ser. A*, **442**, 97–111.
Harris, K. D. M. & Tremayne, M. (1996). *Chem. Mater.* **8**, 2554–2570.
Holmes, K. C., Stubbs, G. J., Mandelkow, E. & Gallwitz, U. (1975). *Nature (London)*, **254**, 192–196.
Lightfoot, P., Tremayne, M. J., Harris, K. D. M. & Bruce, P. G. (1992). *J. Chem. Soc. Chem. Commun.* p. 1012.
Lightfoot, P., Tremayne, M., Harris, K. D. M., Glidewell, C., Shankland, K., Gilmore, C. J. & Bruce, P. G. (1993). *Mater. Sci. Forum*, **133–136**, 207–212.
McCusker, L. B & Baerlocher, C. (1995). *Microporus Mater.* **6**, 51–54.
McMurdie, H. F., Morris, M. C., Evans, E. H., Paretzkin, B., Wong-Ng, W., Ettlinger, L. & Hubbard, C. R. (1986). *Powder Diffr.* **1**, 64–77.
MacWilliams, F. J. & Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes.* Amsterdam: North-Holland.
Ng, H. N. & Calvo, C. (1973). *Can. J. Chem.* **51**, 2613–2620.
Pearson, E. S. & Hartley, H. O. (1966). Editors. *Biometrika Tables for Statisticians*, Vol. I. Cambridge University Press.
Shankland, K. (1994). *Trans. Am. Crystallogr. Assoc.* **30**, 29–39.
Shankland, K., David, W. I. F. & Csöka, C. (1997). *Z. Kristallogr.* **212**, 550–552.
Shankland, K., Gilmore, C. J., Bricogne, G. & Hashizume, H. (1993). *Acta Cryst.* A**49**, 493–501.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Sivia, D. S. & David, W. I. F. (1994). *Acta Cryst.* A**50**, 703–714.
Stubbs, G. J. & Diamond, R. (1975). *Acta Cryst.* A**31**, 709–718.
Tremayne, M., Kariuki, B. M. & Harris, K. D. M. (1997). *Angew. Chem. Int. Ed. Engl.* **36**, 770–772.
Tremayne, M., Lightfoot, P., Glidewell, C., Mehta, M. A., Bruce, P. G., Harris, K. D. M., Shankland, K., Gilmore, C. J. & Bricogne, G. (1992). *J. Solid State Chem.* **100**, 191–196.
Tremayne, M., Lightfoot, P., Harris, K. D. M., Shankland, K., Gilmore, C. J., Bricogne, G. & Bruce, P. G. (1992). *J. Mater. Chem.* **2**, 1301–1302.