Acta Cryst. (1991). A47, 830-841

A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. IV. The *Ab Initio* Solution of Crystal Structures from their X-ray Powder Data

By C. J. Gilmore and K. Henderson

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

and G. Bricogne*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(*Received* 22 *October* 1990; *accepted* 2 *May* 1991)

Abstract

The multisolution method of phase determination combing entropy maximization and likelihood evaluation, previously developed for and applied to single-crystal X-ray studies [Bricogne & Gilmore (1990). *Acta Cryst.* A46, 284-297; Gilmore, Bricogne & Bannister (1990). *Acta Cryst.* A46, 297-308], is here extended to permit structure determination from X-ray powder diffraction data using the formulae derived in the previous paper [Bricogne (1991). *Acta Cryst.* A47, 803-829]. Traditionally, structures are difficult to solve *ab initio* from powder diffraction data because of peak overlaps, which arise accidentally or are imposed by point-group symmetry. Overlaps reduce both the effective sampling of reciprocal space and the resolution of the data; this makes the application of traditional direct methods difficult. In the method of combined entropy maximization and likelihood evaluation described here, the intensity data are normalized using both the overlapped and non-overlapped reflections by means of a suitably modified version of the *MITHRIL* computer program [Gilmore (1984). *J. Appl. Cryst.* 17, 42-46; Gilmore & Brown (1988). *J. Appl. Cryst.* 22, 571-572]. The data are then passed to the maximum-entropy program *MICE* [Gilmore, Bricogne & Bannister (1990). *Acta Cryst.* A46, 297-308]. Following origin and enantiomorph definition (if relevant), this builds a phasing tree in which nodes of the tree represent phase permutations of basis-set reflections which are used as constraints in entropy maximization. The nodes of this tree are ranked according to a likelihood criterion evaluated by a new expression capable of using the combined intensities of overlapped reflections. Successive nodes are built *via* continuing phase permutation, keeping only those solutions for which the likelihoods are large. Centroid maps are used to determine atomic positions. The method is applied to two data sets from known structures: $KAlP_2O_7$ [McMurdie, Morris, Evans, Paretzkin, Wong-Ng, Ettlinger & Hubbard (1986). *Powder Diffr.* 1(2), 64-77] collected using a conventional X-ray source; and the Sigma-2 clathrasil [McCusker (1988). *J. Appl. Cryst.* 21, 305-310] which is a synchrotron-derived data set. In both cases the structures are solved routinely and show even some of the light-atom positions in the final maps. The Sigma-2 map even shows the positions of some C and N atoms in the disordered 1-aminoadamantane molecule present in the cavity. The phasing tree for $KAlP_2O_7$ reveals the structure after 29 nodes have been computed whilst Sigma-2 shows a complete structure after 17 nodes. Furthermore, the inclusion of overlapped reflections in the likelihood calculations turns out to be essential - nodes which cannot be distinguished when overlaps are not present are readily and correctly ranked when the latter are included. The centroid maps computed with the inclusion of overlapped reflections show a significant improvement in signal-to-noise ratio over those in which overlapped reflections are omitted. We conclude that because of its stability at any resolution range, this method has the potential to be the most powerful technique available for solving structures from their powder diffraction data.

0. Introduction

Conventional direct methods have had only a limited impact on the *ab initio* solution of structures from their X-ray or neutron powder diffraction data. The following quotation appears in a review article by Shirley (1984): 'the state of the art in *ab-initio* powder structure [determination] is now at about the stage that single-crystal structure analysis had reached by *ca* 1950'.

The reasons for this are not difficult to understand. In powder diffraction, a three-dimensional data set is projected into one dimension where it is spherically averaged; as a result, reflections which would otherwise be separately measured overlap and the degree of overlap increases with $(\sin\theta)/\lambda$, thus effectively reducing the resolution of the data to 1·3 Å or greater.

* Permanent address: LURE, Université Paris-Sud, 91405 Orsay, France.

0108-7673/91/060830-12$03.00

The overlap can arise accidentally from the diffraction geometry or as a consequence of point-group symmetry. This state of affairs creates a serious obstacle to the routine application of conventional direct methods, as these need data at atomic resolution. An empirical rule to this effect has been stated by G. Sheldrick: "If less than 50% of the observed data in the resolution range 1·1–1·2 Å are 'observed' the structure will be difficult to solve by conventional direct methods".

There are exceptions to this rule, the most common being very small structures and those containing heavy atoms, but it does follow that the application of direct methods to structure determination from powder diffraction patterns will be a difficult process. There have been, however, some notable successes: McCusker (1988) solved the Sigma-2 clathrasil with 11 non-hydrogen non-disordered atoms in the asymmetric unit using a synchrotron-radiation data set and the direct-methods modules in the *XTAL* system (Hall & Stewart, 1989); Cheetham, David, Eddy, Jakeman, Johnson & Toradi (1986) solved $FeAsO_4$ using high-resolution powder diffraction data from the Rutherford spallation source and the *MITHRIL* computer program (Gilmore, 1984); and Lehmann, Christensen, Fjellvåg, Feidenhans'l & Nielsen (1987) studied $Al_2Y_4O_9$ and I_2O_4 using X-ray and neutron data and the *MULTAN*77 direct-methods program (Main, Lessinger, Woolfson, Germain & Declercq, 1977). In addition, David (1990) has applied maximum-entropy methods to Patterson deconvolution for TiO_2. There are further references to other *ab initio* structure solutions in McCusker (1988).

Sheldrick's rule does not apply to structures solved by maximum entropy (ME) methods, as has been found by Collins & Mahar (1983) in an application to β-lyxose, Bricogne (1984) in a study of phase extension on the small protein crambin and Gilmore & Bricogne (1991) in applications to two small molecules. Indeed, the method works more smoothly in situations of low resolution. This paper thus describes the practical application of the theory of combined entropy maximization and likelihood discussed in the previous paper (Bricogne, 1991; hereafter referred to as I) to two powder diffraction data sets, one collected on a laboratory source and the other collected at a synchrotron. Preliminary reports have already been published (Henderson & Gilmore, 1989; Gilmore & Bricogne, 1991) although there are major differences in this presentation.

§ 1 of this paper outlines the ME-likelihood methods employed to solve structures *ab initio* from their powder data. They are similar to those described by Bricogne & Gilmore (1990) and Gilmore, Bricogne & Bannister (1990), in which a phasing tree is generated after origin and optional enantiomorph definition by the process of phase permutation; each node on the tree represents a set of phase choices for a set of basis reflections which are subjected to constrained entropy maximization. Each node thus has associated with it a unique map $q^{ME}(\mathbf{x})$ which can extrapolate both U magnitudes and their phases for reflections that were not included in the basis set. Likelihood is used to measure the agreement between observed ($|U_\mathbf{h}|^{obs}$) and calculated ($|U_\mathbf{h}^{ME}|$) amplitudes, but, unlike the single-crystal case, the overlapped reflections are included, and, in situations where much of the data are overlapped, this greatly increases the power of discrimination of the likelihood in choosing the correct node(s). The overlap information is also incorporated into the generation of centroid maps, which are used to interpret the structure; again, in cases of severe overlap, there is a significant improvement in the quality of the final maps when overlapped reflections are included. The program used to normalize the data and select the origin-defining reflections is *MITHRIL* (Gilmore, 1984; Gilmore & Brown, 1988), whilst the maximum entropy and likelihood program is *MICE* (Gilmore, Bricogne & Bannister, 1990). The necessary changes to these programs in order to accommodate powder diffraction data are also discussed in § 1.

§ 2 describes the two structures, their crystallographic intensity data and the method of processing these data to get unitary structure factors $|U_\mathbf{h}|^{obs}$ and their associated variances. § 3 describes the structure solution of $KAlP_2O_7$ where the overlapped reflections, which comprise *ca* 50% of the data, play a major role in the likelihood calculations, whilst § 4 discusses the structure solution of the Sigma-2 clathrasil (McCusker, 1988). In this case there are only 13 pairs of overlapped reflections in a data set comprising 245 reflections, but they are nonetheless highly significant. In both Sigma-2 and $KAlP_2O_7$ the data resolution is *ca* 1·3 Å so that Sheldrick's rule is violated. However, the ME method works routinely and the final maps even show both light or disordered atoms in a way that is unusual in conventional direct methods when heavy atoms are present.

Finally, § 5 summarizes the paper and suggests future developments will serve to strengthen the method still further.

1. The method and the computer programs

1.1. *The phasing method*

The technique is an extension of the procedures described by Bricogne & Gilmore (1990) and Gilmore, Bricogne & Bannister (1990).

(i) The data are normalized to give unitary structure factors $|U_\mathbf{h}|^{obs}$. This includes the overlapped reflections. The data are partitioned into two sets $\{N\}$ and $\{O\}$ which are the non-overlapped and overlapped data respectively.

(ii) The origin (and enantiomorph if relevant) is defined by the appropriate reflections which belong

to set $\{N\}$ and satisfy the usual rules and criteria. Convergence mapping, weighted to confer preference to reflections of low resolution, is a suitable technique. These reflections define the current *basis set* $\{H\}$ where $H \subseteq N$; the remaining non-basis-set reflections are assigned to set $\{K\}$.

(iii) The basis-set reflections ($\mathbf{h} \in H$) are used as constraints in the generation of a maximum entropy prior $q^{\mathrm{ME}}(\mathbf{x})$. The latter reproduces the known phases and amplitudes for reflections $\mathbf{h} \in H$, but remains maximally non-committal with respect to the unknown structure factors $\mathbf{h} \notin H$. The construction of $q^{\mathrm{ME}}(\mathbf{x})$ gives rise to phase extrapolation, *i.e.* its Fourier transform can generate $|U_{\mathbf{h}}^{\mathrm{ME}}|$ and $\varphi_{\mathbf{h}}^{\mathrm{ME}}$ for both $\mathbf{h} \in H$ and $\mathbf{h} \in K$, *i.e.*

$$q^{\mathrm{ME}}(\mathbf{x}) \begin{cases} \xrightarrow{\mathrm{FT}} U_{\mathbf{h}\in H}^{\mathrm{ME}} \\ \xrightarrow{\mathrm{FT}} U_{\mathbf{h}\notin H}^{\mathrm{ME}} \end{cases}.$$

The fit between $U_{\mathbf{h}}^{\mathrm{obs}}$ and $U_{\mathbf{h}}^{\mathrm{ME}}$ for $\mathbf{h} \in H$ is of great importance: underfitting reduces the extrapolative power of $q^{\mathrm{ME}}(\mathbf{x})$ whereas overfitting produces artifacts which undermine the likelihood calculations, frequently driving the log-likelihood gains negative [see (v) below]. As a measure of fit we use the reduced-χ^2 statistic:

$$\chi^2 = (2n_a + n_c)^{-1} \sum_{\mathbf{h}\in H} s_{\mathbf{h}}^{-2} \| U_{\mathbf{h}}|^{\mathrm{obs}} - |U_{\mathbf{h}}^{\mathrm{ME}}\|^2.$$

n_a and n_c are the numbers of unique acentric and centric reflections in the basis set respectively. The sum $2n_a + n_c$ is the total number of degrees of freedom. The parameter $s_{\mathbf{h}}$ is a measure of variance and has four components:

$$s_{\mathbf{h}}^2 = \sigma_{\mathbf{h}}^2 + p\varepsilon_{\mathbf{h}}\Sigma \qquad (1)$$

where $\varepsilon_{\mathbf{h}}$ is the standard epsilon factor, $\sigma_{\mathbf{h}}^2$ is the estimated variance of $|U_{\mathbf{h}}|^{\mathrm{obs}}$, Σ is the reciprocal of the refined N_{eff} from likelihood refinement (see § 1.2) and p is an empirical parameter usually set to unity. However, in powder structures N_{eff} can be small since the structures themselves are small, and hence Σ rather large, and in these cases p is adjusted to ensure a sufficient fit.

Entropy maximization is carried out by exponential modelling; a line search was employed until $\chi^2 < 1{\cdot}25$, then the slower plane search algorithm was used with bicubic modelling of both the entropy and the constraint functions to hold the χ^2 at 1·0 [Bricogne & Gilmore (1990), § 2.3].

The larger the product $|U_{\mathbf{h}}|^{\mathrm{obs}}|U_{\mathbf{h}}^{\mathrm{ME}}|$ the more reliable the phase extrapolation will be. The extrapolates can belong to both $\{N\}$ and $\{O\}$.

(iv) When the basis set comprises only three or four reflections, the extrapolation is necessarily weak and most reflections will not have reliably extrapolated magnitudes and phases. As shown by Gilmore, Bricogne & Bannister (1990), adding strongly extrapolated reflections to the basis set causes difficulties because it traps the solution in a local entropy maximum. To proceed, therefore, several reflections $|U_{\mathbf{h}\in N}|$ with large U magnitude are given permuted phases just as in traditional direct methods. These reflections are added to the basis set and thus impose new constraints on entropy maximization. They are chosen on the following basis:

(*a*) Maximum surprise - they are those about which the current $q^{\mathrm{ME}}(\mathbf{x})$ knows least, *i.e.* $|U_{\mathbf{h}}|^{\mathrm{ME}}$ is close to zero.

(*b*) Minimum resolution - we attempt to build molecular boundaries before imposing atomic detail.

(*c*) Optimum increase in the second neighbourhood of the basis-set reflections, *i.e.* the new reflections optimally increase not only the number of reflections in the second neighbourhood, but especially the number which have large or very small U magnitudes.

(*d*) $|U_{\mathbf{h}}|^{\mathrm{obs}}$ as large as possible.

(*e*) Centric reflections are preferred to acentrics.

In this way we build a phasing tree in which each choice of a permuted phase generates a node as shown in Fig. 1. Each set of nodes lying on the same line has the same reflections in the basis set but with different phase angles. They are termed *equivalent nodes*; a set of equivalent nodes is itself derived from an equivalent set.

We are thus constantly updating our $q^{\mathrm{ME}}(\mathbf{x})$ prior. Clearly such a tree could readily become computationally unwieldy since each node has to be subjected to entropy maximization. To prune this tree and identify the promising nodes we use likelihood.

(v) The diagonal form of the likelihood used here has been described for single-crystal data by Bricogne (1984) and later by Bricogne & Gilmore (1990) and has been generalized by Bricogne (1991) to include overlapped reflections. This approximation is adequate for the size of structure accessible to powder structures.

Likelihood also permits the refinement of the Σ parameter which reflects the effective number N_{eff} of atoms in the unit cell. $N_{\mathrm{eff}} = 1/\Sigma$ and is a consequence of both the quality and the resolution of the data. N_{eff} tends to increase as the data resolution increases and as the size of the basis set increases and with it the strength of the extrapolation. Previously we have employed two Σ parameters: Σ_c for centric reflection and Σ_a for acentrics. The generalized likelihood makes this distinction unnecessary. The computational aspects of likelihood estimation and Σ refinement are discussed in § 1.2.

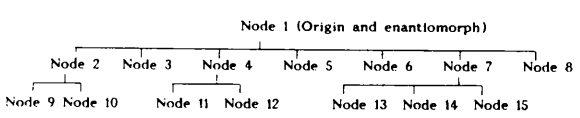

Fig. 1. The early stages of a typical phasing tree.

Nodes with a maximum likelihood and maximum entropy are kept and the others discarded, although *MICE* allows the retracing back to unused nodes if the phasing is proving problematic.

(vi) The process continues until all strong U magnitudes have large extrapolated values, in which case the phasing procedure is complete. It needs to be emphasized that the $q^{ME}(\mathbf{x})$ is not a map in the traditional sense; to produce maps from which atomic coordinates can be extracted, $q^{ME}(\mathbf{x})$ is used to generate a *centroid map* in which reflections belonging to both sets H and K are used and assigned weighted Fourier coefficients; overlapped reflections are also included. The computational aspects of centroid maps are discussed in § 1.3 and the benefits of incorporating overlaps into them are demonstrated in § 3.2.

1.2. *Likelihood*

This is fully described in § 3 of I, but it is useful to summarize some of the salient features, so that the practical implementation of likelihood and Σ refinement can be discussed.

Within an overlap set comprising m reflections (m is unity for a non-overlap situation), let R_i, $i = 1, 2, \ldots, m$, be the observed intensities and r_i, $i = 1, 2, \ldots, m$, be the extrapolated values both with multiplicities μ_i. Define

$$R = \sum_{i=1}^{m} \mu_i R_i^2$$

and

$$r = \sum_{i=1}^{m} \mu_i r_i^2.$$

Let

$$z = Rr/\Sigma$$

and $n = 2n_a + n_c$ be the total number of degrees of freedom. The likelihood, L, for this overlap is then

$$L \propto \{\exp[-(R^2 + r^2)/2\Sigma]/\Sigma^{n/2}\} \times {}_0F_1(-; n/2; z^2/4) \quad (2)$$

(I, § 3) where ${}_0F_1$ is a confluent hypergeometric function. This function has the approximate properties of an exponential in z and to maintain stability for all z (and remember that z is a function of Σ which can take on a wide range of values as phasing progresses), a function $X_n(z)$ is defined:

$$X_n(z) = e^{-z}\,{}_0F_1(-; n/2; z^2/4).$$

The log-likelihood can now be written

$$\log L = -n/2 \log \Sigma - (R - r)^2/2\Sigma + \log X_n(z). \quad (3)$$

It can be visualized as a residual term between observed and calculated magnitudes, with $\log X_n(z)$ acting as an essential correction term.

In *MICE*, $X_n(z)$ is calculated as follows:

(i) By a series expansion for small arguments.

(ii) From closed expressions when n is odd. This is possible because of the relationship between generalized hypergeometric and the modified Bessel functions of half-odd integer order which are themselves expressible in terms of hyperbolic functions.

(iii) *via* numerical approximations to the modified Bessel function of integer order when n is even.

To obtain the total log-likelihood, $L\mathcal{H}$, for the hypothesis $\mathcal{H}$ that the vector of extrapolated structure factors $\mathbf{U}_H^* = \mathbf{U}_H$, (2) is summed over all the symmetry-unique extrapolated reflections $\mathbf{h} \notin H$ both overlapped and non-overlapped:

$$L\mathcal{H} = \sum_{\substack{\text{all} \\ \text{extrapolates}}} [-n/2 \log \Sigma - (R - r)^2/2\Sigma + \log X_n(z)]. \quad (4)$$

For the null hypothesis, termed $\mathcal{H}_0$, $\mathbf{U}_H = \mathbf{0}$, r, and hence z, is set to zero:

$$L\mathcal{H}_0 = \sum_{\substack{\text{all} \\ \text{extrapolates}}} (-n/2 \log \Sigma - R^2/2\Sigma). \quad (5)$$

The log-likelihood gain, abbreviated to likelihood in the remainder of this paper, is obtained as $L\mathcal{H} - L\mathcal{H}_0$, and this is used as the discriminator of nodes.

Both $L\mathcal{H}$ and $L\mathcal{H}_0$ are maximized with respect to Σ. (The values quoted in Tables 2 and 4 are those for $L\mathcal{H}$.) The starting value of Σ is taken from Wilson statistics. The optimization method used is that due to Brent (1973) as implemented by Press, Flannery, Teukolsky & Vetterling (1987). It is a simple yet very robust method that does not employ any calculation of derivatives. The function to be minimized is bracketed between two points, a and b, with x the minimum function point found to date; there are two previous minima, v and w, and u is the point at which the function was last evaluated. The values of x, v and w are used for parabolic interpolation. To be accepted, the parabolic step must:

(i) fall within the limits $[a, b]$;

(ii) involve a movement from x that is less than half the movement of the step before last. This ensures convergence and prevents oscillation.

Because of its simplicity, this method was introduced as an interim measure in program development, but because of its robust qualities we have retained its use. It also prevents the necessity of filtering derivatives which can sometimes be needed when using likelihood in a Newton or quasi-Newton method. The disadvantage is the number of iterations required to achieve convergence, which is usually 10–20, but the calculations take *ca* 1–5 s on our 68030 based workstations, and so this is not important.

For the null hypothesis in the form given, $L\mathcal{H}_0$ is maximized under variation of Σ by

$$\sum(-n/2\Sigma + R^2/2\Sigma^2) = 0$$

which gives the classical estimate of a variance from data with a normal distribution.

The variances of the individual contributors in the form given by (1) in § 1.1 can be incorporated into the likelihood calculations as an option in the *MICE* program. Although the net effect is small, it is considered a useful safeguard and all the results quoted in this paper use these variances.

1.3. *Centroid maps*

There is a full derivation and discussion in I, § 6, but again it is worthwhile to summarize the salient points for discussion in a practical context. The weighted Fourier coefficients q_i have the same phase as the extrapolates r_i and are given as

$$q_i = \frac{R^2}{n\Sigma} \frac{{}_0F_1(-; n/2+1; z^2/4)}{{}_0F_1(-; n/2; z^2/4)} r_i$$

where Σ is the refined value from the maximization of $L\mathcal{H}$. The hypergeometric functions are computed as in § 1.2 above. For the case of a single-crystal reflection this reduces to the standard Sim weighting scheme (Bricogne, 1991) using (4) in § 1.2. In § 3.2 there is a demonstration of how the use of this formula produces maps which are significantly cleaner than those from which the overlaps are omitted. The *MICE* program allows the computation of both overlapped and non-overlapped maps, with the former as the default.

2. The crystal structures and data preparation

Two known structures were studied as a test of the method:

(i) $KAlP_2O_7$. This structure was solved initially from single-crystal X-ray diffraction data (Ng & Calvo, 1973). The powder data were extracted from the JCPDS database (McMurdie, Morris, Evans, Paretzkin, Wong-Ng, Ettinger & Hubbard, 1986). Since JCPDS stores only the largest magnitudes, the smaller structure factors unlisted in the database were assigned zero intensity. There were 84 unique non-overlapped non-zero reflections, 59 unique non-overlapped zero-intensity reflections and 53 sets of overlaps totalling 133 reflections with as many as five reflections under a given envelope. The data resolution is *ca* 1·28 Å. Fig. 2 shows a simulated powder diffraction pattern for these data using peak shapes which conform to Pearson VII functions (Young & Wiles, 1982) and a width at half height of 0·2° held constant across the θ range. Each reflection was assigned a nominal relative variance (σ_n^2) from counting statistics. The data set thus processed represents one which can be readily collected on a good laboratory instrument, although it is somewhat inferior because of the way in which weak reflections have been processed. Because of the large number of overlapped reflections, it is a good test of the likelihood formalism. $KAlP_2O_7$ is not easily solved by routine direct methods; the E maps reveal the Al atom but have a large number of spurious peaks which makes the location of the P atoms especially difficult.

The crystallographic data are as follows: monoclinic, space group $P2_1/a$; $a = 8{\cdot}046$, $b = 9{\cdot}657$, $c = 7{\cdot}331$ Å, $\beta = 106{\cdot}93°$; $Z = 4$. There are 11 atoms in the asymmetric unit: 1×K, 1×Al, 2×P and 7×O.

(ii) The second data set is a considerable contrast – the clathrasil Sigma-2 (McCusker, 1988). A clathrasil is similar in its structure to the more common zeolites but has only cages and no channel system. However, it has all the crystallographic problems posed by zeolites, including a large unit cell with high symmetry, the presence of large voids in the framework as well as disordered organic molecules. Sigma-2 has the molecular formula $[Si_{64}O_{128}] \cdot 4C_{10}H_{17}N$. The data were collected on the synchrotron at Daresbury and comprise 232 unique non-overlapped reflections to a resolution of *ca* 1·3 Å, plus 13 pairs of overlapped reflections. This is representative of the best data sets that can be collected. Indeed, this structure is one of the largest solved *ab initio* from its X-ray powder diffraction pattern, although the process was not routine since it required five attempts using the *XTAL* program. As is often the case for sparse data sets, quartets played a major role in the phasing process.

The crystallographic data are as follows: tetragonal, space group $I4_1/amd$; $a = 10{\cdot}2378$, $c = 34{\cdot}3829$ Å. There are 17 atoms in the asymmetric unit: 4×Si, 7×O, 5×C and 1×N, but the 1-aminoadamantane is disordered so that the C and N atoms

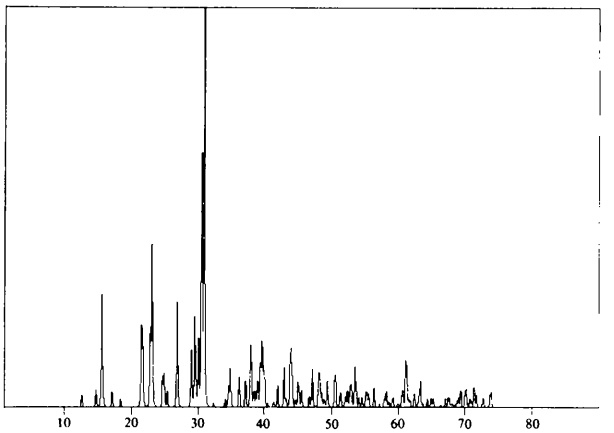

Fig. 2. The simulated powder diffraction pattern for $KAlP_2O_7$. The peak shapes are Pearson VII functions with a width at half height of 0·2°. Since the data were extracted from JCPDS no background modelling is possible.

Table 1. *The phasing tree for* $KAlP_2O_7$

The likelihoods are log-likelihood gains. $\langle|\Delta\varphi|\rangle$ is the U-weighted mean absolute phase error in degrees. Sets of equivant nodes are delineated by single lines ——. A double line ═══ indicates a level in the phasing tree at which the node with the optimum likelihood shows most of the molecular structure *via* a centroid map in which the overlapped reflections are included.

Node	To node	Entropy	Σ	Likelihood (no overlaps)	Likelihood (overlaps)	$\langle\|\Delta\varphi\|\rangle$
1		−2·39	0·03074	0·20	0·96	0
2	1	−3·96	0.04104	0·09	1·30	77
3	1	−4·11	0·04245	−0·04	1·13	46
4	1	−4·12	0·04238	0·08	1·15	105
5	1	−3·96	0·04122	0·03	1·29	74
6	1	−3·68	0·04337	−0·05	0·94	51
7	1	−4·13	0·04421	−0·05	0·91	20
8	1	−4·09	0·04423	0·03	0·91	80
9	1	−3·72	0·04337	0·03	0·95	48
10	1	−4·31	0·04520	0·00	0·75	56
11	1	−3·81	0·04473	0·08	0·75	25
12	1	−3·84	0·04464	0·09	0·76	84
13	1	−4·29	0·04514	0·00	0·74	53
14	1	−3·82	0·04185	−0·12	1·23	31
15	1	−3·78	0·03985	0·00	1·50	0
16	1	−3·79	0·03963	0·00	1·51	59
17	1	−3·85	0·04180	−0·12	1·22	28
18	2	−16·23	0·03099	0·63	3·84	92
19	2	−26·39	0·03877	0·30	2·00	65
20	5	−14·03	0·03446	0·29	3·23	90
21	5	−30·05	0·03717	0·55	2·27	63
22	14	−34·48	0·03608	−0·08	2·52	53
23	14	−11·49	0·03426	−0·39	3·05	26
24	15	−24·37	0·03452	0·05	2·49	27
25	15	−12·61	0·06021	−0·15	3·72	0
26	16	−23·47	0·03528	0·02	2·56	77
27	16	−13·60	0·03077	0·17	3·50	50
28	17	−34·16	0·03611	−0·06	2·90	51
29	17	−13·09	0·03402	−0·17	2·93	24
30	23	−9·69	0·06357	−0·33	3·44	33
31	23	−9·45	0·06404	−0·22	3·77	46
32	23	−9·27	0·05958	−0·56	2·69	20
33	23	−9·02	0·06004	−0·51	3·10	34
34	25	−10·10	0·06072	−0·23	4·13	12
35	25	−10·09	0·06340	−0·28	3·39	25
36	25	−10·02	0·05291	−0·05	4·44	0
37	25	−9·79	0·05524	0·03	3·94	13
38	27	−10·44	0·05410	0·23	2·79	51
39	27	−10·22	0·05389	0·59	3·74	65
40	27	−10·54	0·06309	−0·04	3·88	39
41	27	−10·38	0·06313	0·15	4·29	52
42	34	−17·43	0·05536	−0·01	3·44	21
43	34	−17·17	0·05546	0·07	3·94	32
44	34	−16·84	0·05947	0·60	5·03	10
45	34	−14·93	0·06209	0·50	4·33	22
46	36	−16·04	0·05139	0·25	2·31	10
47	36	−14·99	0·05177	0·56	2·74	22
48	36	−15·53	0·05333	1·05	6·10	0
49	36	−15·14	0·05336	1·41	6·13	11
50	41	−16·88	0·06052	0·38	4·35	57
51	41	−16·79	0·05989	0·52	4·50	68
52	41	−16·34	0·06141	0·52	4·11	46
53	41	−15·29	0·06107	0·71	4·03	57
54	48	−25·21	0·02265	3·15	8·79	7
55	48	−19·44	0·02380	1·55	6·88	0
56	49	−23·09	0·02338	2·18	6·95	18
57	49	−19·40	0·02386	3·25	9·12	10
58	49	−21·97	0·02401	2·11	6·24	16

cannot reasonably be expected to be located by direct methods.

Both data sets were normalized using *MITHRIL* in its latest release (Gilmore & Brown, 1988). In this version overlapped reflections can be included in the normalization process: they are simply partitioned according to reflection multiplicities, since the normalization process depends only on $(\sin^2\theta)/\lambda^2$ and the multiplicity, not on reflection index.

For $KAlP_2O_7$ the inclusion of overlapped reflections made a dramatic change in the normalization. When overlapped reflections were not included, the process was unstable and gave an overall isotropic displacement factor, B, of $-2{\cdot}20\ \text{Å}^2$, whereas inclusion of the overlaps resulted in a value of $1{\cdot}15\ \text{Å}^2$ and realistic intensity statistics. For Sigma-2, with only 13 pairs of overlapped reflections, their inclusion had only a marginal impact. However, there was a problem in normalizing these data - the overall isotropic displacement factor was calculated at $-2{\cdot}52\ \text{Å}^2$. This is probably, in part, a consequence of absorption effects. The radiation used had $\lambda = 1{\cdot}5468$ Å; at this wavelength μ/ρ for Si is 61 cm^{-1}; the sample was held in a 1 mm capillary so that sample absorption is very high. No attempt was made to impose a different overall temperature factor on these data as it was deemed important to process these structures with minimal user intervention.

MITHRIL was also used to estimate the variance of the E magnitudes (and U magnitudes) by the approximate method of Hall & Subramanian (1982). These variances are needed in the *MICE* program for inclusion in the plane search algorithm and in the computation of the reduced-χ^2 statistic.

In both cases, normalization was followed by triplet and quartet generation and a $(\sin\theta)/\lambda$-weighted convergence map in which reflections of low resolution are given a higher weight than those at high resolution. In ME methods it is always advantageous to commence with low-resolution reflections as constraints. These define a molecular envelope upon which atomic detail can be imposed later by successive increases in data resolution. When this is done the whole process becomes much more stable than situations in which reflections at atomic resolution are incorporated into the basis set at the very beginning, although for some structures the distribution of intensity as a function of $(\sin\theta)/\lambda$ makes it impossible to proceed this way. The origin thus selected by convergence mapping was used to generate the first node of the phasing tree. To facilitate comparison between phases extrapolated by entropy maximization and the true values, origin-defining reflections were assigned the true phase angles from the refined structure.

3. Phasing $KAlP_2O_7$

3.1. *The ab initio phasing of* $KAlP_2O_7$

Table 1 shows the phasing tree for $KAlP_2O_7$. This section describes the phasing in detail.

(i) Three reflections (144, 071 and 261) with $|U_h|^{obs}$ values of 0·55, 0·53 and 0·32 respectively and having a resolution of 1·33 Å were used to define the origin. Entropy maximization using a p factor of 0·5 [(1),

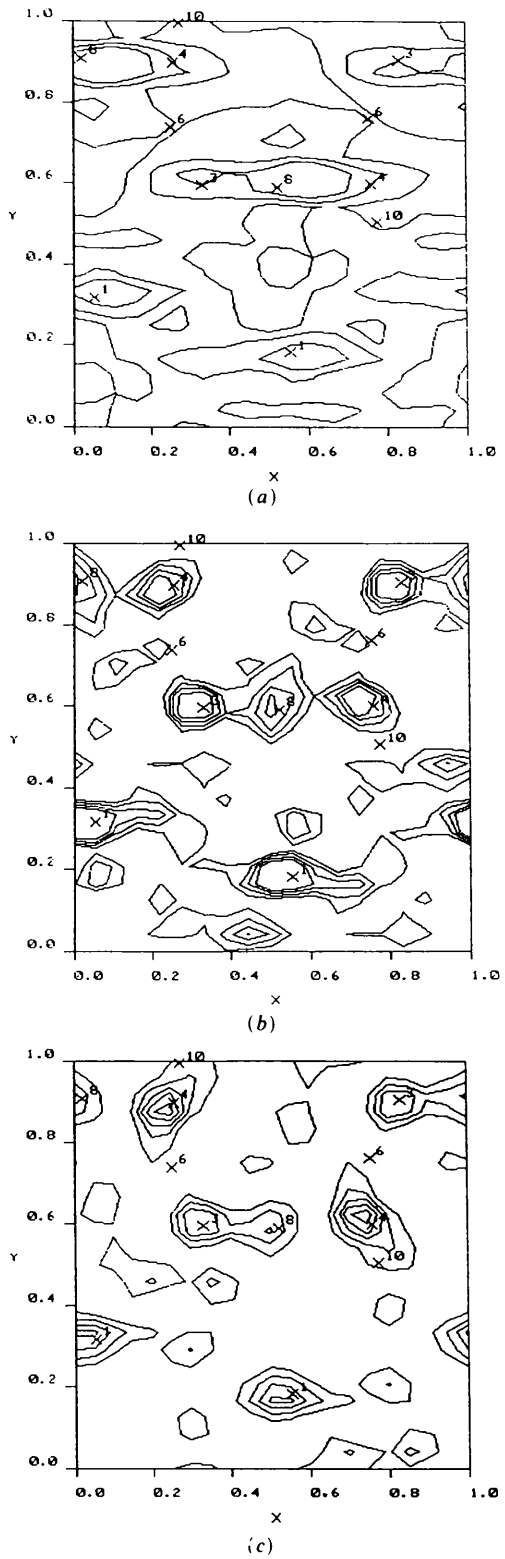

Fig. 3. Sections of centroid maps centred on $z = 0{\cdot}0$ for (*a*) node 1, (*b*) node 15, (*c*) node 48 for $KAlP_2O_7$. Atom 1 is K; atoms 2 and 3 are P, atom 4 is Al, the remainder are O atoms. Overlaps are included in the calculations. All maps are plotted using the *PLOTQ* program (Henderson, Bannister & Gilmore, 1990).

§ 1.1] generated node 1. Fig. 3(*a*) shows a typical section of the resulting centroid map which used 30 reflections including symmetry equivalents. The position of the K atom is already clearly defined. The quality of the phase extrapolation at this early stage is also quite remarkable: there are 14 extrapolated reflections having $|U_h|^{obs}|U_h^{ME}| > 0{\cdot}001$ and only two of these have incorrect phases. This is partly a consequence of the large U magnitudes that are available for origin definition.

(ii) Using the criteria of § 2, four reflections were then incorporated into the starting set with permuted phases (0 or 180°) as follows:

144 $|U_h|^{obs} = 0{\cdot}544$, resolution = 1·3 Å
$36\bar{2}$ $|U_h|^{obs} = 0{\cdot}494$, resolution = 1·3 Å
400 $|U_h|^{obs} = 0{\cdot}443$, resolution = 1·9 Å
431 $|U_h|^{obs} = 0{\cdot}376$, resolution = 1·5 Å.

This generated 16 nodes (nos. 2–17). It can be seen from Table 1 that log-likelihood gains from which the overlapped reflections are excluded discriminate very poorly, with values ranging from −0·12 to 0·09. However, when the overlapped reflections are included, the log-likelihoods increase by an order of magnitude. Acting conservatively, nodes 2, 5, 14, 15, 16 and 17 were retained. Note that the entropy is not a wholly reliable indicator of phase correctness – the entropy-preferred node has two incorrect phases. Two nodes, 15 and 16, both had likelihoods of *ca* 1·5. A centroid map based on node 15 clearly shows the K and Al positions with the P atoms less well defined, although there are peaks at these atomic sites. [See Fig. 3(*b*).]

(iii) Reflection 442 with $|U_h|^{obs} = 0{\cdot}56$ was permuted for each of the retained nodes in (ii) to give nodes 18–29. The basis set now has the full data resolution of 1·28 Å. The overlapped likelihood selected three possible nodes: 23, 25 and 27. Note that, in this case, the entropies for nodes 18–29 also act as discriminators. As described in Gilmore, Bricogne & Bannister (1990), we use likelihood as the primary discriminator; if there are nodes with similar likelihoods but very contrasting entropies, then the latter can be used as a secondary figure of merit. In this case nodes 18 and 20 have entropies significantly lower than those for nodes 23, 25 and 27 and for this reason the former were not selected. It can be seen, however, that entropy alone is an uneven discriminator: sometimes it is successful, but more often not. At this point the solution with the highest likelihood showed the correct positions of the K, Al and P atoms; however, it was decided to pursue the calculations to search for the O atoms.

(iv) Three further reflections were now permuted: 224, 162 and 214 with U magnitudes of 0·35, 0·35 and 0·33 respectively. This would have generated 24 nodes. However, the $P(\delta q)$ filter was used. [See

Table 2. *The U-weighted mean absolute phase error,* $\langle|\Delta\varphi|\rangle$, *as a function of the product* $|U_h|^{obs}|U_h^{ME}|$ *for nodes* 55 *and* 57 *in the phasing tree for* $KAlP_2O_7$

The final entry gives the values for all extrapolates with a $|U_h|^{obs}|U_h^{ME}|$ product >0·001.

	Node 57		Node 55	
$\|U_h\|^{obs}\|U_h^{ME}\|$ limits	$\langle\|\Delta\varphi\|\rangle$ (°)	Number of entries	$\langle\|\Delta\varphi\|\rangle$ (°)	Number of entries
0·0010-0·0049	89	19	93	17
0·0050-0·0099	65	11	104	11
0·0100-0·0199	61	15	63	15
0·0200-0·0249	0	6	0	3
0·0250-0·0299	0	4	0	3
0·0300-0·0449	0	5	0	5
0·0500-0·0689	0	5	0	6
0·0010-0·0689	42	65	49	60

Gilmore, Bricogne & Bannister (1990), § 5.1.] In this method a set of reflections, such as the three cited above, are chosen and their phases permuted in the usual way. Each permutation gives rise to $\delta q(\mathbf{x})$ which is a Fourier synthesis using $U_h^{obs} - U_h^{ME}$ as coefficients. $P(\delta q)$ is calculated as

$$P(\delta q) \propto \int \delta q(\mathbf{x})^2 / q^{ME}(\mathbf{x})\, d^3\mathbf{x}$$

(Bricogne, 1984). A minimum value is required for P which then acts as a filter: only those sets below a specified maximum value are passed to entropy maximization and become part of the phasing tree. The method is very fast compared to the time taken for entropy maximization, requiring only one Fourier synthesis and a map division. The method, however, is only of value when it produces values of P which have sufficient contrast. In most of the steps in this calculation, this was not the case. Application of the filter to these three nodes with the three permuted reflections produced 12 nodes, 30–41, which were then subjected to entropy maximization. Three nodes had likelihoods >4·0 and were retained.

(v) Two further reflections (421 and 135) with U magnitudes 0·34 and 0·32 were permuted. The P filter gave no useful discrimination so that 12 further nodes were generated numbered 42–53. Only two had log-likelihood gains in excess of 6·0. A section from node 48 is shown in Fig. 3(*c*). The map is very clean with the positions of four of the O atoms now clearly indicated.

(vi) A final permutation of 023, $|U_h|^{obs} = 0{\cdot}26$, and 150, $|U_h|^{obs} = 0{\cdot}20$, coupled with the P filter gave five nodes numbered 54–58. There are two points of interest here. First the refined Σ parameter shows a dramatic fall, which is typical behaviour when a great deal of correct phase information has accumulated and the ME extrapolation becomes more and more exact. Secondly, the node selected as best by the likelihood has one wrong phase in the basis set. However, this node (57) produces better maps and has better-quality extrapolation than node 55 in which all the basis-set phases are correct. For node 57 there are 65 extrapolated U magnitudes having $|U_h|^{obs}|U_h^{ME}| > 0{\cdot}001$ of which 19 are wrong, whilst for node 55 there are 60 such extrapolates of which 21 are wrong. The extrapolation results are tabulated in Table 2.

The final map shows the K, Al and P atoms with the correct relative peak heights, plus the positions of five of the seven O atoms; the rest can easily be placed by simple model building. This is a remarkable result. Traditional direct methods have difficulty in locating light atoms in the presence of strong X-ray scatterers, but here even with relatively poor and limited data most of the light atoms have been located.

3.2. *The inclusion of overlapped reflections in centroid maps*

Since *ca* 50% of the reflections in this data set are overlapped, it is a good test of the usefulness of including overlaps in the centroid maps. Fig. 4 shows

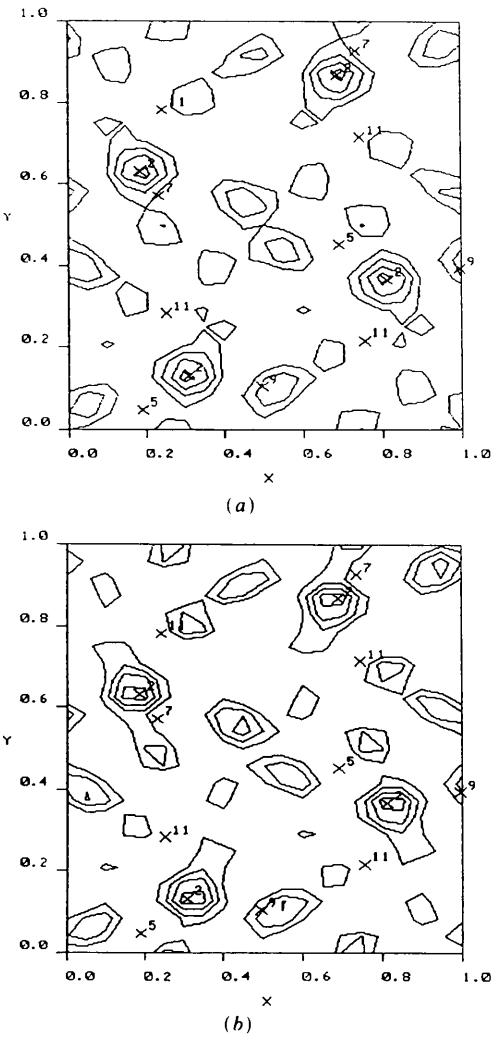

Fig. 4. Sections of centroid maps for $KAlP_2O_7$ centred on $z = 0{\cdot}58$ for node no. 25. (*a*) Overlapped reflections included; (*b*) overlapped reflections excluded.

a typical section of a centroid map for node 25 centred at $z = 0{\cdot}58$; Fig. 4(*a*) includes the overlaps whereas Fig. 4(*b*) excludes them. The former map is significantly cleaner with fewer and weaker spurious peaks. Where maps are subjected to automatic peak search and interpretation, the inclusion of overlapped reflections could be essential, and it is obviously worthwhile to use them as a default.

4. The phasing of Sigma-2

It is a reflection of the quality of this data set that, although Sigma-2 is a more complex structure than $KAlP_2O_7$, it was quite simple to solve using ME-likelihood techniques. Indeed it is relatively easy to solve *via* routine direct methods. The phasing tree is shown in Table 3. The *P* filter was employed at each phase permutation in the following calculations, but proved to be of very limited value, with only weak indication of the preferred sets. This could be a consequence of the very sharp atomic resolution maps that entropy maximization produced even in its earliest stages. Thus no results based on $P(\delta q)$ appear in this section.

(i) The space group is $I4_1/amd$, so that the origin was defined *via* a single reflection (549) with $|U_h|^{obs} = 0{\cdot}21$ and a resolution of 1·5Å. A *p* factor [(1), § 1.1] of 0·3 was used in entropy maximization. The resulting centroid map, based on 16 reflections in the hemisphere, is shown in Fig. 5(*a*). Although the map is noisy, there are peaks corresponding to all the Si and five of the O atoms.

(ii) Reflections 116, 224, 211 and 516 having *U* magnitudes of 0·27, 0·20 and 0·19 respectively and a resolution range of 1·9–4·5 Å were permuted to generate nodes 2–17. The results again show the power that the overlapped reflections add to the likelihood. There are only 13 pairs of overlaps, but some of these reflections are both strong and belong to the second neighbourhood of the basis set, thus playing a key role in the likelihood calculations, particularly in the early stages of phasing. With overlaps excluded, all the nodes have negative likelihoods, but with their inclusion, nodes 6 and 13 have positive likelihoods and are thus retained. Fig. 5(*b*) shows a typical section of the centroid map based on 416 reflections. Its quality is remarkable: it is almost noise free and shows all but two of the ordered Si and O atoms with a clear distinction in peak height between Si and O except for O(1) which has too large a peak. Nonetheless, Sigma-2 is clearly solved at this point, but we continued the calculations to see how much more information could be extracted.

(iii) For each of the nodes 6 and 13, three reflections (549, 4,4,16 and 448) with *U* magnitudes of 0·20, 0·19 and 0·18 respectively and a maximum resolution of 1·38 Å were given permuted phases, thus generating nodes 18–33. On likelihood grounds, nodes 22, 24, 27 and 31 were retained. However, these nodes have entropies of −2·01, −1·90, −2·65 and −2·53 respectively. Using entropy as a secondary indicator, only nodes 22 and 24 were retained with node 24 much preferred. Node 24 is indeed the correct one; extrapolation at this point shows only one wrong phase in the top 21 extrapolates.

Table 3. *The phasing tree for Sigma-2*

See Table 1 for explanatory notes.

Node	To node	Entropy	Σ (×10⁴)	Likelihood (no overlaps)	Likelihood (overlaps)	⟨|Δφ|⟩ (°)
1	—	−0·31	3·6	0·02	0·02	0
2	1	−1·45	64·2	−0·54	−0·46	32
3	1	−1·55	75·6	−0·19	−0·13	78
4	1	−1·29	57·4	−0·88	−0·91	66
5	1	−1·56	77·4	−0·18	−0·24	112
6	1	−1·45	72·9	−0·13	0·04	0
7	1	−1·53	76·6	−0·33	−0·33	47
8	1	−1·34	72·5	−0·55	−0·44	34
9	1	−1·57	75·8	−0·10	−0·09	80
10	1	−1·56	75·5	−0·17	−0·15	64
11	1	−1·41	71·8	−0·31	−0·16	98
12	1	−1·51	76·2	−0·56	−0·56	144
13	1	−1·45	72·5	−0·15	0·00	32
14	1	−1·59	77·1	−0·22	−0·27	78
15	1	−1·33	62·6	−0·58	−0·62	66
16	1	−1·53	76·0	−0·38	−0·32	118
17	1	−1·38	68·3	−0·46	−0·39	112
18	6	−2·01	86·5	−0·83	−0·59	40
19	6	−2·40	89·1	−1·07	−1·03	62
20	6	−1·95	87·7	−0·66	−0·48	19
21	6	−2·32	89·8	−0·55	−0·46	41
22	6	−2·01	87·2	−0·01	0·20	20
23	6	−2·30	88·2	−0·37	−0·23	43
24	6	−1·90	86·8	0·32	0·51	0
25	6	−2·22	88·0	−0·14	0·07	22
26	13	−2·68	89·6	−0·29	−0·13	134
27	13	−2·65	85·7	0·06	0·17	156
28	13	−2·75	90·6	−0·85	−0·82	113
29	13	−2·62	86·5	−0·58	−0·45	135
30	13	−2·42	88·1	−0·78	−0·65	114
31	13	−2·53	83·9	0·05	0·15	136
32	13	−2·44	87·3	−1·43	−1·18	93
33	13	−2·46	83·3	−0·68	−0·44	116
34	22	−3·55	69·6	−4·66	−5·35	47
35	22	−2·75	73·0	0·03	0·09	31
36	22	−3·00	71·6	−1·88	−1·87	33
37	22	−3·01	71·4	0·72	1·02	17
38	24	−3·64	70·2	−2·58	−3·03	30
39	24	−2·78	73·9	0·54	0·44	14
40	24	−2·97	67·8	−1·74	−1·74	16
41	24	−2·96	68·0	1·76	2·08	0
42	41	−4·42	74·3	−0·48	−0·38	37
43	41	−4·45	72·7	0·26	0·35	24
44	41	−4·39	74·0	−0·48	−0·65	24
45	41	−4·15	74·0	−0·16	−0·25	11
46	41	−3·08	71·2	2·49	2·92	26
47	41	−3·36	68·4	4·17	4·71	12
48	41	−3·37	68·9	3·83	4·53	13
49	41	−2·90	69·9	4·50	5·20	0
50	49	−3·67	67·5	2·42	2·51	20
51	49	−4·20	65·0	4·19	4·54	9
52	49	−5·31	68·4	0·44	0·07	30
53	49	−5·30	67·4	2·62	2·78	19
54	49	−3·39	65·9	1·09	0·04	10
55	49	−3·51	62·5	4·52	4·33	0
56	49	−4·42	68·2	2·23	2·18	20
57	49	−4·33	65·0	4·21	4·43	10
58	55	−4·03	58·6	5·28	4·98	7
59	55	−4·47	59·0	4·01	3·47	14
60	55	−3·97	56·0	7·49	6·80	0
61	55	−4·60	55·8	4·05	2·27	7

The real-space nature of these calculations makes the investigation of nodes *via* computer graphics an easy and natural activity; wrong nodes can be spotted by visual inspection of the centroid maps where chemical knowledge is brought into use as well as using likelihood and entropy. As an example of a centroid map for a wrong node, a section is shown in Fig. 5(*d*) for node 32. It can be seen that the signal-to-noise ratio is poor and that the peaks define an Si–O cage structure that is too small to be chemically sensible.

(iv) For each of the nodes 22 and 24, reflections 631 and 3,2,21 having U magnitudes of 0·18 and 0·16 were permuted, generating nodes 34–41. The p parameter was reduced to 0·20 and each node subjected to entropy maximization. Node 41 with a likelihood of 2·08 is strongly preferred and retained. A map at this point is very clear with only the peak corresponding to O(2) being rather weak.

(v) Reflections 4,0,18, 7,0,11 and 635 ($|U_{\mathbf{h}}|^{\text{obs}}$ = 0·18, 0·18 and 0·16 respectively) were given permuted phases. The calculations now extend to the full data resolution. This generates nodes 42–49, of which no. 49 is clearly indicated on both likelihood and entropy grounds. The corresponding centroid map with 1290 contributors is shown in Fig. 5(*c*). The position of O(2) is now indicated thus completing the clathrasil cage structure. However, two further sets of equivalent nodes were generated to see if any indications were available concerning the position of the disordered 1-aminoadamantane molecule.

(vi) Reflections 3,0,17, 721 and 723, all with $|U_{\mathbf{h}}|^{\text{obs}} = 0{\cdot}16$, were permuted to generate nodes 50–57. On likelihood ground, nodes 51, 55 and 57 survive,

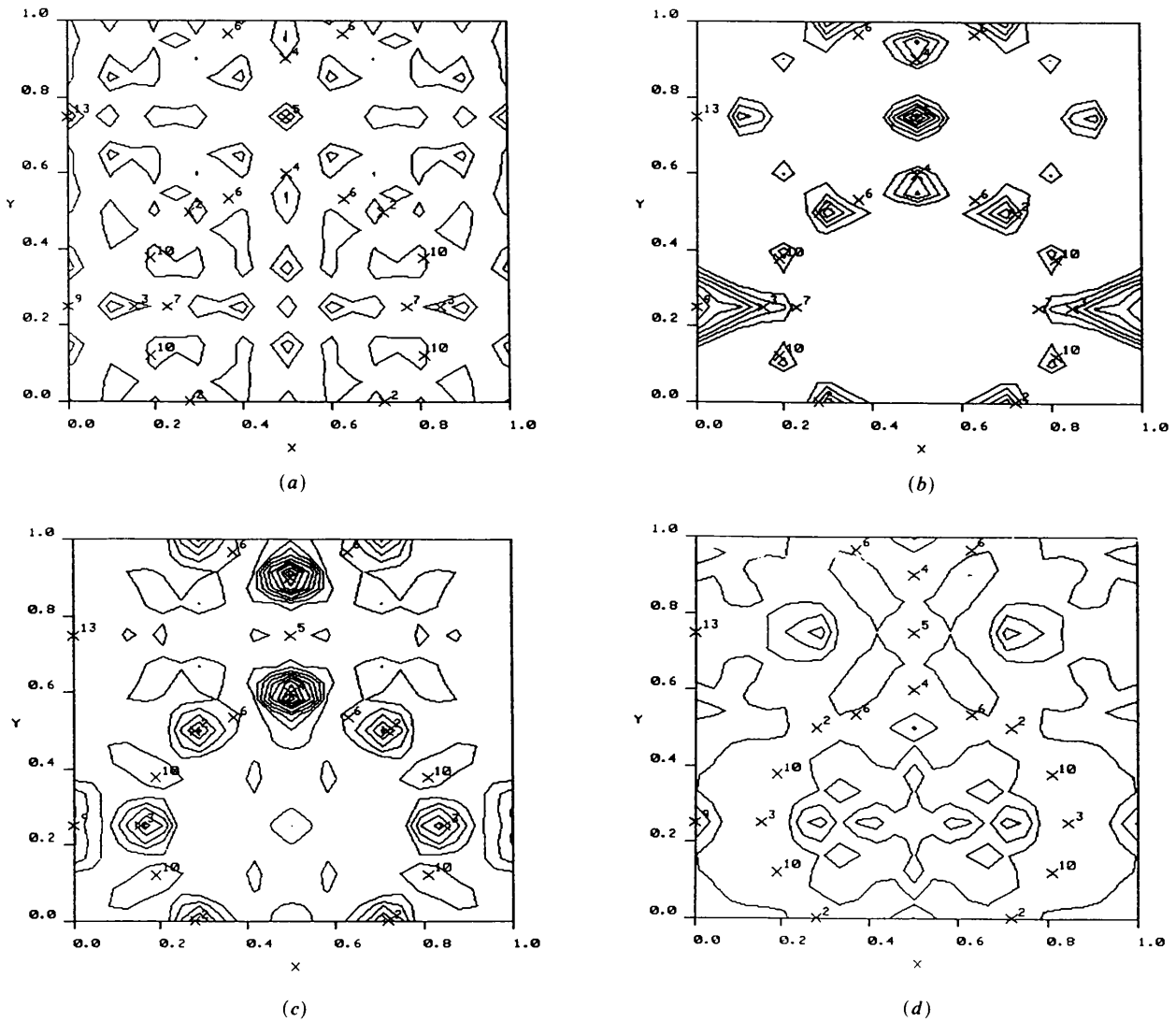

Fig. 5. Sections of centroid maps centred on $z = 0{\cdot}0$ for (*a*) node 1, (*b*) node 6, (*c*) node 60, (*d*) node 32 (an incorrect node) for Sigma-2. Atoms numbered 1–4 are Si, 5–11 are O and the remainder represent the disordered 1-aminoadamantane molecule.

but the entropies are −4·20, −3·51 and −4·33 respectively; thus only node 55 was retained. Phase extrapolation for this node shows that for the 100 reflections having $|U_h|^{obs}|U_h|^{ME} > 0{\cdot}001$ only 16 are incorrect.

(vii) Finally, reflections 3,1,24 and 3,1,16 with U magnitudes of 0·14 and 0·13 respectively were permuted to give nodes 58–61. The p parameter is still 0·20. Node 60 is clearly indicated at this point. A centroid map based on 1570 reflections shows not only the Si–O cluster but a smear of density indicating the location of the 1-aminoadamantane molecule (see Fig. 6). The positions of four of the adamantane atoms are also indicated. Given that this is powder diffraction data and that the basis set comprises only 18 unique reflections, this is quite remarkable. Table 4 summarizes the quality of the phase extrapolation at this point.

There are three further features of this analysis which are worthy of comment:

(*a*) The Σ parameter starts at *ca* 75×10^{-4} (excluding node 1 where the very low value is unreliable because of the weakness of the extrapolation). It rises to 88×10^{-4} for nodes 18–33 as the data resolution and hence the N_{eff} increase; then it falls steadily to *ca* 58×10^{-4} for the last nodes as the phase extrapolation becomes stronger and as the basis set accumulates more and more correct phase information.

(*b*) In later stages of the analysis the likelihood estimate in which overlapped reflections are included sometimes falls *below* that for which overlaps are excluded. This is not unexpected. The overlapped reflections belong to the second neighbourhood of the basis set. As extrapolation increases in strength, some of these reflections may well be overextrapolated, *i.e.* $|U_h^{ME}| > |U_h|^{obs}$, and this depresses the likelihood estimates. The strength of the method is, however, in no way reduced.

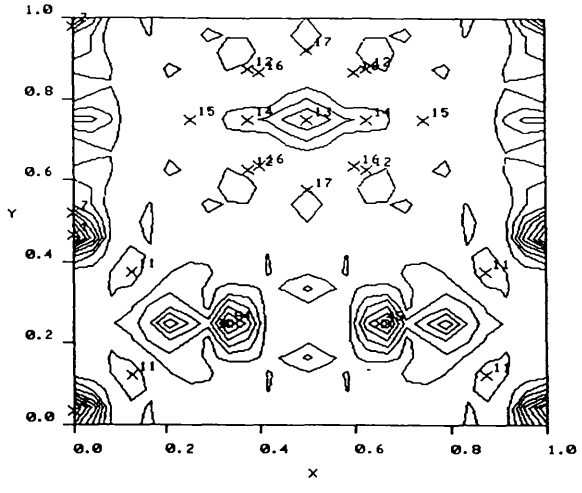

Fig. 6. A section of a centroid map centred at $z = 0{\cdot}93$ for node 60 for Sigma-2. The peaks numbered 12 and above are the refined positions of the disordered 1-aminoadamantane molecule.

Table 4. *U-weighted mean absolute phase error as a function of the product* $|U_h|^{obs}|U_h^{ME}|$ *for node* 60 *in the phasing tree of Sigma-2*

$\|U_h\|^{obs}\|U_h^{ME}\|$ limits	$\langle\|\Delta\varphi\|\rangle$ (°)	Number of entries
0·0010–0·0019	77	35
0·0020–0·0029	59	13
0·0030–0·0039	31	9
0·0040–0·0049	17	9
0·0050–0·0059	0	6
0·0060–0·0069	0	6
0·0070–0·0079	0	5
0·0110–0·0080	0	8
0·0439–0·0109	0	7
0·0010–0·0109	28	98

(*c*) Taking *strongly* extrapolated reflections into the basis set is as disastrous for powder data sets as for the single-crystal case. It pins the entropy into a false maximum, building artefacts into the resulting maps. As an example, for node 55, 36 extrapolated reflections having $|U_h|^{obs}|U_h^{ME}| > 0{\cdot}05$ were incorporated into the basis set with phase angles taken from U_h^{ME}. All these angles were correct. After entropy maximization, however, the likelihood fell to −9·20 and the centroid map showed very large peaks at the centres of ring systems; it was difficult to interpret.

5. Summary and concluding remarks

We have demonstrated the applicability of the combined use of entropy maximization and likelihood estimation to the determination of crystal structures from their X-ray powder diffraction patterns even when the data are relatively poor in quality. The single-crystal likelihood, calculated from non-overlapped reflections, is a poor discriminator in these circumstances, but the new likelihood incorporating the overlap information has proved to be a uniquely powerful figure of merit. Entropy serves as a valuable secondary indicator of preferred phase sets.

One disadvantage of this method is an increased computing requirement when compared to traditional direct methods. However, relatively few nodes were needed for both $KAlP_2O_7$ and Sigma-2: the former was solved after 29 nodes and the latter after only 17; the remaining nodes were generated in an attempt to see how much further information could be extracted from the diffraction data. This was successful: light atoms that would not normally be resolved in traditional direct methods became clearly visible. The computing time was a matter of hours on our Concurrent workstations, which are UNIX based 68030 machines running at 33 MHz. We are currently working to reduce this time substantially by performing the entire process of exponential modelling in memory with a minimum of disk transfers. This will replace the current method which stores all maps on disk and uses intermediate scratch files in the Fourier calculations. Initial indications show that this will

speed up the calculations by a factor of 5-10. We can thus expect all these calculations to take less than one hour of processor time; on the new generation of workstations, from *e.g.* IBM or MIPS, the total time would be less than 15 min. Computer time is thus not an important issue.

Another point needs to be made concerning data quality and completeness. To solve structures *ab initio*, the data quality and its mode of processing are still of paramount importance. Weak reflections, in particular, must be measured carefully and included in the data set with their observed values, not a zero value. The effects of not doing so are well documented for the single-crystal case: it introduces a bias which may upset the whole phasing procedure. The effect would be worse in the powder case, where the data are much more limited and where the sophisticated analysis which underlies the likelihood criterion would be even more vulnerable to systematic bias in the data.

As far as future developments are concerned, paper I has described some other theoretical results which could further enhance the capabilities here. In particular:

(i) The use of a multichannel entropy formalism (Bricogne, 1988) in which the heavy atoms and light atoms are assigned separate channels in the entropy maximization. This would improve our ability to resolve light atoms in the presence of heavy scatterers. Since atoms may have negative scattering factors in this formalism, the program would then also be able to deal with neutron diffraction data in a very natural fashion. This is discussed in detail by Bricogne (1988).

(ii) Hyperoctant phase permutation (I, § 6) would permit the inclusion of *overlapped* reflections into the basis set. For Sigma-2, some very strong reflections were involved in overlaps and it would be valuable to incorporate these into the basis set, thus using them in an active way. Many data sets are predominantly overlapped and this would offer a mechanism of solving such structures.

(iii) The generalization of the 'heavy-atom' method by using likelihood to detect and exploit the presence of known fragments (I, § 5).

(iv) New methods of data normalization for non-uniform distributions and in the presence of known fragments (I, § 5.1).

(v) New criteria for conducting crystal structure refinement against powder data (I, § 8).

The possibility of these future enhancements together with the methods described in this paper make the routine solution of complex structures from powder data a realistic prospect.

This research was initiated following discussions with Dr Rhona Highcock. It was supported by a grant from BP Research *via* an extra-mural research award; we are particularly grateful to Dr M. J. Tricker, Professor K. J. Packer and Dr R. C. Ward at BP Research for their continuing support. GB is indebted to Trinity College, Cambridge, for unfailing logistic support. Dr Lynne McCusker and Dr Michael Esterman at ETH, Zurich, supplied the Sigma-2 data as well as useful discussions on the data processing.

References

BRENT, R. P. (1973). *Algorithms for Minimisation without Derivatives*, Ch. 5. Englewood Cliffs, NJ: Prentice Hall.

BRICOGNE, G. (1984). *Acta Cryst.* **A40**, 410-445.

BRICOGNE, G. (1988). *Acta Cryst.* **A44**, 517-545.

BRICOGNE, G. (1991). *Acta Cryst.* **A47**, 803-829.

BRICOGNE, G. & GILMORE, C. J. (1990). *Acta Cryst.* **A46**, 284-297.

CHEETHAM, A. K., DAVID, W. I. F., EDDY, M. M., JAKEMAN, R. J. B., JOHNSON, M. W. & TORADI, C. C. (1986). *Nature (London)*, **320**, 46-48.

COLLINS, D. M. & MAHAR, M. C. (1983). *Acta Cryst.* **A39**, 252-256.

DAVID, W. I. F. (1990). *Nature (London)*, **346**, 731-734.

GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42-46.

GILMORE, C. J. & BRICOGNE, G. (1991). In *Crystallographic Computing* 5, edited by D. MORAS, A. PODJARNY & J. C. THIERRY. Oxford: Clarendon Press.

GILMORE, C. J., BRICOGNE, G. & BANNISTER, C. (1990). *Acta Cryst.* **A46**, 297-308.

GILMORE, C. J. & BROWN, S. R. (1988). *J. Appl. Cryst.* **22**, 571-572.

HALL, S. R. & STEWART, J. M. (1989). Editors. *XTAL* 2.6 *User's Manual.* Univs. of Western Australia, Australia and Maryland, USA.

HALL, S. R. & SUBRAMANIAN, V. (1982). *Acta Cryst.* **A32**, 598-600.

HENDERSON, K., BANNISTER, C. & GILMORE, C. J. (1990). *J. Appl. Cryst.* **23**, 143-144.

HENDERSON, K. & GILMORE, C. J. (1989). In *Maximum Entropy and Bayesian Methods*, edited by J. SKILLING, pp. 233-239. Amsterdam: Kluwer.

LEHMANN, M. S., CHRISTENSEN, A. N., FJELLVÅG, H., FEIDENHANS'L, R. & NIELSEN, M. (1987). *J. Appl. Cryst.* **20**, 123-129.

MCCUSKER, L. (1988). *J. Appl. Cryst.* **21**, 305-310.

MCMURDIE, H. F., MORRIS, M. C., EVANS, E. H., PARETZKIN, B., WONG-NG, W., ETTLINGER, L. & HUBBARD, C. R. (1986). *Powder Diffr.* **1**(2), 64-77.

MAIN, P., LESSINGER, L., WOOLFSON, M. M., GERMAIN, G. & DECLERCQ, J.-P. (1977). *MULTAN77. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.

NG, H. N. & CALVO, C. (1973). *Can. J. Chem.* **51**, 2613-2620.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1987). *Numerical Recipes. The Art of Scientific Computing*, pp. 283-287. Cambridge Univ. Press.

SHIRLEY, R. (1984). In *Methods and Applications in Crystallographic Computing*, edited by S. R. HALL & T. ASHIDA, pp. 411-437. Oxford: Clarendon Press.

YOUNG, R. A. & WILES, D. B. (1982). *J. Appl. Cryst.* **15**, 430-438.