hence the derivatives:

$$\frac{\partial L_c}{\partial r} = -\frac{r}{\Sigma} + \frac{R}{\Sigma}\tau\left(\frac{rR}{\Sigma}\right), \qquad (A7a)$$

$$\frac{\partial L_c}{\partial \Sigma_c} = -\frac{\varepsilon}{2\Sigma} + \frac{\varepsilon}{2}\frac{r^2+R^2}{\Sigma^2} - \frac{\varepsilon rR}{\Sigma^2}\tau\left(\frac{rR}{\Sigma}\right), \qquad (A7b)$$

$$\frac{\partial^2 L_c}{\partial \Sigma_c^2} = +\frac{\varepsilon^2}{2\Sigma^2} - \varepsilon^2\frac{r^2+R^2}{\Sigma^3} + \left(\frac{\varepsilon rR}{\Sigma^2}\right)^2\tau'\left(\frac{rR}{\Sigma}\right) + \frac{2\varepsilon^2 rR}{\Sigma^3}\tau\left(\frac{rR}{\Sigma}\right). \qquad (A7c)$$

These derivatives are used as described in the acentric case.

References

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.
BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410-445.
BRICOGNE, G. (1988*a*). *Acta Cryst.* A**44**, 517-545.
BRICOGNE, G. (1988*b*). In *Crystallographic Computing* 4, edited by N. W. ISAACS & M. R. TAYLOR, pp. 60-79. Oxford Univ. Press.
BRYAN, R. K. & SKILLING, J. (1980). *Mon. Not. R. Astron. Soc.* **191**, 69-79.
CARTER, C. W. JR, CRUMLEY, K. V., COLEMAN, D. E., HAGE, F. & BRICOGNE, G. (1990). A**46**, 57-68.
COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1-12.
COLLINS, D. M. & MAHAR, M. C. (1983). *Acta Cryst.* A**39**, 252-256.
DIAMOND, R. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 266-272. Oxford Univ. Press.
GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42-46.
GILMORE, C. J., BRICOGNE, G. & BANNISTER, C. (1990). *Acta Cryst.* A**46**, 297-308.
GILMORE, C. J. & BROWN, S. R. (1988). *J. Appl. Cryst.* **22**, 571-572.
GOEDKOOP, J. A. (1950). *Acta Cryst.* **3**, 374-378.
HARKER, D. (1953). *Acta Cryst.* **6**, 731-736.
HAUPTMAN, H. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 151-197. New York: Plenum.
LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Vols. 1 and 2. Cambridge Univ. Press.
MAIN, P. (1977). *Acta Cryst.* A**33**, 750-757.
NEYMAN, J. & PEARSON, E. (1933). *Philos. Trans. R. Soc. London Ser. A*, **231**, 289-337.
RICE, S. O. (1944). *Bell Syst. Tech. J.* **23**, 283-332.
RICE, S. O. (1945). *Bell Syst. Tech. J.* **24**, 46-156.
SHANNON, C. E. & WEAVER, W. (1949). *The Mathematical Theory of Communication.* Urbana: Univ. of Illinois Press.
WAX, N. (1954). Editor. *Selected Papers on Noise and Stochastic Processes*, pp. 133-294. New York: Dover.
WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 53-56.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.
WILSON, A. J. C. (1950). *Acta Cryst.* **3**, 258-261.

Acta Cryst. (1990). A**46**, 297-308

A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. II. Application to Small Molecules

BY C. J. GILMORE

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

G. BRICOGNE

LURE, Université Paris-Sud, 91405 Orsay, France

AND C. BANNISTER

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

(*Received* 8 *August* 1989; *accepted* 15 *November* 1989)

Abstract

The approach described by Bricogne & Gilmore [*Acta Cryst.* (1990). A**46**, 284-297] (I) is applied to three small organic molecules. Phase extension for sucrose octaacetate ($C_{28}H_{38}O_{19}$) from a basis set of 300 correctly phased U magnitudes confirms the stability of the exponential modelling and plane-search algorithms under very demanding conditions; the extrapolated phases are of comparable quality with those produced by the tangent formula, although it is possible, by overfitting the observed and calculated U magnitudes, to obtain results that are better than tangent refinement. The *ab initio* phasing of two small molecules, one (diamantan-4-ol, $C_{14}H_{20}O$) centrosymmetric and the other [(−)-platynecine, $C_8H_{15}NO_2$] non-centrosymmetric, shows that the likelihood function is a more powerful discriminator between phase choices than any figure of merit hitherto available in conventional direct methods; correct discrimination of phase sets arising from phase-angle permutation is readily achieved even in

0108-7673/90/040297-12$03.00

situations where less than ten phased reflexions are in the basis set. In the non-centrosymmetric case, the conditional probability criterion $P(\delta q) \propto \int \delta q(\mathbf{x})^2/q^{ME}(\mathbf{x})\, d^3\mathbf{x}$ [I, equation (1.4)] plays a vital rôle in exploring the multimodality of statistical phase indications and is used as a filter in building the phasing tree. Finally, the likelihood function is successfully used in phase refinement in (−)-platynecine where the mean absolute phase error is reduced by 6·1° for the acentric reflexions using a basis set of only 27 reflexions. Such a calculation would be impossible in traditional direct methods.

0. Introduction

This paper describes the practical application of the approach presented in the previous paper (Bricogne & Gilmore, 1990; hereafter referred to as I) to three small crystal structures. A preliminary report has already been published (Bannister, Bricogne & Gilmore, 1989) although there are significant differences with this paper. Throughout, the term *likelihood* is used to refer to the logarithm of the likelihood ratio $L = L(\mathbf{U}_H^*) - L(0)$ (I, § 1.4) and the symbol Σ [see I, equation (2.10)] is used as a generic term for Σ_a or Σ_c whichever is appropriate; the context should make the usage clear. § 1 describes the structures under study and discusses the data-preparation methods used, in particular the estimation of $\sigma_{\mathbf{h}}^2$, the variance of $|U_{\mathbf{h}}|^{obs}$. § 2 outlines the computer program developed for constrained entropy maximization and likelihood estimation. This program, *MICE*, incorporates the theory developed in I and includes exponential modelling with plane search (I, § 2.3), likelihood estimation (I, § 2.4), Σ refinement (I, § 2.4 and Appendix) and phase refinement (I, § 2.5). § 3 discusses the application of the maximum-entropy algorithms to phase extension in sucrose octaacetate. Although the primary thrust of this paper is *a priori* phasing methods, the problem of phase extension described is a challenging one for any entropy-maximization method. Indeed, it is unlikely that any other entropy-maximization algorithm described in the literature could have coped with it. It also confirms the validity of the underlying theory of the behaviour of Σ refinement (I, § 3.3) and the variation of error in extrapolated phases as a function of the product of the observed and extrapolated **U** magnitudes. § 4 describes the *ab initio* phasing of a small centrosymmetric structure using the concept of a phasing tree coupled with the use of likelihood as a method of selecting the correct nodes of that tree. Likelihood is shown to be a figure of merit of immense power even when very little phase information is available. § 5 extends the method to a small non-centrosymmetric structure. Here the use of $P(\delta q) \propto \int \delta q(\mathbf{x})^2/q^{ME}(\mathbf{x})\, d^3\mathbf{x}$ [I, equation (1.4)] proves to be a necessary filter before likelihood by providing indications of multimodality which are not given by likelihood alone. Likelihood now also takes on an active role in the process by providing a means of phase *refinement*. With only 27 unique basis-set reflexions for this structure, this phase-refinement procedure was able to reduce the mean absolute phase error by 6·1°. The paper is summarized in § 6 including a discussion of extending the method to structures of greater complexity. Although both structures studied in the present work are small and readily solved by traditional direct methods, they provide a valuable test of the theory given in I allowing one to learn how the process of combined maximization of entropy and likelihood works in practice and how it might best be later extended to more complex structures and to poorer data sets (*e.g.* powder diffraction patterns and electron crystallography).

1. The crystal structures and data preparation

The three structures studied were:

(i) Sucrose octaacetate (Oliver & Strickland, 1984), $C_{28}H_{38}O_{19}$, which crystallizes in space group $P2_12_12_1$ with $a = 18{\cdot}350$, $b = 21{\cdot}441$, $c = 8{\cdot}350$ Å and $Z = 4$. This structure is found in the database of difficult structures, compiled by G. Sheldrick, and was used here as a test of phase extension.

(ii) Diamantan-4-ol, $C_{14}H_{20}O$, which crystallizes in space group $P4_2/n$ with $a = 16{\cdot}704$, $c = 7{\cdot}922$ Å and $Z = 8$. This structure is also to be found in the Sheldrick database where it is included principally as an example of a tetragonal space group. Its structure is shown in Fig. 1(*a*). Note that the rings cluster around an effective molecular centroid. This has some significance later.

(ii) A pyrrolizidine necine base (−)-platynecine (Freer, Kelly & Robins, 1987), $C_8H_{15}NO_2$, which crystallizes in space group $P2_12_12_1$ with $a = 7{\cdot}810$, $b = 8{\cdot}348$, $c = 12{\cdot}459$ Å and $Z = 4$. It has the structure shown in Fig. 1(*b*).

For both data sets the structure factors $|F_{\mathbf{h}}|^{obs}$ were normalized using the method of Wilson (1942) to generate normalized structure factors $|E_{\mathbf{h}}|^{obs}$ which were then converted to unitary structure factors

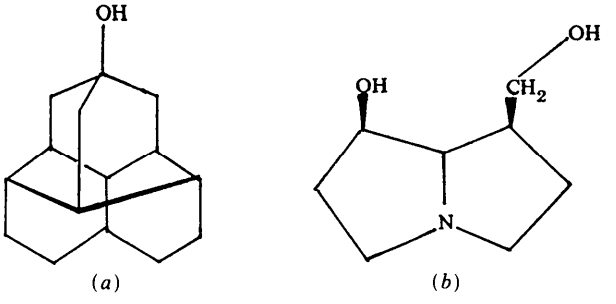


Fig. 1. The structures of (*a*) diamantan-4-ol and (*b*) (−)-platynecine.

$|U_h|^{obs}$. A variance of $|U_h|^{obs}$, σ_h^2, was estimated using the method of Hall & Subramanian (1982) in which contributions to σ_h^2 arise from $\sigma(|F_h|^{obs})$, errors in both the scale factor K [$\sigma(K)$] and the overall isotropic temperature factor B [$\sigma(B)$] and their correlation $r(K, B)$. The latter three parameters are readily estimated from the linear regression that the Wilson plot utilizes. The final expression for σ_h^2 is

$$\sigma_h^2 = |U_h|^2 \{\sigma^2(|F_h|)/|F_h|^2 + \sigma^2(K)/K^2 + s^4\sigma^2(B) + 2s^2 r(K, B)\,\sigma(K)\sigma(B)/K\},$$

where $s = (\sin\theta)/\lambda$.

Although this method of variance estimation is not strictly valid, and it produces error estimates which are much too high, the problem is not serious. This is because σ_h^2 is always added to Σ (Σ_a or Σ_c) whenever it is used and, in general, $\sigma_h^2 \ll \Sigma$ except in situations where Σ has refined to a very small value (see § 3). In these circumstances a solution to the phase problem has been achieved anyway so that there is no practical difficulty. The use of a σ_h^2 estimate still remains useful, however, in giving individual reflexions the correct relative weight even though the absolute values of these weights may be wrong. Better methods of estimating σ_h^2 are under study.

The normalization and σ_h^2 estimation is carried out using a modified version of the *MITHRIL* direct-methods computer program (Gilmore, 1984; Gilmore & Brown, 1988). *MITHRIL* is also used to generate triplets and quartets followed by a convergence map which is used to propose suitable origin- and enantiomorph- (if relevant) defining reflexions. This starting set is used to define the root node (node 1) for the problem. An option in the *CONVERGE* module allows convergence to be resolution dependent, upweighting reflexions at low angle. This is because the entropy/likelihood maximization procedure works best in a situation where low-resolution molecular boundaries are built first, and higher-resolution detail built in subsequent nodes of the phasing tree. In some cases, however, especially for small molecules, it is impossible to find suitable U magnitudes at low angle which are sufficiently large and which interact with other reflexions *via* triplets and quartets; in these circumstances one is compelled to use data at atomic resolution from the outset, which does make phasing more difficult.

2. The *MICE* computer program

2.1. *An outline of the MICE program*

All the calculations carried out in this paper were performed using a program *MICE* (Maximum entropy In a Crystallographic Environment) written in standard Fortran 77 for both VAX VMS and UNIX operating-system environments. It consists of the following modules linked together by a central control routine.

(i) A constrained entropy-maximization program using exponential modelling coupled with plane- and line-search routines. All the damping factors, bumpers and other checks discussed in I (§ 2.3.2.4) are employed. In all cases, default values are set for these parameters in the program; these rarely need to be changed except for handling a large number of basis-set reflexions. A general-purpose FFT routine, derived from the programs of Ten Eyck (1973), is used for all forward and inverse Fourier calculations.

(ii) A likelihood module incorporating likelihood estimation, Σ refinement and phase refinement. Σ refinement uses a Newton method on the logarithm of the likelihood to maintain the convexity of the Σ functions and prevent convergence on a false minimum, which is usually at a negative value of Σ. Likelihood is used to monitor the entropy optimization process for each node at every cycle.

(iii) A module to choose which reflexions are to be incorporated into the basis set for phase permutation using a method of optimal second-neighbourhood extension. (See § 4.1.)

(iv) A set of routines that employs $P(\delta p)$ (I, § 1.2.1) coupled with a filter to examine phase sets produced by phase permutation before they are passed on to exponential modelling. (See § 5.1.)

(v) A module to produce a centroid map (I, § 1.6) coupled with peak search and interpretation routines.

MICE has a similar design philosophy to *MITHRIL* with a mixture of interactive and batch/background modes of operation. The interactive use is menu driven, and although entropy maximization itself is, in general, unsuitable for interactive use of a computer, likelihood calculations, phase refinement, reflexion selection and centroid evaluation are all eminently suited for interactive computation on a modern workstation. Indeed all the calculations in this paper were carried out on MassComp scientific workstations.

MICE operates with an interface to the *MITHRIL* direct-methods program. This permits the input of U magnitudes and their variances from *MITHRIL* into *MICE*, but there are two further interfaces which permit the phases of the basis-set reflexions to be output to the convergence-mapping and tangent-refinement modules of *MITHRIL*. This allows the option of phasing a starting set of 10–20 reflexions *via* maximum entropy and then continuing the phasing using tangent refinement, which is much faster although prone to other problems of overfitting invariants *etc.*

An important feature of entropy methods is the real-space nature of much of the calculation; this permits the investigation of suitable maps by inspecting their contoured densities. A program *PLOTQ* (Henderson, Bannister & Gilmore, 1990) was used

for this purpose. It permits the viewing of 2D sections and projections along any axis and 3D contoured boundaries viewed from any direction to be produced on a wide variety of Tektronix-compatible graphics devices in both colour and monochrome. *PLOTQ* can also operate with the Fourier module of *MITHRIL* to inspect poor-quality E maps; it is sometimes possible to obtain information about atomic positions *via* inspection of contoured maps in a way that is not possible with a simple peak list. The *MICE* program structure with its *MITHRIL* and *PLOTQ* interfaces is shown diagrammatically in Fig. 2.

Two permanent files are used by *MICE*. The first is the output file from the normalization module of *MITHRIL* and contains a complete set of U magnitudes and their variances, the unit-cell dimensions and contents plus symmetry information. The second contains the full phasing tree. For each node the basis-set reflexions and their associated Lagrangian multipliers are stored along with information concerning likelihood, Σ values, entropy *etc.* Such a file is very compact. The only time it is necessary to store a complete map *permanently* is when passing data from *MICE* to *PLOTQ*. Even large tree structures are thus very modest in their disk-space requirements. It is also possible when storing the whole tree to backtrack when an incorrect node selection has been made.

2.2. *Line search and plane search*

Two algorithms are provided for stabilizing the exponential modelling process:

(i) line search which optimizes the parameter t and maintains a pull-back (s) of zero (I, § 2.3.2.1) subject to the usual damping factors and bumpers;

(ii) plane search which introduces a second search direction and bi-cubic modelling of both the constraint and the entropy.

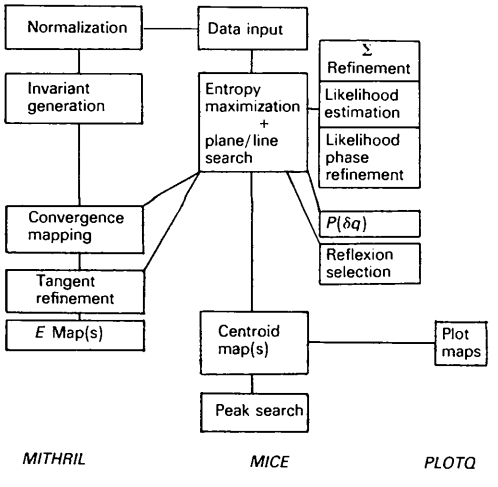


Fig. 2. Flowchart of the *MICE* program and its *MITHRIL* and *PLOTQ* interface.

Line search is *ca* five times faster than plane search but lacks the stability of the latter. However, the full power of the plane-search method is only needed when approaching $\chi^2 = 1{\cdot}0$ at a point where the dynamic range of $\omega(\mathbf{x})$ is growing large. *MICE* offers an option of mixing both methods by using line search while $\chi^2 > 1{\cdot}5\chi^2_{\text{target}}$ and then subsequently utilizing the plane-search algorithm. This is a recent development and all the calculations described in this paper use the plane-search method exclusively.

There is one final practical point to discuss. A basis set of reflexions is refined to convergence (usually $\chi^2 = 1{\cdot}0$). The $|U_\mathbf{h}|^{\text{obs}}$ are then well fitted to $|U_\mathbf{h}^{\text{ME}}|$ *via* the Lagrangian multipliers $\zeta_\mathbf{h}$. When new reflexions are added to this set, initially they have $\zeta_\mathbf{h} = 0{\cdot}0$, and so a situation arises in which new U's have to be fitted from zero in the presence of well fitted U values. This is not a favourable situation, and to overcome any problems that could arise, all the non-zero $\zeta_\mathbf{h}$ are multiplied by 0·5, so as to reduce the fitting of $|U_\mathbf{h}|^{\text{obs}}$ and $|U_\mathbf{h}^{\text{ME}}|$ before exponential modelling commences.

3. Phase extension in sucrose octaacetate

As a test of the plane-search algorithm, and of the quality of phase extrapolation in a maximum-entropy environment, phase extension was carried out on sucrose octaacetate. The top 300 U magnitudes were given correct phases and used as input to *MICE*. Experimental data were used throughout. This is a very severe test of the entropy maximization algorithm; the $q^{\text{ME}}(\mathbf{x})$ map produced using coefficients that attempt to fit U magnitudes is very sharp with a dynamic range exceeding e^{10}, and this can pose massive stability problems. Indeed, most previous reports using the ME method have used $|F_\mathbf{h}|$ on an absolute scale (*i.e.* unsharpened) to alleviate these problems. With the plane-search algorithm described in I with suitable damping parameters, there is no difficulty in the *MICE* program. The maximum s and t shifts (I, § 2.3.2.3) were both set to 0·1 and a temperature factor of 7·5 Å^2 was applied to the shifts to allow for the mixture of low- and high-resolution terms. This forces a better fit initially of the low-resolution terms and prevents an excessively fast build up of atomic detail. Because of this damping, more cycles of refinement are needed than would be the case when dealing with a smaller basis set. Refinement was followed to $\chi^2 = 1{\cdot}0$ where

$$\chi^2 = 1/(2n_a + n_c) \sum_\mathbf{h} (U_\mathbf{h}^{\text{obs}} - U_\mathbf{h}^{\text{ME}})^2/\sigma_\mathbf{h}^2,$$

where n_a and n_c are the number of acentric and centric reflexions, respectively. Entropy maximization was then allowed to continue while keeping $\chi^2 = 1{\cdot}0$, until the change in entropy, ΔS, was less than 0·01%.

Table 1 summarizes the results. For the first 15 cycles, the build up of detail is slow enough that the

Table 1. *Entropy maximization for sucrose octaacetate using the top 300 U magnitudes with correct phase angles in the basis set*

Cycle no.	χ^2	cos ($\nabla S, \nabla C$)	L	Entropy
1	4·10	0·9719	$-0{\cdot}195\times10^{-2}$	$-0{\cdot}305\times10^{-2}$
5	3·60	0·8520	$0{\cdot}253\times10^{0}$	$-0{\cdot}183\times10^{0}$
10	2·70	0·7893	$0{\cdot}546\times10^{2}$	$-0{\cdot}133\times10^{1}$
15	1·51	0·6237	$0{\cdot}743\times10^{3}$	$-0{\cdot}944\times10^{1}$
20	1·11	0·7297	$0{\cdot}1589\times10^{4}$	$-0{\cdot}310\times10^{2}$
25	1·00	0·7819	$0{\cdot}1845\times10^{4}$	$-0{\cdot}393\times10^{2}$
30	1·00	0·8178	$0{\cdot}1919\times10^{4}$	$-0{\cdot}400\times10^{2}$
35	1·00	0·8391	$0{\cdot}1954\times10^{4}$	$-0{\cdot}403\times10^{2}$

Table 2. *The U-weighted $\langle|\Delta\varphi|\rangle$ as a function of $|U_{\mathbf{h}}|^{\text{obs}}|U_{\mathbf{h}}^{\text{ME}}|$ for the extrapolated phases in sucrose octaacetate*

$\|U_{\mathbf{h}}^{\text{obs}}\|\|U_{\mathbf{h}}^{\text{ME}}\|$ limits	U-weighted $\langle|\Delta\varphi|\rangle$	Number of reflexions
0·001–0·0019	17	10
0·002–0·0039	14	109
0·004–0·0059	11	199
0·006–0·0079	9	135
0·008–0·0099	7	34
>0·0099	4	8

pull-back parameter s is always calculated as zero, but in subsequent cycles it plays a vital role in controlling the process. The dynamic range of the final map is enormous: $\omega(\mathbf{x})$ has maximum and minimum values of 9·14 and −6·51, respectively; this generates a dynamic range of $6{\cdot}85\times10^{2}$ to $1{\cdot}11\times10^{-4}$ in $q^{\text{ME}}(\mathbf{x})$. It is a clear vindication of this algorithm that such a huge range can be readily controlled, although at the expense of computer time, and some loss of alignment between the gradients ΔS and ΔC of the entropy and the constraint. It is possible, however, to force alignment by continuing refinement at $\chi^2 = 1{\cdot}0$ beyond the convergence point of $\Delta S = 0{\cdot}01\%$. Table 1 shows that there is a slow increase in cos ($\Delta S, \Delta C$) towards unity when maximizing the entropy at $\chi^2 = 1{\cdot}0$, but it is not worth the computer time that it would require to achieve a cosine value of 1·0, since there is no significant improvement in map quality nor in the quality of phase extrapolation.

Likelihood shows the expected increase from cycle to cycle as detail is built up in $q^{\text{ME}}(\mathbf{x})$, and the strength of extrapolation of U magnitudes increases. Entropy, too, shows the expected fall associated with the increase in atomic detail from cycle to cycle before stabilizing at −403.

By cycle 30, 494 non-basis-set reflexions are extrapolated with a U-weighted mean absolute phase error $\langle|\Delta\varphi|\rangle$ of 11°. (The examination of this restricted set of reflexions is merely a limitation of the current program and has no intrinsic significance.) The variation of $\langle|\Delta\varphi|\rangle$ as a function of $|U^{\text{obs}}||U^{\text{ME}}|$ is shown in Table 2. It can be seen that this conforms to the distribution developed in I (§ 1.2.4). The final values of Σ_a and Σ_c are $1{\cdot}05\times10^{-4}$ and $6{\cdot}85\times10^{-4}$, respectively, at the end of refinement. This represents an effective N of 3920; this is just the behaviour discussed in I (§ 3.3) for a situation where a great deal of correct phase information has accumulated. The tangent formula (Karle & Karle, 1966) gives a mean phase error of 12° under these conditions, but it is worth pointing out that tangent refinement uses the values of *all moduli* outside the 300 reflexions of the basis set, while the procedure used here *does not.*

In these calculations, each basis-set reflexion was given a weight of $1/(p\Sigma + \sigma_{\mathbf{h}}^2)$ in the maximization procedure where p is normally set to unity. It is, however, possible in the *MICE* program to set $p = 0{\cdot}0$ for the purposes of weighting. When this was done for sucrose octaacetate, the mean absolute phase error of the extrapolated reflexions fell to 9°. This is a viable option when a great deal of correct phase information is available for the basis-set reflexions, but not recommended for other situations since it will tend to drive log-likelihood gains negative because of the overfitting of U^{obs} and U^{ME}.

4. Phasing of diamantan-4-ol

4.1. *The ab initio phasing of diamantan-4-ol*

Despite claims to the contrary (Gull, Livesey & Sivia, 1987), centrosymmetric structures are more easily solved than non-centrosymmetric ones using both traditional and ME methods. Diamantan-4-ol is a typical small organic molecule, but it also possesses a strongly clustered set of six six-membered rings (Fig. 1*a*), and there is also a preference for phase reflexions having $l = 0 (\text{mod } 3)$ in the early stages of direct methods since such reflexions form a strong well connected phase island. Because this structure is centrosymmetric, relatively easy to solve and because of a lack of suitable reflexions, the phasing was not carried out using resolution ranges. This section needs to be read in conjunction with Table 3.

The intensity data were normalized as described in § 1. The *MITHRIL* program was run as far as the *CONVERGE* module, and the origin thus selected used to generate node 1 *via* reflexions $2\bar{1}6$ and 973 with U magnitudes of 0·275 and 0·273, respectively.

Node 1: There was some extrapolation taking place even with only two reflexions in the basis set. Indeed, seven reflexions having $|U^{\text{obs}}||U^{\text{ME}}|/2\Sigma_c > 1{\cdot}0$ were correctly extrapolated. However, none of them were incorporated into the basis set. (See § 3.2.) Instead, reflexions were then selected for phase permutation using the following criteria:

(i) the associated U magnitudes must be as large as possible;

(ii) the reflexions must interact with the basis set *via* triplets, quartets *etc.* to make a maximum number of new phases reliably accessible, *i.e.* the reflexions must optimally enlarge the second neighbourhood of the basis-set reflexions (Hauptman, 1980);

Table 3. *The phasing tree for diamantan-4-ol*

Node	From	Number of reflexions	Number wrongly phased	S	L	$\Sigma_c \times 10^3$
1	-	2	0	−1·40	0·13	8·04
2	1	7	2	−0·92	0·55	7·57
3	1	7	1	−0·92	0·50	7·59
4	1	7	1	−0·92	0·50	7·59
5	1	7	0	−0·92	0·55	7·57
6	2	10	2	−2·09	1·48	9·23
7	2	10	3	−2·01	1·23	9·21
8	2	10	1	−2·20	3·16	9·10
9	2	10	2	−2·18	2·44	9·11
10	2	10	3	−2·07	2·20	9·05
11	2	10	4	−2·03	2·45	9·15
12	2	10	2	−2·11	2·09	9·17
13	2	10	3	−2·19	2·22	9·21
14	5	10	1	−2·09	1·45	9·23
15	5	10	2	−2·01	1·22	9·22
16	5	10	0	−2·13	3·14	9·10
17	5	10	1	−2·17	2·41	9·13
18	5	10	2	−2·08	2·23	9·04
19	5	10	3	−2·05	2·47	9·14
20	5	10	3	−2·13	2·12	9·15
21	5	10	2	−2·20	2·24	9·20
22	8	12	3	−2·77	5·63	7·02
23	8	12	2	−3·07	1·51	7·38
24	8	12	2	−5·42	4·42	7·11
25	8	12	1	−5·07	4·77	7·69
26	16	12	2	−5·18	4·13	7·22
27	16	12	1	−5·45	4·38	6·99
28	16	12	1	−3·05	1·49	7·39
29	16	12	0	−2·77	5·62	7·02
30	22	14	2	−3·71	13·57	7·52
31	22	14	3	−3·71	3·73	8·16
32	22	14	4	−3·77	2·61	8·14
33	22	14	3	−3·76	5·78	7·83
34	29	14	0	−3·76	12·82	7·53
35	29	14	1	−3·76	3·87	8·17
36	29	14	1	−3·76	2·84	8·13
37	29	14	2	−3·76	6·14	7·87
38	34	100	3	−3·37	144·20	3·08
39	30	100	42	−8·30	−106·84	3·02
40	38	186	6	−10·41	200·07	2·71

(iii) the reflexions should be inaccessible or only weakly accessible to current extrapolation;

(iv) all other effects being equal, reflexions should be chosen at the lowest possible resolution;

(v) it is necessary to minimize the number of new reflexions in order to prevent the tree from becoming too extensive.

A module in *MICE* provides the necessary facilities with a command *NEXT* which explores the second neighbourhood of the basis-set reflexions and seeks its optimal, resolution and U-magnitude driven extension. For this structure, it was also necessary to force the inclusion of reflexions having $l \neq 0 (\mathrm{mod}\ 3)$. In this way, two reflexions were then selected for phase permutation, 107 ($|U| = 0{\cdot}234$) and 854 ($|U| = 0{\cdot}184$). Their phase permutation gave nodes 2–5.

Nodes 2–5: All four nodes had identical entropy values of −0·92. The likelihoods, however, were 0·55, 0·50, 0·50 and 0·55, respectively. For a centrosymmetric structure of this complexity and with such a small basis set, experience has shown that these differences may be considered significant. Accordingly, nodes 2 and 5 were selected. Using the same criteria as before, reflexions 18, 0, 0, 18, 2, 0 and 530 having U magnitudes of 0·411, 0·391 and 0·259, respectively, were permuted for both nodes. This gave rise to 16 nodes numbered 6–21 of which nodes 6–13 were connected to node 2 and the remainder to node 5.

Nodes 6–21: At this point there are only seven basis-set reflexions, but the power of the likelihood function, even in its diagonal approximation, is considerable. All the nodes have approximately the same entropy but the likelihoods vary from 1·22 to 3·16. Two nodes, 8 and 16, have likelihoods significantly higher than any others at 3·16 and 3·14. (Node 16 is indeed the correct one.) No other figure of merit in direct methods has this ability to discriminate correctly knowing only the phases of so few reflexions. Indeed, earlier experimentation in direct methods with early figures of merit used in tangent refinement showed all figures of merit to be unreliable at this level of phase information. Reflections 992 ($|U| = 0{\cdot}267$) and 10, 5, 4 ($|U = 0{\cdot}253$) were then permuted for both these nodes to give nodes 22–25 connected to node 8 and 26–29 connected to 16.

Nodes 22–29: Node 22 has a likelihood of 5·63 and that of node 29 is 5·62. These two nodes also have much higher entropies than any other in this set. Nodes 24, 25, 26 and 27 also have high likelihood values in the range 4·13–4·77, but their entropies vary between −5·07 and −5·45 which is much lower than that of −2·77 for nodes 22 and 29. Note also that the mean value of Σ_c has fallen to 7·23 for this set of nodes from a mean value of 9·15 for the previous level. Indeed, the correct node has one of the smallest values of this parameter although the value of Σ_c is not a reliable indicator of the strength of an individual node depending as it does on the nature of the crystal structure, and the data resolution (I, § 3.2.4). Nodes 16 and 22 were then passed to the next level by the permutation of reflexions $8\bar{4}0$ and $9\bar{8}3$ with U magnitudes of 0·265 and 0·260, respectively. This gave nodes 30–33 connected to node 22 and nodes 34–37 connected to 29.

Nodes 30–37: There are now 11 reflexions in the basis set and the power of the likelihood is enormous. Just two nodes, 30 and 34, with L values of 13·57 and 12·82 are kept. Note their relatively small values of Σ_c. At this point the extrapolation is very strong and almost error free for node 34. Indeed, inspection of a centroid map revealed the whole structure in the top 20 peaks of the map (based on only 11 basis-set reflexions!). However, the phasing process was continued by incorporating a total of 89 extrapolated reflexions having $|U|^{\rm obs}|U^{\rm ME}|/2\Sigma_c > 0{\cdot}3$ into the basis set for both nodes 30 and 34. The incorporated reflexions were given weights $w_{\rm h}$ derived from their centroids such that

$$w_{\rm h} = \tanh\left(|U_{\rm h}|^{\rm obs}|U_{\rm h}^{\rm ME}|/2\Sigma_c\right).$$

This gave nodes 38 and 39. For the former, 42 phases in the basis set are incorrect, whilst for the latter there are only 3.

Nodes 38 *and* 39: These nodes involved a two-stage process of entropy maximization in which the extrapolated reflexions were:

(i) incorporated into the basis set with centroid weights $w_{\rm h}$ from (1) above and subjected to entropy maximization;

(ii) re-weighted according to the new values of $|U^{\rm ME}|$ produced by (i) and the entropy optimization repeated.

For node 38 the result was a collapse of the likelihood to −106·84 and a massive entropy decrease to −8·30, whereas node 39 refined to $L = 144{\cdot}20$ and an entropy value of −3·37. It is noteworthy that the three wrongly phased reflexions for this node had their centroid weights *reduced* during the re-weighting procedure (ii) above. The value of Σ_c fell to $3{\cdot}02 \times 10^{-3}$ which, as for sucrose octaacetate, is an indication of correct phasing. Node 38 was discarded; for node 39 an extra 86 extrapolated reflexions were accepted to give node 40.

Table 4. *Peak heights in the final centroid map for diamantan-4-ol*

Peak no.	Peak assignment	Peak height
1	O	102
2	C	92
3	C	89
4	C	87
⋮	⋮	⋮
15	C	59
16	Noise	34
17	Noise	15

Node 40: Of the 186 basis-set reflexions six were incorrectly phased. However, the final solution after the standard re-weighting calculation gave a likelihood of 200·07 and an entropy of −10·41 with $\Sigma_c = 2{\cdot}71 \times 10^{-3}$. The quality of the final centroid map was excellent with the O atom distinguished from the C atoms and with no trace of a peak at the ring centroid. Table 4 lists the top 17 peaks in the map and their interpretation. The best E map from *MITHRIL* for diamantan-4-ol was also very good but it did not distinguish atom type.

4.2. *Accepting extrapolated reflexions - a caution*

Extrapolation is very strong even at early stages of the phasing process for diamantan-4-ol. It is, however, essential to refrain from incorporating any extrapolated reflexions into the basis set until *all* strong $U_{\rm h}$ are accessible to the extrapolation procedure. As an example, consider node 34. There were ten reflexions having $|U^{\rm obs}||U^{\rm ME}|/2\Sigma_c > 0{\cdot}7$ correctly extrapolated. The centroid map showed the outline of the molecule, but it was dominated by a large peak at the centroid of the system of six six-membered rings. Acceptance of these ten extrapolated phases into the basis set, even though they are correct, and performance of entropy maximization reinforced this central peak and greatly weakened the extrapolation. The likelihood function proved a very sensitive indicator of this by giving a value of $L = -75{\cdot}63$ whereas the entropy remained uncritical at −3·28. This is not an isolated example: in one case it was possible to produce a set of reflexions for diamantan-4-ol in which there are 150 basis-set reflexions with only 19 incorrectly phased that produced a map with the strong central peak and only small ghost atoms at the atomic sites around it. The ME method will, in general, always behave in this fashion until most of the large U magnitudes are accessible *via* extrapolation, at which point the centroid map usually shows the crystal structure anyway so that the incorporation of extrapolates into the basis set is not needed. This problem was probably the reason that Gull, Livesey & Sivia (1987) were only able to locate the Sn and Cl atoms in $C_{10}H_{14}Cl_7O_4Sn$ (Miller & Schlemper, 1978) despite phasing 159 basis-set reflexions.

There is, however, one situation where augmentation of the basis set with extrapolates can be valuable. If two or more equivalent nodes have very similar likelihoods and entropies such that they are difficult to distinguish, it is possible to collect some extrapolated reflexions and carry out the two-stage process of entropy maximization coupled with centroid weighting as described above. In general, all such nodes will acquire negative likelihoods but they should be distinguishable on the grounds of their new likelihoods and entropies as long as the same reflexions were used to augment the basis set in all cases, so that like is being compared with like. These augmented nodes are then discarded so that they are used only in a passive way. In fact, this procedure was used to distinguish nodes 38 and 39. This is possibly an important technique in solving more complex structures.

4.3. *The use of computer graphics*

It is, of course, also possible to monitor the phasing process *via* the inspection of centroid maps. Fig. 3 shows the development of the structure in real space in x sections centred around $z = 0{\cdot}2$. The first of these (Fig. 3a) is based on only two reflexions, but it already contains some correct structure; Fig. 3(b) is the equivalent Fourier section for node 18 and is based on a basis set of ten reflexions. There are some spurious peaks but most represent atom sites, and Fig. 3(c) shows the final map. The peaks are very sharp and locate the atoms very precisely. Again the stability of the exponential modelling-plane-search algorithm in dealing with such maps is demonstrated. A problem does arise with aliasing errors in these circumstances necessitating very finely sampled maps, but this is only a problem of computer resour-

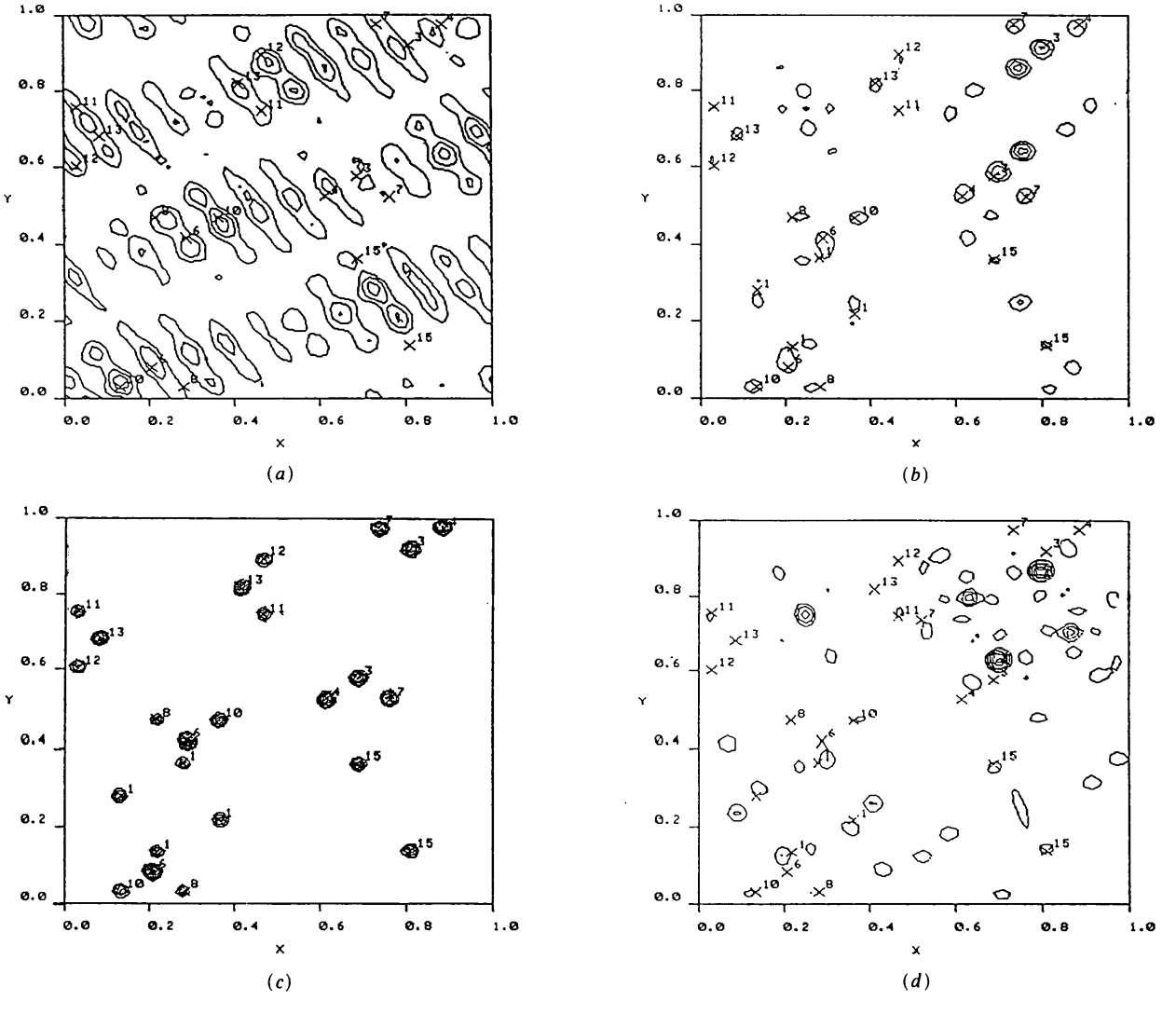


Fig. 3. Sections of centroid maps centred on $z = 0{\cdot}2$ for (a) node 1, (b) node 16, (c) node 40 and (d) node 20 (an incorrect node) in diamantan-4-ol.

ces. In contrast to Figs. 3(*a*)-(*c*), Fig. 3(*d*) represents node 20 which has ten basis-set reflexions of which three are incorrect; there is an excessive build up of large peaks at the centres of rings making no chemical sense. In difficult cases, this monitoring of centroid maps is an invaluable tool that is unavailable to traditional direct methods.

5. Phasing of (−)-platynecine

5.1. *The use of* $P(\delta q)$

This section introduces the use of $P(\delta q)$ (I, § 1.2.1) so a preliminary discussion concerning its use is necessary. $P(\delta q)$ can be used with any $q^{ME}(\mathbf{x})$. A set of reflexions is chosen according to certain criteria discussed below, and their phases permuted (using magic integers for the acentric reflexions). Each permutation gives rise to a $\delta q(\mathbf{x})$ which is a Fourier synthesis using the coefficients $U_{\mathbf{h}}^{obs}-U_{\mathbf{h}}^{ME}$. $P(\delta q)$ is calculated by computing the integral of $\delta q(\mathbf{x})^2/q^{ME}(\mathbf{x})$. A *minimum* value of $P(\delta q)$ is required. The *EXTEND* module in *MICE* performs the necessary calculations. This technique has several important features:

(i) Each phase permutation requires only one Fourier synthesis and a map division and so is very fast. (It is possible to devise a much faster reciprocal-space algorithm but this has not yet been programmed.)

(ii) It acts as a filter (the P filter) to exponential modelling; the reflexions to be permuted are first subjected to the P filter and only those with a certain minimum P are passed to the much slower exponential modelling step.

(iii) As discussed in I (§ 1.2.1), the likelihood function uses only moduli, but $P(\delta q)$ incorporates phases into the calculation. It therefore acts as a useful tool in exploring structure-factor space from the current node before moving to various P minima.

$P(\delta q)$ cannot be used indiscriminately; there are certain necessary conditions for its successful implementation:

(i) $q^{ME}(\mathbf{x})$ must have developed sufficient detail. Thus maps based on a very small basis set or utilizing only small U magnitudes may have insufficient contrast for its successful use. This is made manifest by a set of $P(\delta q)$ values that are virtually constant. In addition, division by $q^{ME}(\mathbf{x})$ is inherently unstable and the same constraints in division must be employed that are used in exponential modelling. In general, the method is not sensitive to limits in $q^{ME}(\mathbf{x})$ although these can be set by the user.

(ii) The reflexions which act as coefficients for $\delta q(\mathbf{x})$ should be chosen using the same criteria of optimum second-neighbourhood extension discussed in § 4.1. It is also advantageous to choose reflexions for which there is a small but finite extrapolated magnitude from the current $q^{ME}(\mathbf{x})$.

Table 5. *The use of* $\int \delta q^2(\mathbf{x})/q^{ME}(\mathbf{x})\,d^3\mathbf{x}$ *as a filter in the phasing of* (−)-*platynecine*

Set no.	$\int \delta q^2(\mathbf{x})/q^{ME}(\mathbf{x})\,d^3\mathbf{x}$	U-weighted $\langle\lvert\delta\varphi\rvert\rangle$
(*a*) Four centric phases permuted		
1	1·79	132
2	1·66	180
3	1·83	85
4	1·72	133
5	2·07	88
6	1·87	136
7	1·48	41
8	1·30	89
9	1·48	91
11	1·51	44
12	1·41	92
13	1·76	47
14	1·57	95
15	1·16	0
16	1·00	48
(*b*) Three centric phases permuted		
1	2·40	180
2	3·37	107
3	2·38	116
4	3·32	43
5	3·33	137
6	2·38	64
7	3·37	73
8	2·40	0

(iii) Table 5(*a*) shows a typical successful use of $P(\delta q)$. It is taken from node 21 in the phasing of (−)-platynecine. Two sets (15 and 16) define a clear minimum with $P=0{\cdot}99$ and 1·16, respectively. The remaining sets have P values between 1·30 and 2·07. The correct set has $P=1{\cdot}16$. Table 5(*b*) shows another successful use of P filtering. Again, the results come from (−)-platynecine but were not used subsequently. In this case two sets of results are produced: one has P around 2·4, the other 3·3. In this case the four sets at $P\simeq 2{\cdot}4$ are passed to exponential modelling and half the sets are rejected.

(iv) Situations where there is no clear minimum or where all the $P(\delta q)$ values are constant are clear indications that either $q^{ME}(\mathbf{x})$ has insufficient contrast or that unsuitable reflexions have been chosen as coefficients for $\delta q(\mathbf{x})$.

5.2. *The ab initio phasing of* (−)-*platynecine*

This section must be read in conjunction with Table 6. The intensity data were processed as described in § 1.

Node 1: Unlike diamantan-4-ol, the distribution of intensities as a function of $(\sin^2\theta)/\lambda^2$ permits the use of resolution shells to phase (−)-platynecine. At a resolution of 2·24 Å, four reflexions (310, 012, 303 and 231) having U magnitudes of 0·282, 0·243, 0·177 and 0·167, respectively, were used to define the origin and the enantiomorph. These reflexions were given their true phases to facilitate comparison between observed and calculated phases and to save on the initial phase permutation of 231. For entropy

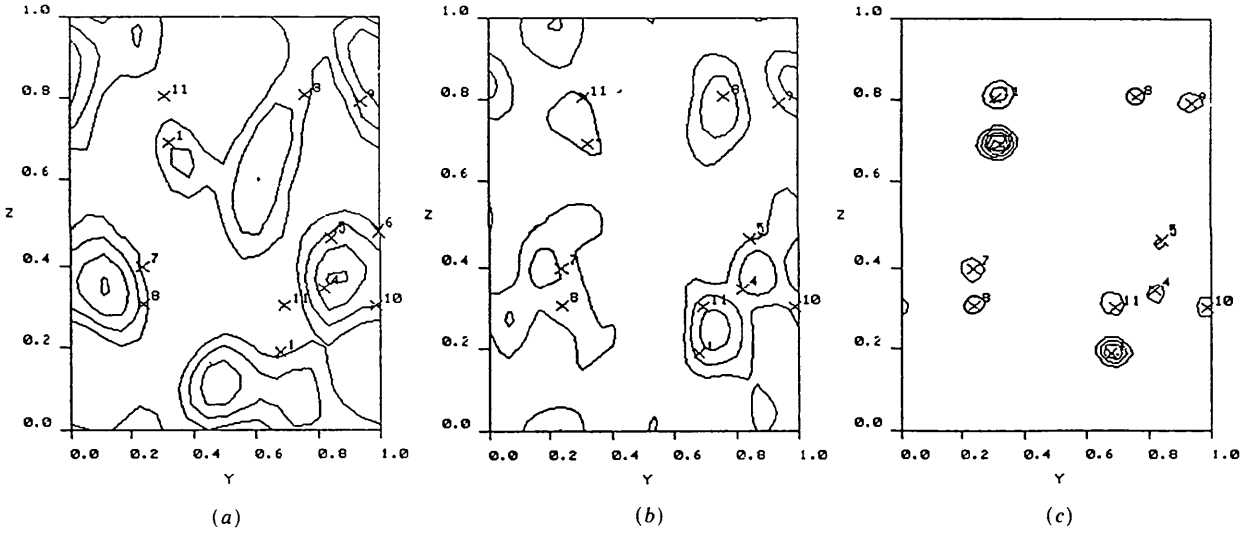
Table 6. *The phasing tree for (−)-platynecine*

Node	From	$\langle\|\Delta\varphi\|\rangle$	R.m.s. $\langle\|\Delta\varphi\|\rangle$	Number of reflexions	S	L	$\Sigma_a \times 10^3$	$\Sigma_c \times 10^3$	Resolution (Å)
1	-	0·0	0·0	4	−0·17	0·01	7·51	1·83	2·24
2	1	50·0	90·0	8	−0·95	−0·35	5·17	8·60	2·24
3	1	21·8	63·7	8	−1·34	−1·62	5·25	8·40	2·24
4	1	28·7	63·7	8	−0·96	−0·29	5·18	8·65	2·24
5	1	52·1	90·0	8	−1·35	−1·57	5·27	8·42	2·24
6	1	21·8	63·7	8	−0·96	0·05	5·08	8·10	2·24
7	1	45·1	90·0	8	−1·37	−0·29	5·11	7·87	2·24
8	1	0·5	1·8	8	−0·95	0·06	5·06	8·10	2·25
9	1	23·8	63·7	8	−1·37	−0·28	5·10	7·89	2·24
10	1	72·9	110·2	8	−0·98	−0·22	5·07	6·88	2·24
11	1	96·3	127·3	8	−1·53	0·07	4·97	7·14	2·24
12	1	51·6	90·0	8	−0·98	−0·25	5·12	6·97	2·24
13	1	75·0	110·2	8	−1·53	0·06	5·01	7·25	2·24
14	1	44·7	90·0	8	−0·95	−0·40	5·20	8·26	2·24
15	1	68·0	110·2	8	−1·45	−0·67	5·16	8·57	2·24
16	1	23·3	63·7	8	−0·95	−0·32	5·18	8·25	2·24
17	1	46·7	90·0	8	−1·45	−0·58	5·12	8·58	2·24
18	8	4·7	15·8	11	−1·18	0·29	4·07	8·28	1·50
19	6	21·4	56·5	11	−1·44	0·70	3·90	8·20	1·50
20	18	11·3	23·4	16	−2·01	1·21	4·20	6·40	1·40
21	20	9·0	21·5	19	−2·15	2·82	7·61	2·82	1·40
22	21	7·4	19·5	23	−2·86	9·27	4·63	6·04	1·40
23	21	14·7	42·2	23	−2·45	6·60	4·86	6·30	1·40
24	23	14·7	42·2	27	−4·12	15·02	5·81	8·81	1·10
25	24	9·4	21·1	32	−5·98	36·73	6·38	11·12	0·78

maximization, this is a favourable starting point with large U magnitudes available at low resolution. A section of the current centroid map is shown in Fig. 4(*a*). It can be seen that most atoms lie on regions of relatively high density.

Nodes 2–17: With the resolution maintained at 2·24 Å, four reflexions (302, 202, 220 and 102) (U magnitudes of 0·242, 0·220, 0·237 and 0·239) were given permuted phases. Since all are centric reflexions in space group $P2_12_12_1$, this generated 16 nodes. After entropy maximization only four nodes had a likelihood >2·0, and, of these, two had much smaller entropies than the others. Thus two nodes 6 and 8, of which the latter is marginally preferred, were selected.

Nodes 6 *and* 8: The resolution was now extended to 2·0 Å for both nodes by permuting reflexions 132, 320 and 233 with U magnitudes of 0·203, 0·193 and 0·163, respectively. Magic integers were used to permute the two acentric reflexions so that 28 phase sets were produced. Nodes 6 and 8 both had a $q^{ME}(\mathbf{x})$ that contained sufficient contrast to permit the use of $P(\delta q)$ so that these nodes were first P filtered. For node 6 the best set had $P(\delta q) = 1{\cdot}12$ with the next having $P(\delta p) = 1{\cdot}17$. For node 8 these figures were 1·07 and 1·15, respectively. In both cases these



(*a*) (*b*) (*c*)

Fig. 4. Sections of centroid maps centred on $x = 0{\cdot}20$ for (*a*) node 1, (*b*) node 18 and (*c*) node 25 in (−)-platynecine.

minima are sufficiently well defined that only the best set was carried forward in each case to a full entropy maximization thus giving nodes 18 and 19.

Nodes 18 *and* 19: An interesting situation arises here in distinguishing the best node. Node 18 has $L=0{\cdot}29$ and $S=-1{\cdot}18$ whereas node 19 has $S=-1{\cdot}44$ and $L=0{\cdot}70$. Thus there is some conflict between the requirement of maximum entropy and maximum likelihood. In these circumstances $NS+L$ is used as a joint indicator, where N is derived from a weighted average of $1/\Sigma_c$ and $1/2\Sigma_a$ (I, § 3.2.4). Node 18 had $NS+L=-82$ whereas the value for node 19 was -104. Thus node 18 was selected; a section from the relevant centroid map is shown in Fig. 4(*b*). It can be seen that the correct atomic detail is emerging. The resolution was increased to 1·5 Å by permuting reflexions (422, 216, 151, 117 and 241) with U's of 0·238, 0·238, 0·206, 0·191 and 0·168, respectively. 48 phase sets were produced, and these were filtered *via* $P(\delta q)$. The best solution had $P(\delta q)=2{\cdot}9$ with the next-ranked solution having $P(\delta q)=3{\cdot}1$. Thus the first set only was passed to entropy maximization to give node 20.

Node 20: The resolution was increased to 1·4 Å by permuting three centric reflexions (203, 530 and 028) with U's of 0·201, 0·317 and 0·212, respectively. Using $P(\delta q)$ as a filter, the best set had a value of $P(\delta q)=1{\cdot}8$ with the rest having $P(\delta q)>2{\cdot}3$, so only one set was passed to entropy maximization to give node 21.

Node 21: The resolution was maintained at 1·4 Å and four further centric reflexions (430, 530, 052 and 051) were given permuted phases and filtered *via* $P(\delta q)$. There were two permutation sets with much lower values of $P(\delta p)$ than the rest having values of 0·99 and 1·16, respectively, whilst the third-ranked set had $P(\delta p)=1{\cdot}30$. The two best sets were used to generate nodes 22 and 23.

Nodes 22 *and* 23: Node 22 had a log-likelihood gain of 9·27 compared to 6·60 for node 23. However, the two entropy values were $-2{\cdot}86$ and $-2{\cdot}45$ so that the two indicators are contradictory. As usual in these circumstances, inspection of $NS+L$ is used. For nodes 22 and 23 the values are -128 and -109, so that node 23 is clearly preferred. Moreover, a centroid map at this juncture showed over half the structure for node 23 but no recognizable fragment for node 22. For both reasons, the latter was therefore discarded. The resolution was then extended to 1·1 Å by permuting 451, 171, 270 and 527 with U's of 0·283, 0·252, 0·251 and 0·244, respectively. The $P(\delta q)$ filter produced one obvious minimum, and this was used to generate node 24.

Node 24: This node was subjected to phase refinement by likelihood optimization. (Previous nodes had also been phase refined, but there had been no significant change in the phase angles.) The mean absolute phase error was reduced to 7·5° from a value of 8·3°. The r.m.s. value was reduced from 19·0 to 17·7°. However, of the 23 basis-set reflexions which were being refined, only 11 were acentric; the improvement for these reflexions only is 17·3 to 15·7° for the mean absolute phase error and 27·4 to 25·5° for the r.m.s. error. This is only a small improvement, but one reflexion (117) which had the largest phase error had its phase improved from -64 to $-84°$ (with a true value of $-90°$). The resolution was then extended to 0·78 Å by permuting the phases of 860, 911, 808, 707 and 761 with U's of 0·382, 0·321, 0·306, 0·301, 0·293. The best set had $P(\delta q)=10{\cdot}5$ with the next sets having $P=11{\cdot}5$, 11·6, 11·8, ... Thus only one set was passed to entropy maximization to give node 25.

Table 7. *Phase refinement on node* 25 *for* (−)-*platynecine*

There are 19 centric and 13 acentric reflexions. The figures in parentheses denote the $\langle|\Delta\varphi|\rangle$ and r.m.s. $\Delta\varphi$ values for the acentric reflections only. All phase angles are in degrees.

Cycle no.	$\langle\|\Delta\varphi\|\rangle$	R.m.s. $\Delta\varphi$	S	L
0	7·5 (18·4)	17·0 (26·7)	−5·98	36·73
1	6·2 (15·2)	13·8 (21·7)	−6·48	45·47
2	5·3 (13·0)	12·0 (18·9)	−6·23	48·80
3	5·0 (12·3)	11·5 (18·1)	−5·87	49·33

Table 8. *Peak heights in the final centroid map for* (−)-*platynecine*

Peak no.	Peak assignment	Peak height
1	O	143
2	O	115
3	N	68
4	C	52
5	C	50
⋮	⋮	⋮
11	C	23
12	Noise	18

Nodes 25: The centroid map showed the complete structure. The basis-set reflexions were given four cycles of likelihood phase refinement the results of which are summarized in Table 7. It can be seen that there is a very significant improvement in the phase angles which is paralleled by an increase in likelihood from 36 to 49. The resulting centroid map gave a very clean map; the required 11 atoms were the top 11 in the list with peak heights reflecting atomic type as shown in Table 8. This map is of better quality than the best E map from *MITHRIL*. The latter does distinguish the O atoms but not the N and has a spurious peak in the centre of one five-membered ring which is not present in this map. Fig. 4(*c*) shows a typical section of this map with the customary sharp well placed peaks.

There are 289 extrapolated phases at this point having $|U_h|>0{\cdot}14$ and having a mean absolute phase error of 28°. It is also possible to collect extrapolated reflexions, but even greater care is needed in this non-centrosymmetric situation because of phase

errors and their accumulation. The calculation is so sensitive to the number and mode of incorporation of extrapolates that it is not worth the effort since the entire structure is clearly visible in the final map.

6. Summary and concluding remarks

We have demonstrated the viability of the approach presented in paper I. It is clear that the maximum-entropy method is a powerful tool for phase determination and, when combined with likelihood, is capable of solving the crystal structures of small organic molecules. Likelihood is a unique discriminator of phase sets when used as a figure of merit in a multisolution environment, and also an equally powerful tool in phase refinement.

The algorithms we have devised are stable even when employing U magnitudes, and, since they are based on the FFT, are generally applicable to any symmetry and any size of problem. The processor time required is a simple function of map size. The ME method clearly requires much more computer time than traditional methods, but there are several features which can compensate:

(i) For small basis sets, where stability is much less of a problem, it is possible to construct much faster algorithms than used here; we have written a *general-purpose* constrained entropy-maximization program that works irrespective of the basis-set size as a test of the method.

(ii) The exponential modelling algorithm is readily vectorized, and can therefore be used with a variety of cheap vector and array processor boards that are currently available for scientific workstations as well as supercomputers. We are currently investigating this.

(iii) For difficult structures it is not unusual for a conventional direct-methods program to consume a very large amount of processor time. If the ME method can be extended to such structures the times could be strictly comparable.

There is an additional benefit that ME confers on the phasing process that has hitherto only been indirectly alluded to when discussing resolution shells: the ME method in its current implementation is stable regardless of data quality, sampling and resolution. Traditional direct methods require data that span at least the Cu sphere and that fully sample reciprocal space; indeed, the spectacular improvements in direct-methods success in the past two decades owe much to the wide availability of the current generation of four-circle diffractometers. However, it means that direct methods still cannot readily deal with:

(i) poor-quality data sets from weakly diffracting or very small crystals;

(ii) X-ray and neutron powder data where there is an effective loss of resolution because of peak overlap, even with synchrotron sources and neutron spallation sources;

(iii) fibre diffraction. There has been one successful application by Marvin, Bryan & Nave (1987) to improve a map initially phased by isomorphous replacement;

(iv) electron crystallography using a combination of phased transform data from the electron microscope image and higher-resolution unphased transform data.

In all these cases, the ME method offers a possible way forward. We are currently researching the application of the ME method to powder diffraction (Bricogne, unpublished results; Henderson & Gilmore, 1989) and electron crystallography. It seems too that there is an applicability of the method to much larger and more complex molecules, although a more complicated phasing tree will be needed as the discrimination between nodes will inevitably be more difficult.

This collaborative research was initiated at LURE thanks to a Senior Ciba Fellowship to CJG; subsequent work at Glasgow was supported by a grant from the SERC (to CJG). GB acknowledges support from CNRS (France) and partial support from Trinity College, Cambridge and the MRC Laboratory of Molecular Biology, Cambridge. Keith Henderson (Glasgow) was the source of the Σ-refinement code used here, and of many fruitful discussions.

References

BANNISTER, C., BRICOGNE, G. & GILMORE, C. J. (1989). In *Maximum Entropy and Bayesian Methods*, edited by J. SKILLING, pp. 225-232. Amsterdam: Kluwer.

BRICOGNE, G. & GILMORE, C. J. (1990). *Acta Cryst.* A**46**, 284-297.

FREER, A. A., KELLY, H. A. & ROBINS, D. J. (1987). *Acta Cryst.* C**43**, 2020-2022.

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

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

GULL, S. F., LIVESEY, A. K. & SIVIA, D. S. (1987). *Acta Cryst.* A**43**, 112-117.

HALL, S. R. & SUBRAMANIAN, V. (1982). *Acta Cryst.* A**32**, 598-600.

HAUPTMAN, H. A. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 151-197. New York: Plenum.

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.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849-859.

MARVIN, D. A., BRYAN, R. K. & NAVE, C. (1987). *J. Mol. Biol.* **193**, 315-343.

MILLER, G. A. & SCHLEMPER, E. O. (1978). *Inorg. Chim. Acta*, **30**, 131-134.

OLIVER, J. D. & STRICKLAND, L. C. (1984). *Acta Cryst.* C**40**, 820-824.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183-191.

WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151-152.