DWIGGINS, C. W. (1975). *Acta Cryst.* A31, 395-396.
EICHHORN, K. (1987). *REDUCE. Program for Data Reduction of Step-Scan Measured Reflection Profiles for both Neutrons and X-rays.* DESY/HASYLAB, Hamburg, Federal Republic of Germany. Unpublished.
HAFNER, S. & NAGEL, S. (1983). *Phys. Chem. Miner.* 9, 19-22.
HALLBERG, J. & HANSON, R. C. (1970). *Phys. Status Solidi*, 42, 305-310.
International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)
KOELLING, D. D. & ARBMAN, G. O. (1975). *J. Phys. F*, 5, 2041.
KRUEGER, H. & MEYER-BERKHOUT, K. (1952). *Z. Phys.* 132, 171-178.
KUPCIK, V., WENDSCHUH-JOSTIES, M., WOLF, A. & WULF, R. (1986). *Nucl. Instrum. Methods*, A246, 624-626.
KURKI-SUONIO, K. (1977). *Isr. J. Chem.* 16, 115-123.
LANDOLDT-BÖRNSTEIN (1966). Edited by K.-H. HELLWEGE. Neue Serie III-1, p. 13. Berlin: Springer Verlag.
LEHMANN, M. S. & LARSEN, F. K. (1974). *Acta Cryst.* A30, 305-310.
LEWIS, J., SCHWARZENBACH, D. & FLACK, H. D. (1982). *Acta Cryst.* A38, 733-739.
MARKSTEINER, P., BLAHA, P. & SCHWARZ, K. (1986). *Z. Phys.* B64, 119-127.
MERISALO, M. & KURITTU, J. (1978). *J. Appl. Cryst.* 11, 179-183.
MULLEN, D. & FISCHER, K. (1981). *Z. Kristallogr.* 156, 85-86.
NAGEL, S. (1985). *J. Chem. Phys. Solids*, 46, 743-756.
NGO THONG & SCHWARZENBACH, D. (1979). *Acta Cryst.* A35, 658-664.
NIELSEN, F. S., LEE, P. & COPPENS, P. (1986). *Acta Cryst.* B42, 359-364.
RESTORI, R. & SCHWARZENBACH, D. (1986). *Acta Cryst.* B42, 201-208.
SMEND, F., SCHAUPP, D., CZERWINSKI, H., MILLHOUSE, A. H. & SCHENK-STRAUSS, H. (1984). DESY Internal Report No. SR-84-003. DESY/HASYLAB, Hamburg, Federal Republic of Germany.
STEWART, R. F. (1976). *Acta Cryst.* A32, 565-574.
STEWART, R. F. (1979). *Chem. Phys. Lett.* 65, 335-342.
STEWART, R. F. & SPACKMAN, M. A. (1981). *VALRAY* Users Manual. Preliminary Draft. Department of Chemistry, Carnegie-Mellon Univ., Pittsburgh, PA 15213, USA.
TSIRELSON, V. G., ANTIPIN, M. YU., GERR, R. G., OZEROV, R. P. & STRUCHKOV, YU. T. (1985). *Phys. Status Solidi A*, 87, 425-433.
WAL, R. J. VAN DER, VOS, A. & KIRFEL, A. (1987). *Acta Cryst.* B43, 132-143.
WENDSCHUH-JOSTIES, M. & WULF, R. (1989). *J. Appl. Cryst.* 22, 382-383.

Acta Cryst. (1990). A46, 284-297

A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. I. Theory, Algorithms and Strategy

BY G. BRICOGNE

LURE, Université Paris-Sud, 91405 *Orsay, France*

AND C. J. GILMORE

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

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

Abstract

A new multisolution method for direct phase determination [Bricogne (1984). *Acta Cryst.* A40, 410-445] has been implemented and tested on small crystal structures. It consists of an organized search for those combinations of phases associated with a 'basis set' of reflexions which have *maximum likelihood, i.e.* which lead to the assignment of the highest conditional probability to the observed moduli belonging to reflexions outside the basis set. Phase choices are made sequentially, progressively enlarging the basis set, and the book-keeping involves a 'multisolution tree' which summarizes the parentage relations between them. The conditional probability distributions (c.p.d.'s) of structure factors used in evaluating the likelihood are derived from joint distributions obtained by the saddlepoint method. The latter involves distributions of atoms which have the *maximum entropy* compatible with all phase choices made, and hence are different for each node of the multisolution tree. These distributions q^{ME} are constructed numerically by exponentially modelling, coupled with a very robust plane search which often simplifies to a line search. C.p.d.'s of small numbers of structure factors not in the basis set are readily calculated from q^{ME}, with correct representation of their multimodality. A further 'diagonal' approximation of these c.p.d.'s allows the log-likelihood to be written as a sum of contributions from individual non-basis reflexions. The phasing process is initiated by specifying the origin-fixing and enantiomorph-defining phases, and forming the corresponding q^{ME}. It progresses by roughly locating the maxima of the c.p.d.'s of additional structure factors by a magic-integer technique, updating q^{ME} separately for each such maximum, and evaluating their respective likelihoods. The most likely phase sets are further refined

0108-7673/90/040284-14$03.00

by numerically maximizing their likelihood and are accepted as enlarged basis sets. This is a first approximate implementation of a general Bayesian theory of phase determination [Bricogne (1988). *Acta Cryst.* A**44**, 517-545]. A companion paper [Gilmore, Bricogne & Bannister (1990). *Acta Cryst.* A**46**, 297-308] describes successful applications to the *ab initio* phasing of small molecules, which demonstrate the viability of this new method and show that likelihood is far superior to any existing figure of merit in discriminating between correct and incorrect phases.

0. Introduction

This paper describes a first step towards the implementation of a Bayesian treatment of direct phase determination put forward in previous papers by one of us (Bricogne, 1984, 1988*a*, *b*; hereafter referred to as I, II and III, respectively). The work presented here aims at testing the most basic mechanism in this approach: the use of joint probability distributions of structure factors built recursively from maximum-entropy distributions of random atoms, driven by a tree-directed search for phase combinations which also maximize a likelihood criterion based on the values of the yet unphased moduli. This first paper gives a detailed account of the theoretical basis of the method, describes the level of approximation used and the algorithms developed to implement it, and outlines the principles according to which the process of phase determination is conducted.

The theoretical basis of the present work is summarized in § 1. As previously discussed, its main novelty resides in the methods used to calculate the various probability distributions (I, §§ 2, 5), in the definition and systematic use of a likelihood criterion (I, §§ 4.2.2, 8.1; II, § 0.6; III, §4), and in the idea that phases can be refined by optimization of this criterion (II, § 0.6; III, § 4). Explicit formulae are given for the forms of conditional probability distributions and of likelihood functions which are invoked in the sequel. Attention is drawn to those properties of the exact expressions which are retained or lost at various levels of approximation, as this is of great practical importance in their subsequent use. A detailed presentation of the algorithms developed for the maximization of entropy and likelihood and for other accessory calculations is given in § 2. The overall multisolution phasing strategy originally proposed in I, § 8.1, is then re-examined step by step and its practical implementation is described in terms of the previously established formulae and algorithms.

A companion paper presents applications of the method to small structures, discusses the results obtained, and concludes that the proposed approach has been validated.

1. Theory

1.1. *Joint distributions of structure factors*

Joint probability distributions (j.p.d.'s for short) of structure factors will be derived from the assumption that a crystal structure consists of a large number N of statistically independent equal atoms randomly distributed in the asymmetric unit with probability density $m(\mathbf{x})$. If $H=\{\mathbf{h}_1,\ldots,\mathbf{h}_n\}$ is a collection of unique non-origin reflexions, the corresponding unitary structure factors will be arranged as a vector

$$\mathbf{U}_H=\begin{pmatrix} U_{\mathbf{h}_1} \\ \vdots \\ U_{\mathbf{h}_n} \end{pmatrix}$$

and their joint distribution will be described by a probability density in structure-factor space which will be denoted $\mathcal{P}(\mathbf{U}_H)$. This is the usual starting point of direct methods, but the latter always make the further assumption that $m(\mathbf{x})$ is a uniform distribution.

In the present approach, the most radical departure from traditional direct methods is not in the definition but in the mode of calculation of j.p.d.'s. The rationale for such a departure has been analysed elsewhere (I, § 2; III, § 2) and the modifications suggested were shown to yield better and more informative conditional distributions (I, §§ 2.3.1, 8.2) which constitute the essential tool of statistical phase determination. For a given set of phased structure-factor values $\mathbf{U}_H^*\neq\mathbf{0}$ containing large moduli, numerical approximations to $\mathcal{P}(\mathbf{U})$ at or near $\mathbf{U}_H^*$ will be calculated by means of the saddlepoint method (I, § 5). The latter is equivalent (I, § 5.5) to applying the central limit theorem starting from a modified distribution of atoms $q^{\mathrm{ME}}(\mathbf{x})$ [instead of $m(\mathbf{x})$] which is characterized by the property that it is the unique distribution which has maximum entropy $\mathcal{S}_m(q)$ relative to $m(\mathbf{x})$, where

$$\mathcal{S}_m(q)=-\int_V q(\mathbf{x})\log[q(\mathbf{x})/m(\mathbf{x})]\,\mathrm{d}^3\mathbf{x}, \quad (1.1)$$

and has Fourier coefficients $\mathbf{U}_H^*$ for the collection of reflexions H.

1.2. *Conditional distributions of structure factors*

Let K be another collection of unique non-origin reflexions, disjoint from H, and let $\mathbf{U}_K$ denote the associated vector of structure-factor values. The conditional probability distribution (c.p.d. for short) of $\mathbf{U}_K$, given that $\mathbf{U}_H$ has the value $\mathbf{U}_H^*$, is defined as

$$\mathcal{P}(\mathbf{U}_K\,|\,\mathbf{U}_H=\mathbf{U}_H^*)=\mathcal{P}(\mathbf{U}_H^*,\mathbf{U}_K)/\mathcal{P}(\mathbf{U}_H^*). \quad (1.2)$$

1.2.1. *Gaussian approximation.* It was shown in I, § 4.2, that the maximum-entropy distribution of atoms q^{ME} constructed to evaluate $\mathcal{P}(\mathbf{U}_H^*)$ provides a means

of approximating this conditional distribution by a multivariate Gaussian with centre $\mathbf{U}_K^{\mathrm{ME}}$ and covariance matrix $\mathbf{Q}_{KK}$:

$$\mathcal{P}(\mathbf{U}_K \,|\, \mathbf{U}_H = \mathbf{U}_H^*) \propto \exp\{-\tfrac{1}{2}(\mathbf{U}_K - \mathbf{U}_K^{\mathrm{ME}})^T \mathbf{Q}_{KK}^{-1}(\mathbf{U}_K - \mathbf{U}_K^{\mathrm{ME}})\}. \quad (1.3)$$

Here, $\mathbf{U}_K^{\mathrm{ME}}$ is the vector of Fourier coefficients of q^{ME} which have been extrapolated from the data $\mathbf{U}_H^*$ by the process of entropy maximization [I, §§ 3.4.2(2), 4.2]; the covariance matrix $\mathbf{Q}_{KK}$ can be calculated exactly from the spectrum of q^{ME} *via* structure-factor algebra [see I, § 4.2.1, equation (4.6); II, § 0.6 and Appendix], and is essentially the Toeplitz matrix associated with q^{ME} (I, § 4.2.1). This approximation to the conditional distribution will be denoted $\mathcal{P}^{\mathrm{SP}}(\mathbf{U}_K \,|\, \mathbf{U}_H = \mathbf{U}_H^*)$.

The quadratic form in the exponent of the above formula may be rewritten (I, § 4.2.1) using Parseval's theorem, to give another approximation in terms of maps in real space:

$$\mathcal{P}^{\mathrm{SP}}(\mathbf{U}_K \,|\, \mathbf{U}_H = \mathbf{U}_H^*) \propto \exp\{-\tfrac{1}{2}N \int_V \{[\delta q(\mathbf{x})]^2 / q^{\mathrm{ME}}(\mathbf{x})\}\, \mathrm{d}^3\mathbf{x}\}, \quad (1.4)$$

where $\delta q(\mathbf{x})$ is the Fourier synthesis calculated from the coefficients $\mathbf{U}_K - \mathbf{U}_K^{\mathrm{ME}}$ (with symmetry expansion according to the space group of the structure under study). This new expression has the following very simple interpretation in real space: the role of the non-uniform distribution q^{ME} is to create a *differential cost* for the addition of new features at different locations in the unit cell, making it less costly (in terms of loss of conditional probability) to introduce new detail where q^{ME} is large than where it is small.

1.2.2. *Multimodality as a function of the phases.* An important characteristic of the conditional distributions above and of their approximate expressions is that, although they are *unimodal* when their arguments $\mathbf{U}_K$ are unrestricted, they are *multimodal* functions of the phases $\{\varphi_{\mathbf{k}}\}_{\mathbf{k}\in K}$ when the moduli $|\mathbf{U}_K|$ are given fixed values. This is a consequence of the fact that, although $\mathcal{P}^{\mathrm{SP}}$ is a convex function, the locus defined by the fixed moduli (a multidimensional torus) is non-convex.

The real-space picture of c.p.d.'s in terms of differential cost provides an alternative illustration of multimodality: the contribution δq of reflexions in K will correspond to a local maximum of $\mathcal{P}^{\mathrm{SP}}$ if its peaks can be made to cluster near the maxima of q^{ME}; but this can in general be achieved for many choices of the phases $\varphi_{\mathbf{k}}$. This picture makes obvious the following property: if q^{ME} contains only low-resolution modulations (hence few local maxima), the multimodality of the conditional distributions built from it will be less severe than if it contains high-resolution modulations (hence numerous local maxima). This remark has great practical importance (§ 3.2.0).

1.2.3. *Comparison with the Wilson distribution.* To compare $\mathcal{P}^{\mathrm{SP}}$ with the usual distribution for $\mathbf{U}_K$ (Wilson, 1949, 1950), it is useful to recall that, for unitary structure factors in the equi-atom case, the quantity Σ (the sum of the squared form factors over the contents of the unit cell) becomes $(1/N)$ [see I, § 4.2.2(1)]. The distribution $\mathcal{P}^{\mathrm{SP}}$ is then seen to differ from the Wilson distribution in two respects:

(i) it is centred around $\mathbf{U}_K = \mathbf{U}_K^{\mathrm{ME}}$, not $\mathbf{U}_K = \mathbf{0}$;

(ii) its covariance matrix is $\mathbf{Q}_{KK}$, not $1/N$ times the diagonal matrix of 'statistical weights' $\boldsymbol{\varepsilon}$.

The Wilson distribution would actually be $\mathcal{P}^{\mathrm{SP}}(\mathbf{U}_K \,|\, \mathbf{U}_H = 0)$, corresponding to a uniform distribution of atoms $q^{\mathrm{ME}}(\mathbf{x}) = 1/V$.

1.2.4. *The diagonal approximation.* In this work we will use a cruder approximation than $\mathcal{P}^{\mathrm{SP}}$ to the c.p.d. of $\mathbf{U}_K$, called the *diagonal approximation* below, which consists of using as a covariance matrix $1/N$ times the diagonal matrix of statistical weights (or ε factors). Thus we retain the first-moment information $\mathbf{U}_K^{\mathrm{ME}}$ generated by the non-uniformity of q^{ME}, but discard the off-diagonal second-moment information associated with that non-uniformity. This decouples the various $U_{\mathbf{k}}$'s so that the resulting $\mathcal{P}_{\mathrm{diag}}^{\mathrm{SP}}$ may be written as

$$\mathcal{P}_{\mathrm{diag}}^{\mathrm{SP}}(\mathbf{U}_K \,|\, \mathbf{U}_H - \mathbf{U}_H^*) = \prod_{\mathbf{k}=K} \mathcal{P}_{\mathrm{diag}}^{\mathrm{SP}}(U_{\mathbf{k}} \,|\, \mathbf{U}_H = \mathbf{U}_H^*), \quad (1.5)$$

each factor being a one- or two-dimensional Gaussian offset from the origin by $U_{\mathbf{k}}^{\mathrm{ME}}$.

The diagonal approximation has the serious disadvantage that it is *unimodal*, so that it is unable to represent the branching behaviour which introduces ambiguities in the process of phase extension. This scrambling of the multimodal structure $\mathcal{P}^{\mathrm{SP}}$ is however less serious when the phases are integrated out to form conditional distributions of moduli (§ 1.3) and likelihood functions (§ 1.4). The advantage of the diagonal approximation is that it yields relatively simple explicit expressions for these various distributions, which we now proceed to derive.

Let $\mathbf{k}$ be an acentric reflexion. Write

$$U_{\mathbf{k}} = A_{\mathbf{k}} + iB_{\mathbf{k}} = |U_{\mathbf{k}}| \exp(i\varphi_{\mathbf{k}}),$$

and let

$$\mathrm{d}A_{\mathbf{k}}\, \mathrm{d}B_{\mathbf{k}} = |U_{\mathbf{k}}|\, \mathrm{d}|U_{\mathbf{k}}|\, \mathrm{d}\varphi_{\mathbf{k}}$$

be the plane measure in the corresponding copy of the complex plane. The conditional probability

distribution $\mathcal{P}^{\rm SP}_{\rm diag}$ is then

$$\begin{aligned}\mathcal{P}^{\rm SP}_{\rm diag}(U_{\mathbf{k}}|\mathbf{U}_H=\mathbf{U}_H^*)\,{\rm d}A_{\mathbf{k}}\,{\rm d}B_{\mathbf{k}}&\\ =(N/\pi\varepsilon_{\mathbf{k}})\exp[-(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}-U_{\mathbf{k}}^{\rm ME}|^2]\,{\rm d}A_{\mathbf{k}}\,{\rm d}B_{\mathbf{k}}&\\ =(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|\exp\{-(N/\varepsilon_{\mathbf{k}})[|U_{\mathbf{k}}|^2+|U_{\mathbf{k}}^{\rm ME}|^2]\}&\\ \times\exp[(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}||U_{\mathbf{k}}^{\rm ME}|&\\ \times\cos(\varphi_{\mathbf{k}}-\varphi_{\mathbf{k}}^{ME})]\,{\rm d}|U_{\mathbf{k}}|{\rm d}\varphi_{\mathbf{k}}/2\pi.&\qquad(1.6)\end{aligned}$$

The conditional distribution of the phase $\varphi_{\mathbf{k}}$ when $|U_{\mathbf{k}}|$ is known is therefore

$$\begin{aligned}P(\varphi_{\mathbf{k}}|\mathbf{U}_H=\mathbf{U}_H^*,|U_{\mathbf{k}}|=|U_{\mathbf{k}}|^{\rm obs})&\\ =[2\pi I_0(X_{\mathbf{k}})]^{-1}\exp[X_{\mathbf{k}}\cos(\varphi_{\mathbf{k}}-\varphi_{\mathbf{k}}^{\rm ME})],&\qquad(1.7a)\end{aligned}$$

where

$$X_{\mathbf{k}}=(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\rm obs}|U_{\mathbf{k}}^{\rm ME}|.\qquad(1.7b)$$

In spite of all the approximations made, this expression already incorporates the use of the triple-phase relationship, as shown in I [equations (3.23) and (4.15)].

Let $\mathbf{k}$ now be a centric reflexion. The phase $\varphi_{\mathbf{k}}$ has only two possible values, π apart, say $\omega_{\mathbf{k}}$ and $\omega_{\mathbf{k}}+\pi$; the quantity $\cos(\varphi_{\mathbf{k}}-\omega_{\mathbf{k}})$ is simply a sign, which will be denoted $s_{\mathbf{k}}$. If ${\rm d}U_{\mathbf{k}}$ denotes the line measure along the corresponding real line in the complex plane, then the equivalent of the polar coordinate relations is

$${\rm d}U_{\mathbf{k}}={\rm d}|U_{\mathbf{k}}|^{\frac{1}{2}}[\delta(s_{\mathbf{k}}-1)+\delta(s_{\mathbf{k}}+1)],$$

where the sum of δ functions (a discrete measure representing a choice of sign) is the centric equivalent of ${\rm d}\varphi_{\mathbf{k}}/2\pi$. It follows that

$$\begin{aligned}\mathcal{P}^{\rm SP}_{\rm diag}(U_{\mathbf{k}}|\mathbf{U}_H=\mathbf{U}_H^*)\,{\rm d}U_{\mathbf{k}}&\\ =(N/2\pi\varepsilon_{\mathbf{k}})^{1/2}\exp[-(N/2\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}-U_{\mathbf{k}}^{\rm ME}|^2]\,{\rm d}U_{\mathbf{k}}&\\ =(2N/\pi\varepsilon_{\mathbf{k}})^{1/2}\exp\{-(N/2\varepsilon_{\mathbf{k}})[|U_{\mathbf{k}}|^2+|U_{\mathbf{k}}^{\rm ME}|^2]\}&\\ \times\exp[(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}||U_{\mathbf{k}}^{\rm ME}|s_{\mathbf{k}}s_{\mathbf{k}}^{\rm ME}]\,{\rm d}|U_{\mathbf{k}}|&\\ \times\tfrac{1}{2}[\delta(s_{\mathbf{k}}-1)+\delta(s_{\mathbf{k}}+1)],&\qquad(1.8)\end{aligned}$$

giving for the conditional distribution of the sign $s_{\mathbf{k}}$ when $|U_{\mathbf{k}}|$ is known:

$$\begin{aligned}P(s_{\mathbf{k}}|\mathbf{U}_H=\mathbf{U}_H^*,|U_{\mathbf{k}}|=|U_{\mathbf{k}}|^{\rm obs})&\\ =[2\cosh(X_{\mathbf{k}})]^{-1}\exp[X_{\mathbf{k}}s_{\mathbf{k}}s_{\mathbf{k}}^{\rm ME}],&\qquad(1.9a)\end{aligned}$$

where

$$X_{\mathbf{k}}=(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\rm obs}|U_{\mathbf{k}}^{\rm ME}|.\qquad(1.9b)$$

A straightforward adaptation to the centric case of equations (3.23) and (4.15) of paper I shows that this approximation, although crude, incorporates the use of the hyperbolic-tangent formula of Cochran & Woolfson (1955).

1.3. *Conditional distributions of moduli*

The calculations carried out so far convey the (correct) impression that entropy maximization has a useful built-in ability to extrapolate phase information. Indeed, even in the case of the small protein crambin, very significant phase extension was shown to take place in conditions where direct methods are powerless (I, § 7.3). But, in this case, the starting phases to 3 Å resolution were exact, and the question immediately arose of knowing whether, in a real situation, errors in the starting phases would not be amplified so as to make the extrapolated phases worthless. In other words, maximum-entropy phase extension looks like an inherently divergent process, capable of some useful extrapolation from good starting phases, but perhaps incapable of refining bad starting phases prior to their extrapolation. This section will be devoted to showing that this is not the case: phase refinement is possible through the use of *likelihood functions* introduced in I (§ 4.2.2).

The basic observation is the following: if we integrate $\mathcal{P}^{\rm SP}(\mathbf{U}_K|\mathbf{U}_H=\mathbf{U}_H^*)$ with respect to the (unobservable) phases in $\mathbf{U}_K$, we obtain a conditional marginal distribution for the (observable) moduli $\mathcal{P}^{\rm SP}(|\mathbf{U}_K||\mathbf{U}_H=\mathbf{U}_H^*)$ which differs from the standard Wilson distribution of moduli, the latter being $\mathcal{P}^{\rm SP}(|\mathbf{U}_K||\mathbf{U}_H=\mathbf{0})$.

This consideration is of fundamental importance in the present work: through the use of the saddle-point approximation, phase choices for the 'basis' reflexions in H induce a *deformation* of the conditional marginal distribution of the moduli in K away from the usual (Wilson) distribution. In other words, the maximum-entropy extrapolation which relates $\mathbf{U}_H^*$ to $\mathbf{U}_K^{\rm ME}$ acts as a *transducer*, converting hypotheses about phase values in $\mathbf{U}_H^*$ (which cannot be tested directly from measured intensities) into hypotheses about a change in the statistical distribution of the moduli $|\mathbf{U}_K|$ (which *can* be tested).

This transduction effect persists (although somewhat weakened) even if we use the diagonal approximation $\mathcal{P}^{\rm SP}_{\rm diag}$, because of the origin offset $\mathbf{U}_K^{\rm ME}$. Explicit expressions for this level of approximation are easily obtained by integrating out phases or signs.

For an acentric reflexion $\mathbf{k}$ the above derivations immediately yield

$$\begin{aligned}\mathcal{P}^{\rm SP}_{\rm diag}(|U_{\mathbf{k}}||\mathbf{U}_H=\mathbf{U}_H^*)\,{\rm d}|U_{\mathbf{k}}|&\\ =(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|\exp\{-(N/\varepsilon_{\mathbf{k}})[|U_{\mathbf{k}}|^2+|U_{\mathbf{k}}^{\rm ME}|^2]\}&\\ \times I_0[(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}||U_{\mathbf{k}}^{\rm ME}|]\,{\rm d}|U_{\mathbf{k}}|&\qquad(1.10)\end{aligned}$$

[a distribution first derived by Rice (1944, 1945; reprinted in Wax, 1954) in another context], while for a centric reflexion

$$\begin{aligned}\mathcal{P}^{\rm SP}_{\rm diag}(|U_{\mathbf{k}}||\mathbf{U}_H=\mathbf{U}_H^*)\,{\rm d}|U_{\mathbf{k}}|&\\ =(2N/\pi\varepsilon_{\mathbf{k}})^{1/2}\exp\{-(N/2\varepsilon_{\mathbf{k}})[|U_{\mathbf{k}}|^2+|U_{\mathbf{k}}^{\rm ME}|^2]\}&\\ \times\cosh[(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}||U_{\mathbf{k}}^{\rm ME}|]\,{\rm d}|U_{\mathbf{k}}|.&\qquad(1.11)\end{aligned}$$

1.4. *Likelihood functions, likelihood ratios*

For the purpose of testing hypotheses, we use the customary *likelihood* criterion where the likelihood of a hypothesis is defined as the probability it assigned to the actual result of an observation before that observation was performed. In the case at hand, we therefore define the likelihood Λ of the phase choices in $\mathbf{U}_H^*$ as the conditional probability of the observed values $|\mathbf{U}_K|^{\text{obs}}$:

$$\Lambda(\mathbf{U}_H^* \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}) = \mathcal{P}^{\text{SP}}(|\mathbf{U}_K|^{\text{obs}} \,|\mathbf{U}_H = \mathbf{U}_H^*). \tag{1.12}$$

In practice, it is convenient to normalize this quantity with respect to the null hypothesis (H_0) that the distribution of atoms is uniform (*i.e.* that $\mathbf{U}_H = \mathbf{0}$) and to use the *likelihood ratio*

$$\frac{\Lambda(\mathbf{U}_H^* \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}})}{\Lambda(\mathbf{0} \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}})} = \frac{\mathcal{P}^{\text{SP}}(|\mathbf{U}_K|^{\text{obs}} \,|\mathbf{U}_H = \mathbf{U}_H^*)}{\mathcal{P}^{\text{SP}}(|\mathbf{U}_K|^{\text{obs}} \,|\mathbf{U}_H = \mathbf{0})} \tag{1.13}$$

or its logarithm, the *log-likelihood gain*

$$L(\mathbf{U}_H^* \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}) - L(\mathbf{0} \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}) \tag{1.14}$$

where $L = \log \Lambda$. For brevity, we will often simply write $\Lambda(\mathbf{U}_H^*)$ or $\Lambda(\mathbf{0})$ when the observation results used in forming the likelihood (here, $|\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}$) are unambiguously defined by the context.

According to the fundamental work of Neyman & Pearson (1933), the likelihood ratio just defined constitutes the most powerful statistical criterion for identifying the best set of phases for $\mathbf{U}_H^*$ on the basis of the information $|\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}$. This criterion measures the extent to which the observed values of the yet unphased moduli $|\mathbf{U}_K|^{\text{obs}}$ are made more probable by the assumption that $\mathbf{U}_H = \mathbf{U}_H^*$ than by the assumption that $\mathbf{U}_H = \mathbf{0}$.

Explicit expressions for likelihoods are easily obtained in the diagonal approximation from the c.p.d.'s of moduli derived above. For **k** acentric,

$$\begin{aligned}\Lambda(\mathbf{U}_H^* \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}}) &= (2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}} \\ &\times \exp\{-(N/\varepsilon_{\mathbf{k}})[(|U_{\mathbf{k}}|^{\text{obs}})^2 + |U_{\mathbf{k}}^{\text{ME}}|^2]\} \\ &\times I_0[(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}} \,|U_{\mathbf{k}}^{\text{ME}}|]\end{aligned} \tag{1.15a}$$

while

$$\Lambda(\mathbf{0} \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}}) = (2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}} \exp\{-(N/\varepsilon_{\mathbf{k}})(|U_{\mathbf{k}}|^{\text{obs}})^2\}. \tag{1.15b}$$

Hence

$$\frac{\Lambda(\mathbf{U}_H^* \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}})}{\Lambda(\mathbf{0} \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}})} = \exp\{-(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}^{\text{ME}}|^2\} I_0[(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}}|U_{\mathbf{k}}^{\text{ME}}|] \tag{1.15c}$$

so that the global log-likelihood criterion from all the acentric moduli in K reads

$$\begin{aligned}L(\mathbf{U}_H^*) - L(\mathbf{0}) &= \sum_{\substack{\mathbf{k}\in K \\ \mathbf{k}\ \text{acentric}}} \{\log I_0[(2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}}\,|U_{\mathbf{k}}^{\text{ME}}|] \\ &\quad -(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}^{\text{ME}}|^2\}.\end{aligned} \tag{1.15d}$$

Similarly, for **k** centric,

$$\begin{aligned}\Lambda(\mathbf{U}_H^* \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}}) &= (2N/\pi\varepsilon_{\mathbf{k}})^{1/2} \\ &\times \exp\{-(N/2\varepsilon_{\mathbf{k}})[(|U_{\mathbf{k}}|^{\text{obs}})^2 + |U_{\mathbf{k}}^{\text{ME}}|^2]\} \\ &\times \cosh[(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}}|U_{\mathbf{k}}^{\text{ME}}|]\end{aligned} \tag{1.16a}$$

while

$$\Lambda(\mathbf{0} \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}}) = (2N/\pi\varepsilon_{\mathbf{k}})^{1/2} \exp\{-(N/2\varepsilon_{\mathbf{k}})(|U_{\mathbf{k}}|^{\text{obs}})^2\}. \tag{1.16b}$$

Hence,

$$\begin{aligned}\frac{\Lambda(\mathbf{U}_H^* \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}})}{\Lambda(\mathbf{0} \,||U_{\mathbf{k}}| = |U_{\mathbf{k}}|^{\text{obs}})} &= \exp\{-(N/2\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}^{\text{ME}}|^2\} \\ &\times \cosh[(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}}|U_{\mathbf{k}}^{\text{ME}}|]\end{aligned} \tag{1.16c}$$

so that the global log-likelihood criterion from all the centric moduli in K reads:

$$\begin{aligned}L(\mathbf{U}_H^*) - L(\mathbf{0}) &= \sum_{\substack{\mathbf{k}\in K \\ \mathbf{k}\ \text{centric}}} [\log\cosh[(N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\text{obs}}|U_{\mathbf{k}}^{\text{ME}}|] \\ &\quad -(N/2\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}^{\text{ME}}|^2].\end{aligned} \tag{1.16d}$$

Because the diagonal approximation has been used, these likelihood functions are sensitive *only to the extrapolated moduli* $|\mathbf{U}_K^{\text{ME}}|$, and not to the associated phases. An exact expression for the likelihood, incorporating the full covariance matrix $\mathbf{Q}_{KK}$ (§ 1.2.1), has been derived by one of us (GB) and will be described elsewhere; it is sensitive to the extrapolated phases as well. Other forms of this likelihood function are suitable for diffraction intensities from fibres or powders rather than from single crystals.

1.5. *Phase refinement by Bayes's theorem*

The *a priori* probability $\mathcal{P}^{\text{SP}}(\mathbf{U}_H^*)$ may be combined with the likelihood $\Lambda(\mathbf{U}_H^* \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}})$ by means of Bayes's theorem (see *e.g.* Lindley, 1965), yielding the *a posteriori* probability

$$\mathcal{P}_{\text{post}}(\mathbf{U}_H^* \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}) \propto \mathcal{P}^{\text{SP}}(\mathbf{U}_H^*)\Lambda(\mathbf{U}_H^* \,||\mathbf{U}_K| = |\mathbf{U}_K|^{\text{obs}}) \tag{1.17}$$

whose maximization with respect to the phases in $\mathbf{U}_H^*$ provides a procedure for phase refinement (II, § 0.6; III, pp. 68–72). It was shown in I (§ 4.2.2) and III (pp. 70–72) that this approach is an enhancement of the use of a 'quartet figure of merit' which can be implemented without any explicit manipulation of phase invariants.

Using the approximation $\mathcal{P}^{\mathrm{SP}}(\mathbf{U}_H^*) \simeq \exp[N\mathcal{S}_m(q^{\mathrm{ME}})]$, it is clear that Bayes's theorem leads to consideration of the compound criterion

$$N\mathcal{S}_m(q^{\mathrm{ME}}) + L(\mathbf{U}_H^* \,|\, |\mathbf{U}_K| = |\mathbf{U}_K|^{\mathrm{obs}}) \quad (1.18)$$

as the best discriminator between phase sets for $\mathbf{U}_H^*$ on the basis of the information contained in the observed moduli associated to H and K.

In practice, L is often much more sensitive to $\mathbf{U}_H^*$ than $\mathcal{S}_m$, especially at the early stages of phase determination. As a result, we often carry out phase refinement by maximization of likelihood alone.

1.6. *Centroid extrapolated structure factors*

Once the trial phases assigned to the moduli $|\mathbf{U}_H^*|$ have been refined by likelihood maximization against the unphased moduli $|\mathbf{U}_K|^{\mathrm{obs}}$, the phase dependence of the conditional probability distributions $\mathcal{P}^{\mathrm{SP}}_{\mathrm{diag}}(U_{\mathbf{k}} \,|\, \mathbf{U}_H = \mathbf{U}_H^*)$, which was integrated out in order to form the likelihood function (§§ 1.3 and 1.4), may be preserved and used to calculate the centroid value $\langle U_{\mathbf{k}} \rangle$ of each structure factor $U_{\mathbf{k}}$. These centroid values are analogous to the 'best' structure-factor values defined by Blow & Crick (1959) in another context, and have the property that they minimize the mean-square error in the corresponding maps caused by the residual phase uncertainties.

Using the expressions derived in § 1.2 for the conditional distribution of the phase $\varphi_{\mathbf{k}}$, we get

(i) for $\mathbf{k}$ acentric,

$$\langle U_{\mathbf{k}} \rangle = |U_{\mathbf{k}}|^{\mathrm{obs}} [I_1(X_{\mathbf{k}})/I_0(X_{\mathbf{k}})] \exp(i\varphi_{\mathbf{k}}^{\mathrm{ME}})$$

with

$$X_{\mathbf{k}} = (2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\mathrm{obs}}|U_{\mathbf{k}}^{\mathrm{ME}}|; \quad (1.19)$$

(ii) for $\mathbf{k}$ centric,

$$\langle U_{\mathbf{k}} \rangle = |U_{\mathbf{k}}|^{\mathrm{obs}} \tanh(X_{\mathbf{k}}) \exp(i\varphi_{\mathbf{k}}^{\mathrm{ME}})$$

with

$$X_{\mathbf{k}} = (N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\mathrm{obs}}|U_{\mathbf{k}}^{\mathrm{ME}}|. \quad (1.20)$$

Maps calculated from this extended set of structure factors will show enhanced resolution without any new phase choices having been made.

2. Algorithms

2.1. *Data preparation*

The conversion of structure-factor amplitudes ($|F|$'s) to unitary ($|U|$'s) and quasi-normalized ($|E|$'s) amplitudes, together with the estimation of the associated variances, is described in § 1 of the companion paper (Gilmore, Bricogne & Bannister, 1990). However, a remark on these two types of normalization is in order here. The normalization of $|F|$'s to $|U|$'s is appropriate for *first-order* statistics (*e.g.* expectation values), while that of $|F|$'s to $|E|$'s is appropriate for *second-order* statistics (*e.g.* variances). Some blurring of this distinction occurs in conventional direct methods because the assumption of uniformly distributed atoms causes the first-order expectation of each structure factor to vanish, leaving only a second-order (fluctuation) term to contribute to the intensity distribution. Here, the use of non-uniform distributions requires that this distinction be maintained. In all formulae derived in § 1 of this paper, it is clear that second-order quantities involved as arguments of exponentials (or of special functions obtained by integrating out phases or signs) would be more simply expressed in terms of $|E|$'s because of the presence of the statistical weight ε. However, when we come to calculate maps for visual inspection, we use centroid maps (§ 1.6) in which the *weights* involve $|E|$'s but the *coefficients* are $|U|$'s. In our opinion, E maps are a dubious hybrid in which second-order quantities ($|E|$'s) are used to calculate a first-order quantity (the expectation value of the sharpened electron-density map): the statistical weight ε has no rôle to play in the latter, and its use may induce substantial distortions in space groups of high symmetry. This viewpoint is supported by the occasional occurrence of crystal structures for which the E map is not interpretable while the corresponding F map is.

2.2. *The maximum-entropy equations and exponential modelling*

The hallmark of the multisolution approach used in this work (I, §§ 2.4, 8.1) is the constant updating of the distribution $q(\mathbf{x})$ of random atoms in the light of all the phase choices made. To serve as the basis for the calculation of saddlepoint approximations to j.p.d.'s and c.p.d.'s near the point $\mathbf{U}_H = \mathbf{U}_H^*$, $q(\mathbf{x})$ must be the unique distribution $q^{\mathrm{ME}}(\mathbf{x})$ with maximum entropy $\mathcal{S}_m(q)$ relative to the 'prior prejudice' $m(\mathbf{x})$ which would be used in the absence of phase assumptions.

It was shown in I (§ 3.3.1) that when $q(\mathbf{x})$ has maximum entropy relative to $m(\mathbf{x})$ under the constraint $\mathbf{U}_H = \mathbf{U}_H^*$, then it can be represented exactly by an exponential model

$$q^{\mathrm{ME}}(\mathbf{x}) = [m(\mathbf{x})/Z(m, \omega)] \exp[\omega(\mathbf{x})], \quad (2.1)$$

where Z is defined in terms of m and ω by

$$Z(m, \omega) = \int_V m(\mathbf{x}) \exp[\omega(\mathbf{x})] \, \mathrm{d}^3\mathbf{x} \quad (2.2)$$

so that the division by Z automatically ensures that $q^{\mathrm{ME}}(\mathbf{x})$ is normalized. The ω map is defined (*e.g.* in *P*1) by

$$\omega(\mathbf{x}) = 2\,\mathrm{Re} \sum_{\mathbf{h}\in H} \zeta_{\mathbf{h}} \exp(-2\pi i \mathbf{h}\,.\,\mathbf{x}) \qquad (2.3)$$

where the complex Lagrange multipliers $\zeta_{\mathbf{h}}$ are determined by the condition that

$$\int_V q^{\mathrm{ME}}(\mathbf{x}) \exp(+2\pi i \mathbf{h}\,.\,\mathbf{x})\, \mathrm{d}^3\mathbf{x} = U_{\mathbf{h}}^* \qquad (2.4)$$

for each $\mathbf{h}\in H$.

A simpler exponential model [with $m(\mathbf{x})$ a uniform distribution] was used by Collins & Mahar (1983) as a means of enforcing positivity, but the latter criterion on its own fails to justify the crucial property of the ω map (equation 2.3) that it should have non-zero Fourier coefficients $\zeta_{\mathbf{h}}$ *only for* $\mathbf{h}\in H$. This justification requires the maximum-entropy formalism developed in I, §§ 3.3–3.5.

Equations (2.4) are referred to as the maximum-entropy equations, and several methods have been proposed for their solution (see *e.g.* I, § 7.1). We have chosen to use an algorithm based on an exponential model because

(i) its computational complexity increases only as $\mathcal{N}\log\mathcal{N}$ for $\mathcal{N}$ reflexions, so that it can be applied to large structures; a Newton method (I, § 7.1.1), whose complexity increases as $\mathcal{N}^2$, would be quicker for small $\mathcal{N}$ but would become prohibitively expensive for large $\mathcal{N}$;

(ii) the amount of storage required to hold a complete description of q^{ME} is minimal (two words per reflexion in H, to store $\zeta_{\mathbf{h}}$); while algorithms such as that of Bryan & Skilling (1980) store q^{ME} itself as a map, which has to be highly oversampled in order to accommodate the extrapolated Fourier coefficients and is therefore very large.

A naïve algorithm for solving the ME equations by exponential modelling (I, § 7.1.2) consists of the following iterative procedure (written for *P*1 for simplicity), where the superscript $^{(i)}$ [not to be confused with $i=(-1)^{1/2}$ in the exponents] indicates quantities pertaining to the ith iteration:

$$(a)\quad \omega^{(i)}(\mathbf{x}) = 2\,\mathrm{Re} \sum_{\mathbf{h}\in H} \zeta_{\mathbf{h}}^{(i)} \exp(-2\pi i \mathbf{h}\,.\,\mathbf{x}) \qquad (2.5a)$$

$$(b)\quad q^{(i)}(\mathbf{x}) = [m(\mathbf{x})/Z(m,\omega^{(i)})] \exp[\omega^{(i)}(\mathbf{x})] \qquad (2.5b)$$

with Z given by (2.2)

$$(c)\quad U_{\mathbf{h}}^{(i)} = \int_V q^{(i)}(\mathbf{x}) \exp(+2\pi i \mathbf{h}\,.\,\mathbf{x})\, \mathrm{d}^3\mathbf{x} \qquad (2.5c)$$

$$(d)\quad \delta q^{(i)}(\mathbf{x}) = 2\,\mathrm{Re} \sum_{\mathbf{h}\in H} [U_{\mathbf{h}}^* - U_{\mathbf{h}}^{(i)}] \times \exp(-2\pi i \mathbf{h}\,.\,\mathbf{x}) \qquad (2.5d)$$

$$(e)\quad \delta\zeta_{\mathbf{h}}^{(i)} = \int_V [\delta q^{(i)}(\mathbf{x})/q^{(i)}(\mathbf{x})] \times \exp(+2\pi i \mathbf{h}\,.\,\mathbf{x})\, \mathrm{d}^3\mathbf{x} \qquad (2.5e)$$

$$(f)\quad \zeta_{\mathbf{h}}^{(i+1)} = \zeta_{\mathbf{h}}^{(i)} + \delta\zeta_{\mathbf{h}}^{(i)} \quad \text{for } \mathbf{h}\in \mathrm{H}. \qquad (2.5f)$$

General space-group symmetry is readily incorporated into the Fourier synthesis step (*a*). Note that the correction $\delta\omega^{(i)} = \omega^{(i+1)} - \omega^{(i)}$ is related to $\delta q^{(i)}/q^{(i)}$ by the operation of selecting only those Fourier coefficients with indices in H. In the sequel, this operation will be called *spectrum truncation* and will be denoted T_H.

This naïve algorithm tends to be very unstable, because the division operation in step (*e*) gives rise to uncontrollably large shifts as the dynamic range of $q^{(i)}$ increases.

2.3. *A robust exponential modelling algorithm*

The exponential modelling technique used in this work was developed by one of us (GB) specifically for entropy maximization and was used for phase-extension calculations on crambin (I, § 7.1.2, and unpublished results of GB). The inherent instability of the naïve algorithm is brought under control by the use of a *plane search* to guard against excessive shifts. This technique will now be described in some detail.

2.3.1. *Calculation of functional derivatives.* In the forthcoming calculations we will frequently vary the 'map' $\omega(\mathbf{x})$ (or equivalently its Fourier coefficients $\zeta_{\mathbf{h}}$) so as to optimize some functional expression (call it $F[q]$) depending on the 'map' $q(\mathbf{x})$ or equivalently on its Fourier coefficients. For this purpose we will need to calculate functional derivatives such as $\partial F/\partial q(\mathbf{x})$ and $\partial F/\partial\omega(\mathbf{x})$, which will sometimes be denoted $\nabla_q F$ and $\nabla_\omega F$, respectively. Their definition is the continuous analogue of ordinary partial derivatives: for instance if $q(\mathbf{x})$ is varied by $\delta q(\mathbf{x})$, then the corresponding variation of F is

$$\delta F = \int_V [\partial F/\partial q(\mathbf{x})]\delta q(\mathbf{x})\, \mathrm{d}^3\mathbf{x} = \langle \nabla_q F, \delta q\rangle, \qquad (2.6)$$

where the angle brackets $\langle\,.\,,.\,\rangle$ denote the scalar product of two real-valued functions. To compute $\nabla_\omega F$ when $\nabla_q F$ is known, it suffices to note that if $\omega(\mathbf{x})$ varies by $\delta\omega(\mathbf{x})$, then $q(\mathbf{x})$ varies by

$$\begin{aligned}\delta q(\mathbf{x}) &= q(\mathbf{x})[\delta\omega(\mathbf{x}) - \delta(\log Z)] \\ &= q(\mathbf{x})[\delta\omega(\mathbf{x}) - \langle q, \delta\omega\rangle] \end{aligned} \qquad (2.7)$$

and therefore

$$\delta F = \langle q\nabla_q F, \delta\omega\rangle - \langle q, \nabla_q F\rangle\langle q, \delta\omega\rangle \qquad (2.8)$$

or equivalently

$$(\nabla_\omega F)(\mathbf{x}) = q(\mathbf{x})(\nabla_q F)(\mathbf{x}) - \langle q, \nabla_q F\rangle q(\mathbf{x}). \qquad (2.9)$$

The second term in each of the last two formulae is

a correction term which keeps $q(\mathbf{x})$ normalized as $\omega(\mathbf{x})$ varies.

2.3.2. *Plane search algorithm.* The algorithm used in this work overcomes the inherent instability of naïve exponential modelling by means of two modifications.

(i) The optimal corrections $\delta\zeta_{\mathbf{h}}^{(i)}$ [or, equivalently, the optimal $\delta\omega^{(i)}(\mathbf{x})$] are sought not under the 'hard' constraint $\mathbf{U}_H = \mathbf{U}_H^*$ but under a softer constraint defined in terms of a χ^2-like statistic $\mathscr{C}$. This relaxation is introduced to allow for the existence of measurement errors in the moduli $|\mathbf{U}_H|^{\text{obs}}$ and of errors in the trial phases associated with them in $\mathbf{U}_H^*$. The statistic $\mathscr{C}$ is modelled somewhat heuristically after a log-likelihood, in keeping with the remark at the end of § 1.5; its precise definition is as follows:

$$\mathscr{C} = \frac{1}{2} \sum_{\substack{\mathbf{h}\in H \\ \mathbf{h}\ \text{acentric}}} \frac{|U_{\mathbf{h}} - U_{\mathbf{h}}^*|^2}{\varepsilon_{\mathbf{h}}\Sigma_a + \sigma_{\mathbf{h}}^2} + \frac{1}{2} \sum_{\substack{\mathbf{h}\in H \\ \mathbf{h}\ \text{centric}}} \frac{|U_{\mathbf{h}} - U_{\mathbf{h}}^*|^2}{\varepsilon_{\mathbf{h}}\Sigma_c + \sigma_{\mathbf{h}}^2}, \quad (2.10)$$

where $\sigma_{\mathbf{h}}^2$ is the estimated variance of $|U_{\mathbf{h}}|^{\text{obs}}$ and where initially $2\Sigma_a = \Sigma_c = 1/N$ (the parameters Σ_a and Σ_c are subsequently refined, see § 2.4).

(ii) Another search direction [besides $T_H(\delta q^{(i)}/q^{(i)})$] is used to temper the build up of contrast in $q^{(i)}$. Together, these two search directions span a *plane*, in which the optimal corrections are sought by constructing simultaneous bicubic models of the entropy $\mathscr{S}$ and of the constraint $\mathscr{C}$.

2.3.2.1 *Search directions.* Consider the ith cycle of the iterative determination of $q^{\text{ME}}(\mathbf{x})$, starting for instance from $q^{(0)}(\mathbf{x}) = m(\mathbf{x})$ or equivalently $\omega^{(0)}(\mathbf{x}) = 0$.

For the purpose of decreasing $\mathscr{C}$ we use a search direction $\Delta_t\omega^{(i)}$ analogous to the direction $T_H(\delta q^{(i)}/q^{(i)})$ defined by steps (d) and (e) above, but incorporating the weights present in $\mathscr{C}$; the subscript t indicates that the coordinate along this direction in the search plane will be called t. Thus we define

$$(d') \quad \Delta q^{(i)}(\mathbf{x}) = 2\,\text{Re} \sum_{\mathbf{h}\in H} \frac{U_{\mathbf{h}}^* - U_{\mathbf{h}}^{(i)}}{\varepsilon_{\mathbf{h}}\Sigma + \sigma_{\mathbf{h}}^2} \times \exp(-2\pi i \mathbf{h}\,.\,\mathbf{x}) \quad (2.5d')$$

with $\Sigma = \Sigma_a$ or $\Sigma = \Sigma_c$ according to whether $\mathbf{h}$ is acentric or centric,

$$(e') \quad \Delta\zeta_{\mathbf{h}}^{(i)} = \int_V \frac{\Delta q^{(i)}(\mathbf{x})}{q^{(i)}(\mathbf{x})} \exp(+2\pi i \mathbf{h}\,.\,\mathbf{x})\, \mathrm{d}^3\mathbf{x}, \quad (2.5e')$$

$$(a') \quad \Delta_t\omega^{(i)}(\mathbf{x}) = 2\,\text{Re} \sum_{\mathbf{h}\in H} \Delta\zeta_{\mathbf{h}}^{(i)} \times \exp(-2\pi i \mathbf{h}\,.\,\mathbf{x}), \quad (2.5a')$$

where it is understood that the Fourier summations in (d') and (a'), written here for $P1$, must incorporate any space-group symmetry which may be present.

To control the dynamic range of the successive approximations $q^{(i)}$, we use as a second search direction

$$\Delta_s\omega^{(i)} = -\omega^{(i)} \quad (2.11)$$

along which we define a coordinate s.

In the search plane defined by $\Delta_s\omega^{(i)}$ and $\Delta_t\omega^{(i)}$, the point with coordinates (s, t) represents the new iterate

$$(f') \quad \omega_{s,t}^{(i+1)} = \omega^{(i)} + s\Delta_s\omega^{(i)} + t\Delta_t\omega^{(i)},$$

i.e.

$$\omega_{s,t}^{(i+1)} = (1-s)\omega^{(i)} + tT_H(\Delta_t q^{(i)}/q^{(i)}). \quad (2.5f')$$

2.3.2.2. *Directional derivatives.* Through the exponential modelling relation $(2.5b)$ defining $q_{s,t}^{(i+1)}$ in terms of $\omega_{s,t}^{(i+1)}$, the relative entropy $\mathscr{S}$ and the constraint $\mathscr{C}$ become numerical functions of the values of s and t used in the updating step $(2.5f')$, which we write as

$$\mathscr{S}_m(q_{s,t}^{(i+1)}) = S(s, t) \quad (2.12)$$

$$\mathscr{C}(q_{s,t}^{(i+1)}) = C(s, t). \quad (2.13)$$

The evaluation of these functions and of their partial derivatives with respect to s and t proceeds as follows:

(i) to evaluate S, use the expression

$$\mathscr{S}_m(q_{s,t}^{(i+1)}) = \log Z[m, \omega_{s,t}^{(i+1)}] - \langle q_{s,t}^{(i+1)}, \omega_{s,t}^{(i+1)}\rangle; \quad (2.14)$$

(ii) to evaluate C, follow steps $(2.5b)$ and $(2.5c)$ with $\omega = \omega_{s,t}^{(i+1)}$ and use the definition of $\mathscr{C}$ given in (2.10);

(iii) to evaluate the derivatives of S, it is simplest to obtain directly the functional derivative $\nabla_\omega\mathscr{S}$ according to (2.9), *i.e.*

$$\nabla_\omega\mathscr{S} = -\omega q + \langle q, \omega\rangle q, \quad (2.15)$$

and thus at a general point (s, t)

$$\begin{aligned} \mathrm{d}S/\mathrm{d}s &= \langle \nabla_\omega\mathscr{S}, \Delta_s\omega^{(i)}\rangle \\ &= \langle \omega_{s,t}^{(i+1)} q_{s,t}^{(i+1)}, \omega^{(i)}\rangle \\ &\quad - \langle q_{s,t}^{(i+1)}, \omega_{s,t}^{(i+1)}\rangle\langle q_{s,t}^{(i+1)}, \omega^{(i)}\rangle \quad (2.16) \\ \mathrm{d}S/\mathrm{d}t &= \langle \nabla_\omega\mathscr{S}, \Delta_t\omega^{(i)}\rangle \\ &= -\langle \omega_{s,t}^{(i+1)} q_{s,t}^{(i+1)}, \Delta_t\omega^{(i)}\rangle \\ &\quad + \langle q_{s,t}^{(i+1)}, \omega_{s,t}^{(i+1)}\rangle\langle q_{s,t}^{(i+1)}, \Delta_t\omega^{(i)}\rangle; \quad (2.17) \end{aligned}$$

(iv) to evaluate the derivatives of C, note that the functional derivative $\nabla_q\mathscr{C}$ with respect to q is simply $-\Delta q$ as defined by $(2.5d')$. Therefore, using (2.9), we get

$$\begin{aligned} \nabla_\omega\mathscr{C} &= q\nabla_q\mathscr{C} - \langle q, \nabla_q\mathscr{C}\rangle q \\ &= -q\Delta q + \langle q, \Delta q\rangle q \quad (2.18) \end{aligned}$$

and hence

$$\begin{aligned} dC/ds &= \langle \nabla_\omega \mathscr{C}, \Delta_s \omega^{(i)} \rangle \\ &= \langle q_{s,t}^{(i+1)} \Delta q_{s,t}^{(i+1)}, \omega^{(i)} \rangle \\ &\quad - \langle q_{s,t}^{(i+1)}, \Delta q_{s,t}^{(i+1)} \rangle \langle q_{s,t}^{(i+1)}, \omega^{(i)} \rangle, \end{aligned} \tag{2.19}$$

$$\begin{aligned} dC/dt &= \langle \nabla_\omega \mathscr{C}, \Delta_t \omega^{(i)} \rangle \\ &= -\langle q_{s,t}^{(i+1)} \Delta q_{s,t}^{(i+1)}, \Delta_t \omega^{(i)} \rangle \\ &\quad + \langle q_{s,t}^{(i+1)}, \Delta q_{s,t}^{(i+1)} \rangle \langle q_{s,t}^{(i+1)}, \Delta_t \omega^{(i)} \rangle. \end{aligned} \tag{2.20}$$

2.3.2.3. *Bicubic model and its solution.* Since $\Delta_t \omega^{(i)}$ is a Newton direction for $\mathscr{C}$ (see I, § 7.1.2) and since a unit shift along $\Delta_s \omega^{(i)}$ with $t = 0$ would flatten the map, we can expect to find the optimum values of s and t within the unit square in the search plane; experience shows that it is most frequently contained in a much smaller region, typically

$$0 \le s \le 0{\cdot}1, \qquad 0 \le t \le 0{\cdot}3.$$

We therefore select trial values of s and t (*e.g.* $s_{\text{trial}} = 0{\cdot}1$, $t_{\text{trial}} = 0{\cdot}3$) and compute the following quantities:

$$\begin{aligned} &S(0,0), \quad S(s_{\text{trial}}, 0), \quad S(0, t_{\text{trial}}), \quad S(s_{\text{trial}}, t_{\text{trial}}), \\ &dS/ds(0,0), \quad dS/ds(s_{trial}, 0), \quad dS/ds(0, t_{\text{trial}}), \\ &dS/dt(0,0), \quad dS/dt(s_{\text{trial}}, 0), \quad dS/dt(0, t_{\text{trial}}) \end{aligned} \tag{2.21}$$

and similarly for C. These ten numbers attached to each function define uniquely a bicubic approximation of the form

$$\begin{aligned} S(s,t) \simeq\ & S_0 + S_1 s + S_2 t + S_{11} s^2 + S_{12} st + S_{22} s^2 \\ & + S_{111} s^3 + S_{112} s^2 t + S_{122} s t^2 + S_{222} t^3 \end{aligned} \tag{2.22}$$

(and similarly for C), the correspondence between the two sets of numbers being given by a 10×10 matrix which is easily calculated and inverted so as to give the coefficients in (2.22) from the function and derivative values in (2.21). Once these bicubic models of S and C are set up they are evaluated on a 100×100 grid covering a 'trial patch' (which may include some slightly negative values of both s and t), and level curves of C are traced using the contouring algorithm of Diamond (1982).

The optimization process then consists of

(i) fixing a value C_{aim} of C as a target for the current iteration, the ultimate target being $C_{\text{aim}}^{\infty} = n/2$ if H contains n reflexions. Let C_{cur} denote the current value $C(0,0)$ of C, C_{ext} the minimum (maximum) value of C in the trial patch if $C_{\text{aim}}^{\infty} \le C_{\text{cur}}$ ($C_{\text{aim}}^{\infty} \ge C_{\text{cur}}$), and put

$$C_{\text{test}} = \tfrac{1}{2} C_{\text{cur}} + \tfrac{1}{2} C_{\text{ext}}.$$

Then C_{aim} is chosen as follows:

$$\begin{aligned} C_{\text{aim}} &= C_{\text{aim}}^{\infty} \quad \text{if } |C_{\text{aim}}^{\infty} - C_{\text{cur}}| \le |C_{\text{test}} - C_{\text{cur}}|, \\ C_{\text{aim}} &= C_{\text{test}} \quad \text{if } |C_{\text{aim}}^{\infty} - C_{\text{cur}}| \ge |C_{\text{test}} - C_{\text{cur}}|; \end{aligned}$$

(ii) finding the point with maximum value of S on the level curve $C = C_{\text{aim}}$. This is done by following the corresponding contour and numerically interpolating the values of S along it.

2.3.2.4. *Damping factors, bumpers and other checks.* When fitting constraints $\mathbf{U}_H^*$ where certain U values are a substantial fraction of unity, further precautions are necessary to prevent divergent behaviour. These include

(i) the dynamic range of the ω map is kept within the limits $-4{\cdot}5 \le \omega(\mathbf{x}) \le 4{\cdot}5$, which correspond roughly to $10^{-2} \le q(\mathbf{x}) \le 10^2$; these limits may be overridden if desired;

(ii) whenever division by $q(\mathbf{x})$ is to be performed [*e.g.* at step (2.5d') above], the lower limit of the ω map is reset to $-2{\cdot}5$ prior to this division;

(iii) the trial values s_{trial} and t_{trial} are readjusted as a function of the contrast of the corresponding search directions, as follows:

(*a*) the user supplies $s_{\max}$ (default: 0·1) and $t_{\max}$ (default: 0·5); these limits are never exceeded;

(*b*) the value of s_{trial} is obtained as

$$s_{\text{trial}} = \min\left(s_{\max} / \max_{\mathbf{x}} |\omega^{(i)}(\mathbf{x})|, s_{\max}\right);$$

(*c*) the value of t_{trial} is obtained in two steps *via*

$$t_{\text{temp}} = \min\left(t_{\max} / \max_{\mathbf{x}} |\Delta_t \omega^{(i)}(\mathbf{x})|, t_{\max}\right),$$

$$t_{\text{trial}} = \min\left(r / \|\Delta_t \omega^{(i)}\|_E^2, t_{\text{temp}}\right),$$

where $\|\Delta_t \omega^{(i)}\|_E^2$ is the length of $\Delta_t \omega^{(i)}$ in the entropy metric

$$\|\Delta_t \omega^{(i)}\|_E^2 = \int_V |\Delta_t \omega^{(i)}(\mathbf{x})|^2 \, q(\mathbf{x}) \, d^3\mathbf{x} \tag{2.23}$$

and r is a parameter with a default value of 0·04 which may be overridden;

(iv) once the plane search has yielded the desired correction, expressed in terms of ζ's, a temperature factor may be applied to it so as to attenuate its high-frequency component;

(v) the collinearity of the gradients $\nabla \mathscr{S}$ and $\nabla \mathscr{C}$ may be monitored by computing the cosine of the angle between them.

In this form, the plane search algorithm is fast (reaching a solution in less than 20 cycles, and less than 10 cycles for a small basis set) and extremely robust. It does not always produce as good a collinearity of gradients as that described in I, § 7.2, but this has not so far seemed to be a limiting factor in its usefulness.

In the early stages of the calculation, the plane search may be simplified into a more economical *line search* using only the t direction, without loss of robustness. This simpler search, however, does not allow one to follow a level curve of $\mathscr{C}$ to locate the maximum of $\mathscr{S}$ on that curve.

2.4. *Likelihood evaluation*

To test the hypothesis (H_1) that $\mathbf{U}_H = \mathbf{U}_H^*$ against the hypothesis (H_0) that $\mathbf{U}_H = \mathbf{0}$, we use the observed moduli $|\mathbf{U}_K|^{\text{obs}}$ for the collection K of reflexions where the maximum-entropy extrapolation creates the strongest moduli $|\mathbf{U}_K^{\text{ME}}|$; the optimal K for this purpose was shown in I, § 4.2.2 (see also III, § 4, for a detailed derivation) to coincide with the 'second neighbourhood' of H as defined by Hauptman (1980).

Having chosen K in this way, we combine the expressions of § 1.4 for the likelihoods in the diagonal approximation with the use of the Σ parameters introduced in (2.10). This allows a treatment of the measurement errors present in the moduli data $|\mathbf{U}_K|^{\text{obs}}$ and also absorbs the effect of systematic computational errors due for instance to aliasing (which is unavoidable because of the numerous non-linear operations performed on the various Fourier maps).

For $\mathbf{k}$ acentric, we put

$$\Lambda(\mathbf{U}_H^* \,||\, |U_\mathbf{k}| = |U_\mathbf{k}|^{\text{obs}}) = \frac{|U_\mathbf{k}|^{\text{obs}}}{\varepsilon_\mathbf{k}\Sigma_a + \sigma_\mathbf{k}^2} \exp\left\{-\frac{1}{2}\frac{(|U_\mathbf{k}|^{\text{obs}})^2 + |U_\mathbf{k}^{\text{ME}}|^2}{\varepsilon_\mathbf{k}\Sigma_a + \sigma_\mathbf{k}^2}\right\} \times I_0\left(\frac{|U_\mathbf{k}|^{\text{obs}}\,|U_\mathbf{k}^{\text{ME}}|}{\varepsilon_\mathbf{k}\Sigma_a + \sigma_\mathbf{k}^2}\right) \tag{2.24a}$$

while

$$\Lambda(\mathbf{0} \,||\, |U_\mathbf{k}| = |U_\mathbf{k}|^{\text{obs}}) = \frac{|U_\mathbf{k}|^{\text{obs}}}{\varepsilon_\mathbf{k}\Sigma_a + \sigma_\mathbf{k}^2} \exp\left\{-\frac{1}{2}\frac{(|U_\mathbf{k}|^{\text{obs}})^2}{\varepsilon_\mathbf{k}\Sigma_a\sigma_\mathbf{k}^2}\right\}. \tag{2.24b}$$

We refine the Σ_a parameter separately for the two hypotheses by a simple Newton method so as to maximize the global likelihood Λ calculated over all the subset K_a of K consisting of acentric reflexions. The necessary derivatives are calculated in the Appendix. This refinement is carried out separately for the two hypotheses, so that the test uses the log-likelihood

$$L_a = \log \frac{\max_{\Sigma_a} \Lambda(\mathbf{U}_H^* \,||\, |\mathbf{U}_{K_a}| = |\mathbf{U}_{K_a}|^{\text{obs}})}{\max_{\Sigma_a} \Lambda(\mathbf{0} \,||\, |\mathbf{U}_{K_a}| = |\mathbf{U}_{K_a}|^{\text{obs}})}. \tag{2.24c}$$

Similarly, for $\mathbf{k}$ centric,

$$\Lambda(\mathbf{U}_H^* \,||\, |U_\mathbf{k}| = |U_\mathbf{k}|^{\text{obs}}) = \left[\frac{2}{\pi(\varepsilon_\mathbf{k}\Sigma_c + \sigma_\mathbf{k}^2)}\right]^{1/2} \exp\left\{-\frac{1}{2}\frac{(|U_\mathbf{k}|^{\text{obs}})^2 + |U_\mathbf{k}^{\text{ME}}|^2}{\varepsilon_\mathbf{k}\Sigma_c + \sigma_\mathbf{k}^2}\right\} \times \cosh\left(\frac{|U_\mathbf{k}|^{\text{obs}}\,|U_\mathbf{k}^{\text{ME}}|}{\varepsilon_\mathbf{k}\Sigma_c + \sigma_\mathbf{k}^2}\right) \tag{2.25a}$$

while

$$\Lambda(\mathbf{0} \,||\, |U_\mathbf{k}| = |U_\mathbf{k}|^{\text{obs}}) = \left[\frac{2}{\pi(\varepsilon_\mathbf{k}\Sigma_c + \sigma_\mathbf{k}^2)}\right]^{1/2} \exp\left\{-\frac{1}{2}\frac{(|U_\mathbf{k}|^{\text{obs}})^2}{\varepsilon_\mathbf{k}\Sigma_c + \sigma_\mathbf{k}^2}\right\}. \tag{2.25b}$$

Here too the parameter Σ_c is refined separately for the two hypotheses so as to maximize their likelihood, and the test uses the log-likelihood

$$L_c = \log \frac{\max_{\Sigma_c} \Lambda(\mathbf{U}_H^* \,||\, |\mathbf{U}_{K_c}| = |\mathbf{U}_{K_c}|^{\text{obs}})}{\max_{\Sigma_c} \Lambda(\mathbf{0} \,||\, |\mathbf{U}_{K_c}| = |\mathbf{U}_{K_c}|^{\text{obs}})}, \tag{2.25c}$$

where K_c is the subset of centric reflexions in K.

The global criterion for testing (H_0) against (H_1) in view of all the data in K is then

$$L = L_a + L_c. \tag{2.26}$$

2.5. *Likelihood optimization for phase refinement*

After L has been calculated pointwise for several sets of trial structure-factor values $\mathbf{U}_H^*$, it is natural to seek to improve the trial phases by maximizing L with respect to them (§ 1.5). We have implemented a first version of such a procedure, which will now be described.

In the diagonal approximation, the log-likelihood function L for hypothesis (H_1) is a sum of separate contributions from each reflexion $\mathbf{k}$ in the second neighbourhood K. Thus for each $\mathbf{k}$ we can calculate the complex partial derivative

$$\partial L/\partial U_\mathbf{k}^{\text{ME}} = [\partial L/\partial |U_\mathbf{k}^{\text{ME}}|] \exp(i\varphi_\mathbf{k}^{\text{ME}}) \tag{2.27}$$

(full expressions for these derivatives can be found in the Appendix). These numbers are the components of the gradient of L with respect to the components of $\mathbf{U}_K^{\text{ME}}$. To obtain a Newton direction from this gradient we divide the component belonging to each $\mathbf{k}$ by $\varepsilon_\mathbf{k}\Sigma + \sigma_\mathbf{k}^2$ (with $\Sigma = \Sigma_a$ or Σ_c according to $\mathbf{k}$), which is a crude approximation to the curvature $\partial^2 L/\partial |U_\mathbf{k}^{\text{ME}}|^2$. A set of Fourier coefficients is thus obtained, from which a map $(\Delta_L q)(\mathbf{x})$ can be calculated. We then use the standard method of division by q and spectrum truncation to 'pull back' $\Delta_L q$ into a search direction in ω space:

$$\Delta_L \omega = T_H(\Delta_L q / q^{\text{ME}}). \tag{2.28}$$

This direction is in turn converted into suggested changes to the structure-factor values $\mathbf{U}_H^*$ by Fourier analysis according to

$$\Delta \mathbf{U}_H^* = \int_V q^{\text{ME}}(\mathbf{x})(\Delta_L \omega)(\mathbf{x}) \exp(+2\pi i \mathbf{h}\,.\,\mathbf{x})\, \mathrm{d}^3\mathbf{x}. \tag{2.29}$$

The phases of $\mathbf{U}_H^* + w\Delta\mathbf{U}_H^*$ are then applied to the moduli $|\mathbf{U}_H^*|$ (with w a scale factor which may be used to prevent excessive changes), thus completing the phase-refinement process.

3. Strategy

3.1. *Overview*

The practical task of constantly updating the distribution of atoms $q(\mathbf{x})$ to the maximum-entropy distribution compatible with the phase choices made at

each stage requires a book-keeping scheme capable of recording these various phase choices and the parentage relations between them - in other words, of representing all the phasing paths one may wish to explore. For this purpose it is convenient to use a *multisolution tree* (I, §§ 2.4, 8.1; III, pp. 73-75).

3.1.1. *Growth of the multisolution tree.* Let the nodes of this multisolution tree be labelled by the index ν, let H_ν denote the set of 'basis' reflexions (*i.e.* of reflexions for which trial phase values are being treated as free parameters) attached to that node, and let K_ν denote the second neighbourhood of H_ν. The overall strategy then goes as follows (I, § 8.1). The root node (labelled $\nu = 1$) of the tree consists of the set H_1 of origin-fixing and enantiomorph-defining reflexions. The subsequent growth of the tree is governed by the following sequence of computations at each node ν:

(a_ν) update the prior distribution of atoms $q(\mathbf{x})$ to the maximum-entropy distribution $q_\nu^{\mathrm{ME}}(\mathbf{x})$ compatible with the phase choices (giving rise to structure-factor values $\mathbf{U}^*_{H_\nu}$) attached to that node;

(b_ν) construct the conditional distribution $\mathscr{P}^{\mathrm{SP}}_{\mathrm{diag}}(\mathbf{U}_{K_\nu}|\mathbf{U}_{H_\nu} = \mathbf{U}^*_{H_\nu})$ of the yet unphased structure factors in K_ν;

(c_ν) construct the likelihood function $\Lambda(\mathbf{U}^*_{H_\nu}||\mathbf{U}_{K_\nu}| = |\mathbf{U}_{K_\nu}|^{\mathrm{obs}})$ by integrating the conditional distribution over the unknown phases and substituting the observed values of the moduli in K_ν;

(d_ν) refine the basis phases in $\mathbf{U}^*_{H_\nu}$ by maximizing Λ with respect to these basis phases;

(e_ν) solve the local branching problem for (*i.e.* identify the local maxima of) the conditional distribution with respect to a subset of unknown phases when the values of the corresponding moduli are introduced;

(f_ν) expand the current node by creating a branch leading to a new tip node for each of these maxima; if the conditional distribution is too flat for (e_ν) to be meaningful, just carry out some phase permutation; in both instances the current node ν generates a certain number m_ν of daughter nodes $(\nu^{(1)}, \nu^{(2)}, \ldots, \nu^{(m_\nu)})$; each such node has an enlarged basis set $H_{\nu^{(1)}}, H_{\nu^{(2)}}, \ldots, H_{\nu^{(m_\nu)}}$ with associated trial structure-factor values $\mathbf{U}^*_{H_{\nu^{(1)}}}, \mathbf{U}^*_{H_{\nu^{(2)}}}, \ldots, \mathbf{U}^*_{H_{\nu^{(m_\nu)}}}$, and process ($a_\nu$) may be applied to each of them.

3.1.2. *Optimal scheduling of tree growth.* The multisolution tree would quickly grow to an unmanageable size, and its growth must be supervised so as to maximize the chance of finding the correct set of phases without having to develop too many of its spurious branches. For this purpose, we may use two criteria at each node ν:

(i) the loss of entropy $\mathscr{S}_m(q_\nu^{\mathrm{ME}})$ in going from a uniform distribution of atoms $m(\mathbf{x}) = 1/V$ to the current non-uniform $q_\nu^{\mathrm{ME}}(\mathbf{x})$; according to Shannon (Shannon & Weaver, 1949), this measures the shrinkage of the population of 'reasonably probable structures' when the constraints $\mathbf{U}^*_{H_\nu}$ are enforced;

(ii) the likelihood $\Lambda(\mathbf{U}^*_{H_\nu}||\mathbf{U}_{K_\nu}| = |\mathbf{U}_{K_\nu}|^{\mathrm{obs}})$, which measures the degree to which the phase choices made in $\mathbf{U}^*_{H_\nu}$ are able to anticipate correctly, through their maximum-entropy extrapolation, some of the information present in the yet inactive constraints and hence measures the 'chances of survival' of the current node in the face of these forthcoming constraints.

Intuitively, the entropy of a node measures the 'girth' of the branch directly leading to it from the root node, while its likelihood is an estimate of the sum of the girths of the branches bifurcating away from that node once the values of the unphased moduli in its second neighbourhood are taken into account. One would thus expect likelihood (which looks 'ahead' at the degree of pre-established harmony between its predictions and the actual observations for the yet unphased data) and entropy (which looks 'back' at the cost of accommodating the current phase assumptions) to be complementary measures of the worth of a trial phase set.

It is this very analysis which is embodied in the criterion (1.18) defined in § 1.5 on the basis of Bayes's theorem. This combined criterion, calculated with N replaced by $1/\Sigma_c \simeq 1/2\Sigma_a$, can therefore be used as a score to determine the priority with which each node will be considered for further growth in step (f), thus affording an optimal scheduling of the growth of the multisolution tree.

3.2. *Principles of operation*

We will now examine in detail the procedures used in each of the above steps. Programming and practical considerations are dealt with by Gilmore, Bricogne & Bannister (1990).

3.2.1. *Origin and enantiomorph definition.* The direct-methods program *MITHRIL* (Gilmore, 1984; Gilmore & Brown, 1988) is used to select those reflexions which optimally define the origin and enantiomorph. We have however found that the present phasing strategy is much more effective if an extra resolution criterion is used (in conjunction with the usual $|E|$-magnitude criterion) to choose these reflexions: if relatively strong reflexions are present at low resolution, even though they may not have the strongest $|E|$ values, they should be incorporated preferentially into the initial basis set H_1. A plausible explanation of this behaviour is that the conditional distributions of the phases of strong high-resolution reflexions are less multimodal (hence make the branching problem more tractable) if the non-uniformity in $q(\mathbf{x})$ has been created by low-resolution constraints rather than by other high-resolution terms (see § 1.2.2).

3.2.2. *Updating to a maximum-entropy distribution.* Step (*a*) is carried out by the method described in § 2.3, involving bicubic modelling and plane search. We have found this algorithm extremely robust and capable of dealing with U values close to unity.

3.2.3. *Conditional distributions.* At step (*b*), the conditional distributions of the structure factors in K_ν (in the diagonal approximation) are calculated by the explicit formulae given in § 1.2.4. Centroid values for these structure factors may be calculated as described in § 1.6 and used to calculate a map for visual inspection.

3.2.4. *Likelihood functions, Σ refinement.* Step (*c*) uses the expressions (2.24), (2.25) and (2.26) derived in § 2.4 for the likelihood functions in the diagonal approximation. The second neighbourhood K_ν of the current basis set H_ν is generated and the values of Σ_a and Σ_c are refined by a Newton method. A weighted average of $1/\Sigma_c$ and $1/2\Sigma_a$ is subsequently used as a substitute for the number of atoms N in all formulae involving N.

We have found the likelihood criterion to be of immense power in ranking trial phase sets, even at very early stages when none of the existing figures of merit used in traditional direct methods would give any useful indications. As phasing progresses, the correct node rapidly emerges above the others, its likelihood being orders of magnitudes greater.

3.2.5. *Phase refinement.* Step (*d*) is based on the principle outlined in § 1.5 and is carried out by the algorithm described in § 2.5.

This new method of phase refinement has proved extremely effective, even with very small basis sets which would in no way lend themselves to a tangent refinement.

3.2.6. *Survey of branching behaviour.* As explained in § 1.2.4, the diagonal approximation destroys the multimodality of conditional distributions, and thus the expressions set up at step (*b*) are unsuitable to carry out step (*e*). For the latter purpose, we use the real-space expression (1.4) of the full conditional distribution for the structure factors.

In order to survey this c.p.d. with respect to as many new phases as possible, we use the magic-integer technique (White & Woolfson, 1975; Main, 1977) to encode several phases into fewer numerical symbols. The integral $\int_V \{[\delta q(\mathbf{x})]^2/q^{\mathrm{ME}}(\mathbf{x})\}\, \mathrm{d}^3\mathbf{x}$ is calculated for each point of a grid of simultaneous values of these symbols, and its minima (corresponding to the maxima of the c.p.d.) are located. Absence of contrast at this stage is usually a sign that the current basis set is a bad choice, and that backtracking to another node of the tree is advisable. If no better node can be identified, simple-minded phase permutation is used to create the daughter nodes, whose ranking must then await the evaluation of the likelihood criterion at the next round.

This method of surveying the multimodality of the c.p.d. has proved an effective and reliable one, and an essential aid in controlling the phasing process.

3.2.7. *Expansion of the basis set.* At steps (*e*) and (*f*), it is of vital importance to exercise great care in choosing the set of new reflexions with respect to whose phases the branching behaviour will be surveyed. Choosing those reflexions where the strongest ME extrapolation takes place always leads to an impasse, undoubtedly because the phasing process becomes over-consistent on a strongly coupled subset of the data, and is then unable to incorporate any of the information present in the rest of the observations (this is a well known phenomenon in tangent refinement). We have found that the correct strategy for expanding the basis set consists of choosing those strong reflexions for which no strong extrapolation has yet taken place, *i.e.* whose strength is maximally unexpected to the current statistical model. For equal degrees of unexpected strength, the preferential choice of low-resolution *vs* high-resolution reflexions is helpful, for the same reason as explained in §§ 1.2.2 and 3.2.0.

3.3. *Meaning of Σ*

The use of an empirical and refinable parameter Σ (calculated as the weighted average of $1/\Sigma_c$ and $1/2\Sigma_a$) instead of the number of atoms N deserves further comment. As defined in § 2.3.2 and used in § 2.4, Σ measures the statistical dispersion of the distribution of the observed moduli, taking into account the modulation of the latter by ME extrapolation from the basis set of reflexions (§ 1.3).

If $q(\mathbf{x})$ is the uniform distribution, Σ measures the dispersion of the unphased $|E|$ values around their root-mean-square value of 1·0. If the atoms in the structure at hand are not truly statistically independent at the working resolution but occur as 'globs' (Harker, 1953), then the refined value of Σ will be *greater* than $1/N$, being roughly the inverse of the number of such globs rather than of the number of atoms. This phenomenon has actually been observed in the course of the direct phasing of contrast variation data from a protein crystal at low resolution, and the size of the globs thus indicated correlates well with the existence of a sharp maximum in the radial intensity distribution at about 10·5 Å resolution (Carter, Crumley, Coleman, Hage & Bricogne, 1990). The quantity $1/\Sigma$ is thus N_{eff}, the 'effective N' defined in I, § 8.3, and this observation corroborates the conclusion reached there that the potential strength of statistical phase relations should be gauged from the value of N_{eff} rather than of N. Our practice of using Σ rather than N guarantees that this effect is automatically taken into account.

At the other extreme, when a great deal of correct phase information has accumulated at certain nodes, the ME extrapolation becomes more and more exact, as the Toeplitz determinants approach zero (Goedkoop, 1950). In these circumstances the parameter Σ refines to values much *smaller* than $1/N$, but as the U values are not renormalized the effect is to make the conditional distributions much sharper, and phase extension much stronger. We have repeatedly observed this 'critical behaviour' of Σ, which is accompanied by soaring values of the log-likelihood of the correct node, to be an indication that the phasing process has reached successful completion.

Summary

We have examined in detail a previously proposed multisolution strategy for direct phase determination by combined maximization of entropy and of likelihood, as well as the theoretical results and the computer algorithms on which its implementation is based. Increasingly specific phase assumptions are made as part of a tree-directed search. For each such assumption, saddlepoint approximations to joint and conditional probabilities of structure factors are calculated by means of maximum-entropy distributions of atoms which are obtained by a robust plane-search algorithm. A likelihood criterion is constructed for testing phase assumptions in the light of the observed distribution of the yet unphased moduli. Different trial phase sets are ranked according to a criterion combining entropy and likelihood according to Bayes's theorem, and the phases are refined by optimizing this criterion. Finally, the ambiguity inherent in the process of phase extension (the 'branching problem') is explicitly surveyed by evaluation of a real-space expression which has a simple interpretation in terms of a differential cost for the creation of new detail, and the search tree is expanded accordingly. This procedure works best if initiated at low resolution and if allowed to progress to higher resolution only gradually, and this behaviour can be rationalized within the proposed theory. Practical applications are described by Gilmore, Bricogne & Bannister (1990).

The collaborative effort reported here was initiated at LURE thanks to a Senior Ciba-Geigy Fellowship (to CJG); subsequent work in Glasgow was supported by a grant from the SERC (to CJG). GB acknowledges support from CNRS (France) and partial support from Trinity College and from the MRC Laboratory of Molecular Biology (Cambridge, England) during the course of this work.

APPENDIX

We will collect here the expressions for the derivatives of the log-likelihood contributions of individual reflexions in the diagonal approximation which are used in the Newton refinement of the Σ parameter (§§ 2.4, 3.2.3) and in constructing the search direction for phase refinement (§ 2.5).

A.1. *Acentric case*

Put $r = |U_{\mathbf{k}}^{\mathrm{ME}}|$, $R = |U_{\mathbf{k}}|^{\mathrm{obs}}$, $\Sigma = \varepsilon_{\mathbf{k}}\Sigma_a + \sigma_{\mathbf{k}}^2$, $\varepsilon = \varepsilon_{\mathbf{k}}$ to simplify the notation. The log-likelihood contribution is (§ 2.4)

$$L_a(r, R, \Sigma) = \log R - \log \Sigma - \tfrac{1}{2}(r^2 + R^2)/\Sigma + \log I_0(rR/\Sigma). \qquad (A1)$$

It is convenient to write τ_n for the quotient I_n/I_0 of modified Bessel functions. If differentiation is denoted by a prime ($'$), the recurrence relations for Bessel functions then give the relations

$$(\log I_0)'(z) = \tau_1(z), \quad \tau_1'(z) = 1 - \tau_1(z)/z - [\tau_1(z)]^2. \qquad (A2)$$

We may then calculate the required derivatives as

$$\frac{\partial L_a}{\partial r} = -\frac{r}{\Sigma} + \frac{R}{\Sigma}\tau_1\left(\frac{rR}{\Sigma}\right), \qquad (A3a)$$

$$\frac{\partial L_a}{\partial \Sigma_a} = -\frac{\varepsilon}{\Sigma} + \frac{\varepsilon}{2}\frac{r^2+R^2}{\Sigma^2} - \frac{\varepsilon rR}{\Sigma^2}\tau_1\left(\frac{rR}{\Sigma}\right), \qquad (A3b)$$

$$\frac{\partial^2 L_a}{\partial \Sigma_a^2} = +\frac{\varepsilon^2}{\Sigma^2} - \varepsilon^2\frac{r^2+R^2}{\Sigma^3} + \left(\frac{\varepsilon rR}{\Sigma^2}\right)^2 \tau_1'\left(\frac{rR}{\Sigma}\right) + \frac{2\varepsilon^2 rR}{\Sigma^3}\tau_1\left(\frac{rR}{\Sigma}\right). \qquad (A3c)$$

The first of these is involved in the gradient of § 2.5, while the latter two determine the shift $\Delta\Sigma_a$ of Σ_a in a Newton refinement step *via*

$$\Delta\Sigma_a = -\sum_{\substack{\mathbf{k}\in K \\ \mathbf{k}\ \text{acentric}}} \left(\frac{\partial L_a}{\partial \Sigma_a}\right)_{\mathbf{k}} \Bigg/ \sum_{\substack{\mathbf{k}\in K \\ \mathbf{k}\ \text{acentric}}} \left(\frac{\partial^2 L_a}{\partial \Sigma_a^2}\right)_{\mathbf{k}}. \qquad (A4)$$

Should this shift result in a negative value of Σ_a, it can be converted into a shift of $\log \Sigma_a$ so as to preserve positivity (K. Henderson, personal communication).

A.2. *Centric case*

With the same abbreviations as before, but with $\Sigma = \varepsilon_{\mathbf{k}}\Sigma_c + \sigma_{\mathbf{k}}^2$, the log-likelihood contribution is (§ 2.4)

$$L_c(r, R, \Sigma) = \frac{1}{2}\log\frac{2}{\pi} - \frac{1}{2}\log\Sigma - \frac{1}{2}\frac{r^2+R^2}{\Sigma} + \log\cosh\left(\frac{rR}{\Sigma}\right). \qquad (A5)$$

Writing $\tau(z)$ for the function $\tanh(z)$, we have

$$(\log\cosh)'(z) = \tau(z), \quad \tau'(z) = 1 - [\tau(z)]^2, \qquad (A6)$$

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