Economy of symbols *versus* clarity

Although the specification of an excessive number of symmetry elements in a symbol of Hermann–Mauguin type can be confusing, reduction of the number to an absolute minimum can be mystifying. The symbols proposed, although reasonably concise, are therefore not claimed to have been condensed to the maximum possible extent.

References

BROWN, H., BÜLOW, R., NEUBÜSER, J., WONDRATSCHEK, H. & ZASSENHAUS, H. (1978). *Crystallographic Groups of 4-Dimensional Space.* New York: Wiley.

WHITTAKER, E. J. W. (1983). *Acta Cryst.* A**39**, 123–130.

WHITTAKER, E. J. W. (1984*a*) *Acta Cryst.* A**40**, 58–66; erratum, 484.

WHITTAKER, E. J. W. (1984*b*). *An Atlas of Hyperstereograms of the 4-Dimensional Crystal Classes.* Oxford Univ. Press. In the press.

Acta Cryst. (1984). A**40**, 410–445

Maximum Entropy and the Foundations of Direct Methods*

BY G. BRICOGNE†

Department of Biochemistry, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, NY 10032, USA

(*Received* 18 *March* 1983; *accepted* 10 *January* 1984)

Abstract

A revision of the classical statistical methods of phase determination is presented which widens their theoretical foundations and consolidates their practical implementation, thus bringing about a major increase of their power. In a brief introductory survey (§ 1), the basic concepts and mathematical techniques of direct methods are analysed. Closer scrutiny (§ 2) reveals that severe inadequacies still impair the effectiveness of these methods. The asymptotic character of the series used to approximate joint distributions of structure factors demands that great caution be exercised to guarantee their accuracy, and this requirement can only be fulfilled if they are used within a multisolution algorithm in which the prior distribution of atoms is constantly updated so as to incorporate at every stage all the phase information assumed to that point. Further limitations follow from the traditional practice of approximating joint distributions by products of marginal distributions of single invariants. A scheme for simultaneously overcoming both difficulties is then proposed. The pivotal element of this scheme is a device, based on Jaynes's maximum-entropy principle, for exploiting the prior knowledge of some structure factors in the construction of the joint distributions of others conditional to that knowledge. Jaynes's maximum-entropy formalism is presented and systematically applied to the construction of the requisite non-uniform prior distributions of atoms in § 3. The problem of effectively approximating conditional distributions of very large numbers of structure factors is solved in § 4 by a novel technique of 'maximum-entropy inversion' of Karle–Hauptman matrices, and the result obtained is shown to generalize the most sophisticated probabilistic formulae hitherto obtained. This procedure is proved in § 5 to coincide with an enhancement of the standard method of asymptotic expansions by means of Daniels's saddlepoint approximation. Its relationship to determinantal methods is investigated in § 6. A numerical algorithm for implementing these ideas is presented in § 7, together with an application to data from the small protein Crambin, and a unified strategy for its use *ab initio* is described and discussed in § 8. It is concluded that the phase-determination strategy proposed here will expedite the realization of the full potential of probabilistic direct methods, and is likely to bring macromolecular structures within their reach.

Introduction

Thirty years ago Hauptman and Karle pioneered the use of sophisticated methods of probability theory for directly determining the phases of structure factors from the sole knowledge of their amplitudes (Hauptman & Karle, 1953). After an initial latency period, these probabilistic direct methods underwent a

* *Editorial Note*: Papers exceeding the normal length limitations of the journal are scrutinized particularly carefully to ensure they meet the stated goal of providing the maximum density of information consistent with clarity of presentation. Considerable reductions in length are often achievable in revision without loss of essential information. This very long paper, having passed all normal editorial procedures, constitutes a rare exception to the normal upper bound.

† Present address: LURE, Bâtiment 209C, 91405 Orsay CEDEX, France.

0108-7673/84/040410-36$01.50

remarkable development between the mid-sixties and the mid-seventies, with the advent of the symbolic addition procedure (Karle & Karle, 1966) and of multisolution strategies (Germain & Woolfson, 1968) for the systematic exploitation of the triple phase relationship. These advances resulted in computer programs capable of routinely solving most organic structures containing up to 100 non-hydrogen atoms. During the last decade, steady progress has been made in the further use of triplets, quartets and quintets, and in the derivation of more elaborate formulae for estimating these invariants by means of their neighbourhoods (Hauptman) or phasing shells (Giacovazzo). A review of these topics may be found in Giacovazzo (1980). In spite of these improvements the solution by direct methods of structures containing over 150 non-hydrogen atoms remains the exception rather than the rule. It seems, however, difficult at present to ascertain whether this upper bound represents a fundamental limit in the intrinsic power of the probabilistic approach, or is merely a consequence of some temporary imperfections in its current implementation.

The present paper is an attempt at settling this question in favour of the second alternative through a reappraisal of the mathematical apparatus of probabilistic direct methods and of the procedures by which it is put to use. This study will bring to light an intimate connection between direct methods and a mode of statistical inference based on Jaynes's maximum-entropy (ME) principle (Jaynes, 1957, 1968).

A number of papers have recently appeared on the maximum-entropy approach to direct phase determination. Piro (1983) considers the relation between direct methods and a form of the ME method already studied by Britten & Collins (1982) and Narayan & Nityananda (1982), which is based on a different entropy called here the 'Burg entropy'. The article by Wilkins, Varghese & Lehmann (1983) contains an exposition of Jaynes's ideas on the relevance of his ME principle to all inverse problems, along lines indicated independently by the present author (Bricogne, 1982). Wilkins *et al.* advocate its use as a *de novo* approach to the phase problem, stating without proof that it 'subsumes all currently practised numerical approaches to the crystallographic inversion problem as special cases'. This unsubstantiated claim is partially refuted by the present study: the precise relation between the two methods is shown to be one of complete logical equivalence, whose proof is anything but obvious, and the only difference lies in the quality of their practical implementations. In this respect, all the currently available entropy maximization algorithms are still far inferior to direct methods in that they fail to deal adequately with one of the major difficulties of phase determination, called 'branching' in this article.

The fundamental problem of direct phase determination is the construction of joint or conditional probability distributions (j.p.d.'s or c.p.d.'s) of structure factors. The formal device of the Edgeworth asymptotic series proposed in the early literature, and purported to afford a general solution to this problem (§ 1), is critically re-examined. It is shown to be incorrectly utilized and hence to yield unreliable approximations to conditional distributions when structure factors with large amplitudes are used as arguments. Further loss of phasing power results from the current practice of exploiting large numbers of 'small-base' distributions (involving few structure factors) rather than a single large-base distribution, since the intricate pattern of interactions between phases is only very partially represented (§ 2). To correct these imperfections, a better method must be sought for approximating large-base joint distributions, which will lead to consistently accurate estimates of conditional distributions derived from them. For this purpose, one is compelled by the restricted domain of validity of the Edgeworth asymptotic expansion to adopt a new kind of multisolution approach in which the prior distribution of atoms is constantly updated so as to recentre the asymptotic expansions of the c.p.d.'s around the trial structure-factor values assumed at each stage (§ 2). Jaynes's maximum-entropy principle and the ensuing formalism provide both the criterion necessary to select uniquely such a non-uniform prior distribution, and the analytical methods for deriving it in closed form (§ 3). The effective construction of accurate conditional distributions of very large numbers of structure factors is considered in § 4. A general method is obtained, by which all previously known formulae for estimating phase invariants may readily be derived as approximations and extended, and whose implementation can be made computationally efficient. The uniqueness and optimality of this scheme is then proved in § 5 by establishing its equivalence with a multidimensional version of Daniels's saddlepoint method (Daniels, 1954). The approximate conditional distributions thus obtained bear a close relationship to those afforded by the maximum-determinant method (MDM) of Tsoucaris (1970), but the present procedure has considerable advantages over the latter (§ 6). A fundamental computational process in the implementation of these ideas is the constrained maximization of an entropy functional. An algorithm designed for this purpose, which embodies a complete solution to the branching problem and has been used successfully to carry out phase extension on a small protein, is presented in § 7. Finally, the overall organization of the new multisolution algorithm, its impact on the powers of direct methods, and its potential usefulness in *ab initio* phase determination for large structures are discussed in § 8.

1. Survey of the classical formalism of direct methods

The recalcitrant nature of the phase problem derives from the fact that it is mathematically indeterminate unless sufficient chemical information is brought to bear on its solution to constrain it down to a reasonably unique answer, while the full content of this chemical information cannot be adequately captured by any of the well-developed devices of mathematics until a stereochemical parametrization of the electron density becomes possible.

The statistical theories of the phase problem endeavour to exploit some of the statistical consequences of the basic premise of chemistry – the 'atomic hypothesis'. This property of atomicity is abstracted to the statement that crystal structures may be viewed as consisting of identical atoms placed randomly and independently of each other in accord with the crystal periodicity and symmetry. The Fourier coefficients pertaining to such structures are therefore sums of a large number of random variables (the contributions from individual atoms), and one may invoke the limit theorems of probability theory to obtain estimates of their distribution.

These ideas were first put to use by Wilson (1949) in his study of the distribution of diffracted intensities and of its modulation by crystal symmetry. The next logical step was taken by Hauptman & Karle (1953), who examined the joint distribution of several Fourier coefficients, and made the crucial observation that such distributions could yield phase information if the amplitudes were assumed to be known. Bertaut (1955*a*, *b*, *c*) and Klug (1958) later re-examined their initial formulation, and showed that its mathematical foundations were to be found in the theory of normal approximation by means of the Gram–Charlier or Edgeworth asymptotic expansions.

The mathematical apparatus of direct methods thus consists of three main elements:

(*a*) a random process attempting to mimic the rules of chemistry through the attribute of 'atomicity';

(*b*) a probabilistic formalism capable of deriving conditional distributions of phases for given amplitudes which describe the redundancy relations between structure factors implied by the hypothesis of atomicity;

(*c*) an algebraic formalism which supervises the use of probabilistic formulae on suitably chosen clusters of reflexions, and collates the information contained in the resulting phase probability distributions.

These three constituents will now be examined individually in some detail.

1.1. *Stochastic model*

Since their initial formulation by Hauptman & Karle, direct methods have used as their starting point a stochastic process which generates crystal structures by randomly placing atoms, independently of each other, in the asymetric unit of the crystal, according to some prior distribution. The atoms are often assumed to be identical, and their prior distribution uniform, but these restrictions are not essential (Klug, 1958).

In virtue of Shannon's sampling theorem (Shannon, 1949), this picture may be discretized at any finite resolution by partitioning the asymmetric unit of the crystal into a finite number B of equal boxes. If the boxes are labelled 1 to B, the stochastic process which places atoms randomly may be viewed abstractly as a discrete source of symbols (the labels of the boxes) in the sense of communication theory (Shannon & Weaver, 1949). The prior distribution of atoms specifies that the B symbols of the source have *a priori* probabilities $q_1, q_2, \ldots, q_B$. Under these circumstances, the stochastic model is one of independent Bernoulli trials with B possible outcomes at each trial: outcome i occurs with probability q_i and causes the current atom to be placed in box i.

This model is admittedly a rather crude representation of chemistry, yet it has the merit of leading to tractable mathematics. A Markov process with finite memory would give a better approximation of the statistical character of a source of atomic positions governed by the laws of stereochemistry, but its exploitation would entail much greater effort.

1.2. *Probabilistic formalism*

Let $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$ be the indices of a prescribed finite set of structure factors, whose joint distribution is being sought.

A point atom of unit weight placed at random position $\mathbf{x}$ gives rise to a random vector in $\mathbb{C}^n$,

$$\mathbf{X}(\mathbf{x}) = \begin{bmatrix} \xi_1(\mathbf{x}) \\ \vdots \\ \xi_i(\mathbf{x}) \\ \vdots \\ \xi_n(\mathbf{x}) \end{bmatrix}, \tag{1.1}$$

of complex contributions to the Fourier coefficients at $\mathbf{h}_1, \ldots, \mathbf{h}_n$. For instance, in space group $P1$,

$$\mathbf{X}(\mathbf{x}) = \begin{bmatrix} \exp(2\pi i \mathbf{h}_1 \cdot \mathbf{x}) \\ \vdots \\ \exp(2\pi i \mathbf{h}_n \cdot \mathbf{x}) \end{bmatrix}. \tag{1.2}$$

By the assumptions of the stochastic model, the vector

$$\mathbf{F} = \begin{bmatrix} F_{\mathbf{h}_1} \\ \vdots \\ F_{\mathbf{h}_n} \end{bmatrix} \tag{1.3}$$

of structure factors for an N-atom structure will be

obtained by summing N independent copies of random vector $\mathbf{X}$:

$$\mathbf{F} = \sum_{i=1}^{N} \mathbf{X}_i. \tag{1.4}$$

Geometrically, each $\mathbf{X}_i$ may be viewed as an elementary step (or increment) of a random flight in $\mathbb{C}^n$, so that reckoning the joint distribution of $F_{\mathbf{h}_1}, \ldots, F_{\mathbf{h}_n}$ amounts to estimating the relative frequency of random flights starting at $\mathbf{0}$ and ending at $\mathbf{F}$ amongst all the N-step flights generated by the stochastic process.

1.2.1. *Exact expression of the joint distribution*

The first aim of probabilistic direct methods is to obtain an asymptotic estimate, in the limit of large N, of the distribution $\mathscr{P}(\mathbf{F})$ of vector $\mathbf{F}$, *i.e.* of the joint distribution of $F_{\mathbf{h}_1}, \ldots, F_{\mathbf{h}_n}$. Standard methods are available in probability theory to obtain such estimates, whose steps will now be outlined for further reference. The reader is referred to Klug (1958) for a detailed introduction to characteristic, moment-generating and cumulant-generating functions.

Let the atoms be placed at random with a prior distribution $q(\mathbf{x})$. Then the induced distribution $P(\mathbf{X})$ for random vector $\mathbf{X}$ is the singular measure P defined by

$$\int_{\mathbb{R}^{2n}} P(\mathbf{X}) G(\mathbf{X}) \, \mathrm{d}^{2n}\mathbf{X} = \int_V q(\mathbf{x}) G[\xi_1(\mathbf{x}), \ldots, \xi_n(\mathbf{x})] \, \mathrm{d}^3\mathbf{x} \tag{1.5}$$

for any function G of an argument in $\mathbb{C}^n$. In particular, all the moments and cumulants of P may be calculated by this formula.

Since the random vectors $\mathbf{X}_i$ are identically distributed and independent, their addition gives rise to the convolution of their distributions:

$$\mathscr{P} = P^{*N} \tag{1.6}$$

hence to the multiplication of their characteristic functions ($C = \bar{\mathscr{F}}P$, $\mathscr{C} = \bar{\mathscr{F}}\mathscr{P}$):

$$\mathscr{C} = C^N \tag{1.7}$$

or moment-generating functions [$M(\mathbf{u}) = C(-i\mathbf{u})$, $\mathscr{M}(\mathbf{u}) = \mathscr{C}(-i\mathbf{u})$]:

$$\mathscr{M} = M^N, \tag{1.8}$$

which is equivalent to the addition of their cumulant-generating functions ($K = \log M$, $\mathscr{K} = \log \mathscr{M}$):

$$\mathscr{K} = NK. \tag{1.9}$$

The desired probability density $\mathscr{P}$ may thus be written

$$\mathscr{P} = \bar{\mathscr{F}}^{-1}[C^N] = \bar{\mathscr{F}}_{\mathbf{u}}^{-1}[M^N(i\mathbf{u})] = \bar{\mathscr{F}}_{\mathbf{u}}^{-1}\{\exp[NK(i\mathbf{u})]. \tag{1.10}$$

1.2.2. *The Edgeworth asymptotic series*

The necessary Fourier inversion is usually not possible in closed form, so that some approximation scheme is needed. For this purpose, it is customary to expand the cumulant-generating function around the origin in the space of the carrying variables $\mathbf{u}$. Recentring $\mathscr{P}$ around its vector of first moments $\mathbf{F}^0 = N\mathbf{U}_q$ [where $\mathbf{U}_q$ is the vector of Fourier coefficients of the prior distribution of atoms $q(\mathbf{x})$] causes the first-order cumulants to vanish, and the second-order terms may be grouped separately from the terms of third or higher order. Going back to characteristic functions gives

$$\mathscr{C} = (\text{Gaussian}) \times \exp\{\text{power series starting with terms of order} \geq 3\}.$$

Expanding the exponential expresses $\mathscr{C}$ as a series of terms of the form

(Gaussian) × monomial in the carrying variables.

Each of these terms may now be subjected to Fourier inversion, giving rise to a Hermite function in the centrosymmetric case (Klug, 1958) or a Laguerre function in the non-centrosymmetric case (Naya, Nitta & Oda, 1965). One thus obtains an expansion of $\mathscr{P}(\mathbf{F})$, asymptotic in powers of $N^{-1/2}$ and valid in a neighbourhood of $\mathbf{F}^0$, called the Edgeworth series of $\mathscr{P}(\mathbf{F})$.

The book-keeping of terms in this series is rather complex. Under the usual assumption of a uniform prior distribution $q(\mathbf{x}) = 1/V$, the structures of the moment- (or cumulant-) generating functions are rather sparse. This is easily seen by considering, for example, the expression for the general moment in $P1$:

$$m_{j_1 j_2 \ldots j_n} = \int_V q(\mathbf{x}) \prod_{l=1}^{n} [\exp(2\pi i \mathbf{h}_l . \mathbf{x})]^{j_l} \, \mathrm{d}^3\mathbf{x}, \tag{1.11}$$

which vanishes unless $\sum_{l=1}^{n} j_l \mathbf{h}_l = \mathbf{0}$. As a result, the j.p.d. of $F_{\mathbf{h}_1}, \ldots, F_{\mathbf{h}_n}$ depends only on those combinations of phases $\sum_{l=1}^{n} j_l \varphi(\mathbf{h}_l)$ whose coefficients verify the above relation, the so-called 'phase invariants'. In a general space group, the pattern of non-vanishing moments is richer, and is associated with dependence relations with integral coefficients amongst the indices $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n$ and their images under the point-group operations. If the prior distribution of atoms $q(\mathbf{x})$ is not uniform, however, the generic moments will not vanish, but will be expressible in terms of the Fourier coefficients of $q(\mathbf{x})$ *via* the structure-factor algebra of the space group (Bertaut, 1955*c*). Such generic moments will be typically about $\sqrt{N}$ times weaker than the special ones considered above, but there will be a great number of them.

1.2.3. *Exploitation of joint distributions*

Joint probability distributions of structure factors are not used as such but serve as intermediates in the

construction of *conditional* distributions, which are derived from the j.p.d.'s by 'specializing' some of their arguments to known or assumed values. A typical practice is to substitute into the j.p.d. the observed values of the structure-factor moduli to obtain the conditional distribution of their phases. Such conditional distributions are the key instrument in the process of phase determination, since they enable one to make inferences about phases from the sole knowledge of moduli.

The formal complexity of the Edgeworth series has so far hindered its full-scale use for constructing j.p.d.'s of large numbers of structure factors. The practical implementations of direct methods have instead evolved around the exploitation of small-base j.p.d.'s involving only the 'basis' reflexions giving rise to a single phase invariant; the conditional distribution obtained by specializing the moduli to their observed values is then the probability distribution of that invariant. Estimates obtained from such 'minimal' j.p.d.'s are of limited reliability, and a more elaborate procedure was devised by Hauptman (1975*a*, *b*): one first constructs a j.p.d. involving, besides the same basis reflexions, a set of suitably chosen satellite reflexions; the conditional distribution obtained after substituting into this 'augmented' j.p.d. the observed moduli of all structure factors concerned is then integrated with respect to the satellite phases. By this mechanism the information contained in the moduli of the satellite structure factors is brought to bear on the estimation of the invariant, whose reliability is thereby increased.

Formally, both types of small-base j.p.d.'s, minimal and augmented, are in effect *marginal* distributions of the complete j.p.d. $\mathscr{P}$ of all structure factors (an object so far unamenable to actual construction), obtained from the latter by integrating over the values of all structure factors but those originally used as arguments. This viewpoint, however artificial it may seem at this stage, will be useful in § 2.2.

1.3. *Algebraic formalism*

The task of collating the information contained in the multitude of probability distributions of individual phase invariants compiled by the current implementations of direct methods gives rise to intricate algebraic constructs, due mainly to Hauptman and Giacovazzo. Hauptman's nested neighbourhoods (or Giacovazzo's phasing shells) classify the sets of satellite reflexions suitable for constructing the augmented j.p.d.'s mentioned in § 1.2.3, while Giacovazzo's theory of representations exploits the rich patterns of relations connecting phase invariants and seminvariants to test further the consistency of their estimates. The purpose of this formalism is therefore to describe the structure of those clusters of reflexions whose amplitudes eventually influence the distribution of a given phase invariant, and to organize systematically the process of cross checking and combining the resulting estimates.

This book-keeping of phase invariants is analogous to that which would be involved in collecting the terms of the Edgeworth series to build the j.p.d. of all structure factors under consideration, were this operation attempted. The Hauptman–Giacovazzo formalism thus performs, externally to the conventional (small-base) probability calculations, the same book-keeping tasks which would remain internal to the construction of large-base j.p.d.'s.

2. Assessment and consolidation of the current methodology

It emerges from the above survey that the central problem of direct methods is the construction of sufficiently accurate approximations to joint probability distributions of large numbers of structure factors. The procedures in current use will now be subjected to closer scrutiny and will be shown to fall short of accomplishing this task in the best possible fashion. Two main infelicities are responsible for this lack of optimality: the non-convergence of the Edgeworth series, and the degradation of j.p.d.'s as they are reconstructed from their marginal distributions. These two difficulties will be analysed in some detail, and a scheme which overcomes them both simultaneously will be proposed.

2.1. *Asymptotic expansions and the problem of large deviations*

The Edgeworth series is asymptotic in powers of $N^{-1/2}$, but for a given N it is not convergent. A claim to the contrary found in the early literature (Hauptman & Karle, 1954) is erroneous. As with all asymptotic series, the error is of the order of the first term neglected, so that such a series should be summed only as far as its smallest term: summing more terms will degrade rather than improve the accuracy of the result. The high-order expansions compiled by Klug (1958) and by Naya, Nitta & Oda (1965) are therefore largely of formal interest. As one attempts to use an asymptotic expansion further away from its centre, the magnitude of the error term increases, making the corresponding estimates unreliable. This troublesome behaviour of asymptotic series for sizeable deviations from their centre demands that they should be used with much care if the results are to be trusted.

The time-honoured custom of using a uniform prior distribution of atoms leads to asymptotic expansions centred around $\mathbf{F}^0 = \mathbf{0}$. These will give good approximations of j.p.d.'s for small structure factors, but if structure factors with large magnitudes are involved the c.p.d.'s of their phases obtained by specialization

(§ 1.2.3) may be totally unreliable. For example, a c.p.d. for the phases of seven magnitudes with $|E| \geq 3{\cdot}0$ will consult the values of the Edgeworth series at distances from its centre in excess of eight times the width of the leading Gaussian term, where all accuracy is illusory.

The heart of the matter is clearly that *the best asymptotic expansion of a c.p.d. cannot in general be obtained by specialization of a pre-existing best asymptotic expansion of the initial j.p.d.*: this procedure is legitimate only if the j.p.d. was expanded sufficiently near the point described by the information used in the specialization, so that the qualifier 'best' may retain its validity after specialization.

The difficulty just diagnosed is a consequence of the fact that all current direct methods commit themselves rigidly at the outset to an Edgeworth series centred at $\mathbf{F}^0 = \mathbf{0}$. This practice gives rise to the ironical situation in which the approximate j.p.d.'s are most accurate where least useful, and least accurate where potentially most useful.

2.2. *Limitations due to the use of marginal distributions*

According to the analysis of §§ 1.2.3 and 1.3, current methods do not undertake to approximate large-base j.p.d.'s directly, but rather by piecing together marginal distributions of single invariants. This operation is comparable to the reconstruction of a high-dimensional object from low-dimensional projections, which requires that a large number of well-determined projections be available.

Unfortunately, the latter requirement is impossible to satisfy. To obtain good estimates of individual invariants, large neighbourhoods must be used, which will unavoidably overlap if numerous invariants are involved. Let $\mathbf{h}$ be a satellite reflexion common to the neighbourhoods of n invariants. Then the information contained in $|F_{\mathbf{h}}|$ will be incorporated into all n estimates separately, and hence will carry an incorrect weight in the final result; at the same time, correlations between these estimates will be lost since $\varphi_{\mathbf{h}}$ will be integrated out of the n augmented distributions (§ 1.2.3) as if it were a distinct variable in each of them.

The intrinsic accuracy of this reconstruction procedure is therefore limited. Furthermore, all the augmented j.p.d.'s used in estimating the invariants are inaccurate in the first place (§ 2.1).

2.3. *A robust approximation scheme*

In the light of the above findings, it appears that current direct methods have attained their already remarkable power in spite of a rather imperfect implementation of the concept of joint probability distribution, based on mathematical procedures which overlook two serious difficulties. It is thus clear that the full potential of Hauptman & Karle's original vision is still largely unrealized, and that the intrinsic power of their probabilistic approach is greatly underestimated.

The limitations inherent in the reconstruction of a j.p.d. or c.p.d. from marginal distributions indicate that a solution must be sought in the form of a better analytical method for directly approximating large-base distributions. A large-scale implementation of the Edgeworth series derived from a uniform prior distribution of atoms is inappropriate for this purpose, since it is expanded around a centre ($\mathbf{F}^0 = \mathbf{0}$) equally distant from all points of the locus defined by the phase circles. A more versatile approximation scheme is therefore necessary.

2.3.1. *Recentring of asymptotic expansions*

If a number of reflexions have large amplitudes, then their structure factors will sweep a very large region in $\mathbf{F}$ space as their phases vary, and there may exist no expansion of a prescribed j.p.d. involving these reflexions whose domain of validity would span that whole region. Under these circumstances there exists no uniform analytical representation of the j.p.d. in question which would be equally reliable for all values of the phases of these reflexions. There is therefore no such object, for practical purposes, as 'the' joint distribution of a set of structure factors from which all conditional distributions required would be obtainable by specialization.

This difficulty is formally identical to that encountered in the theory of analytic functions of a complex variable (Whittaker & Watson, 1927; Ahlfors, 1966): only very rarely is it possible to approximate such functions everywhere by means of a single power series centred at the origin; in the general case, they have to be defined through a collection of locally convergent power series related to each other by analytic continuation, whose domains of convergence cover the entire complex plane.

The same idea can readily be applied to the present situation by taking the following steps:

(*a*) the unwieldy locus defined in $\mathbf{F}$ space by the phase circles can be broken up into small regions, each surrounding a point defined by assigning trial phases to large amplitudes;

(*b*) in every such region the j.p.d. can then be approximated by a different asymptotic expansion, recentred away from the origin to the point representative of the phase information assumed, and valid only in that region.

The j.p.d. of a set of structure factors will then be handled not as a single object, but as a *collection* of local c.p.d.'s each derived from a member of a sufficiently comprehensive set of distinct prior phase assignments. This procedure effectively breaks the deadlock associated with the traditional reliance on

an asymptotic expansion rigidly tethered to the origin **0** of **F** space.

In this way, the two difficulties diagnosed in the standard procedure are resolved simultaneously: the approximations obtained are now accurate (because one always uses their asymptotic expansions near their centres) and maximally informative (because they do not result from reassembling marginal distributions).

2.3.2. *Real-space counterpart of recentring*

Since the atoms are assumed to be of unit weight, and to be identically and independently distributed, it follows from the law of large numbers that the j.p.d. of any set of unitary structure factors is always centred around the vector of corresponding Fourier coefficients of the prior distribution of atoms $q(\mathbf{x})$ (sometimes simply called the 'prior' in the sequel). Hence the process of recentring the asymptotic expansion of j.p.d. around assumed values of some unitary structure factors necessarily entails the use of a non-uniform prior $q(\mathbf{x})$ which reproduces these assumed values.

The procedure outlined in § 2.3.1 may then be rephrased in real-space terms as follows:

(*a*) generate trial sets of phases for the very strongest reflexions so as to give a reasonable coverage of all possible joint values of their structure factors, and incorporate the prior knowledge corresponding to each choice into a non-uniform prior attached to that choice;

(*b*) for each such choice, set up the conditional distribution of any other set of structure factors using the associated non-uniform prior.

In these terms, some of the deficiencies of traditional methods are more directly grasped. In actual practice, one never does start from the state of total ignorance represented by a uniform prior: most space groups require that three different phases of signs be specified in order to define the origin uniquely, and one usually chooses the *largest* three reflexions whose indices form a primitive set. Such a choice will already cause j.p.d. estimations to be attempted outside the range where the Edgeworth expansion around **0** could be relied upon to give accurate results.

2.4. *A new multisolution approach*

The guiding principle used above may be applied recursively. Let each initial trial set of phases chosen at step (*a*) be represented by a node of a tree. By examining the c.p.d. of new phases constructed at each node according to (*b*), different choices may be made for these new phases, leading to the construction of additional branches from that node; the tip-node of each branch is then processed according to (*a*), and so on. This multisolution strategy will be formulated in greater detail in § 8.1.

The hall-mark of this scheme is the constant updating of the prior $q(\mathbf{x})$ in the light of all the phase choices made. This constitutes a fundamental break-away from currently used multisolution methods: the latter do start off by generating distinct hypotheses by assigning trial phases to a subset of the strongest reflexions, but then proceed to estimate phase invariants on the basis of j.p.d.'s derived from a *uniform* prior throughout.

To implement these ideas computationally, two main problems must be addressed.

(1) Given some trial set of phases for a number of strong reflexions, how should this knowledge be reflected in the choice of a non-uniform prior distribution of atoms?

(2) Once a non-uniform prior distribution has been chosen, what is the optimal method for constructing the best asymptotic expansion of the j.p.d. of an arbitrary set of structure factors incorporating this prior knowledge?

Problem 1 is addressed in § 3. Its discussion leads naturally to Jaynes's 'maximum-entropy principle' and its associated formalism, and to a shift of emphasis to real-space combinatorial arguments.

Problem 2 may be treated from the 'classical' standpoint, *i.e.* by probability methods in reciprocal space. It may also be treated wholly within the framework of real-space combinatorial methods and of the maximum-entropy formalism. Both approaches are presented in § 4, and are shown to lead to equivalent results. It thus emerges from this study that the method of joint distributions amounts precisely to a reciprocal-space evaluation of the entropy of the electron density function, as indicated earlier (Bricogne, 1982).

Finally, the uniqueness and optimality of the entire scheme is established in § 5 by proving its equivalence with a purely analytical solution to the approximation problem for j.p.d.'s by means of Daniels's saddlepoint method.

3. Choice of a non-uniform prior distribution of atoms

The requirement that the non-uniform prior $q(\mathbf{x})$ should reproduce assumed values of some unitary structure-factor values leaves this function greatly underdetermined since its other Fourier coefficients remain unconstrained. A natural criterion for selecting an 'optimal' $q(\mathbf{x})$ out of all the admissible ones is that it should restrict minimally the range of structures which can be generated under the initial assumptions. A quantitative formulation of this heuristic argument involves an appeal to Shannon's theory of communication, which will provide the basis for the justification of Jaynes's maximum-entropy principle.

3.1. *The prior distribution of atoms and the source entropy*

The most fruitful way to conceptualize the role of $q(\mathbf{x})$ is to regard it as specifying the statistical structure of the source of random atomic positions used as the starting point of direct methods.

This continuous source may be approximated by a discrete source at any finite resolution, as indicated in § 1.1. An important quantity associated with such a discrete source is its entropy per symbol H, defined by

$$H = -\sum_{i=1}^{B} q_i \log q_i \tag{3.1}$$

which measures the 'amount of uncertainty' involved in the choice of a symbol. Two theorems of Shannon (Shannon & Weaver, 1949, Appendix 3) provide a more direct grasp of the meaning of the source entropy:

(1) H is approximately the logarithm of the reciprocal probability of a typical long sequence, divided by the number of symbols in the sequence;

(2) H gives the rate of growth, with increasing sequence length, of the logarithm of the number of reasonably probable sequences, regardless of the precise meaning given to 'reasonably probable'.

The entropy H of a source thus measures the strength of the restrictions placed on the permissible sequences of symbols, greater restrictions leading to lower entropy. In the case at hand, its maximum value $H_{\max} = \log B$ is reached when all symbols are equally probable, *i.e.* for a uniform prior distribution of atoms. When the prior is not uniform, the usage of the different symbols is biased away from this maximum freedom, and the entropy of the source is lower; by Shannon's theorem (2), the number of 'reasonably probable' structures with a given number of atoms decreases accordingly.

The case of continuous distributions is a straightforward extension of the previous results. A non-uniform prior distribution of atoms $q(\mathbf{x})$ gives rise to a source of random atomic positions with entropy

$$H = -\int_V q(\mathbf{x}) \log q(\mathbf{x})\, \mathrm{d}^3\mathbf{x}, \tag{3.2}$$

the maximum value $H_{\max} = \log V$ being reached for a uniform prior $q(\mathbf{x}) = 1/V$.

3.2. *The maximum-entropy criterion*

We are now in a position to address the first problem mentioned at the end of § 2, namely the choice of a non-uniform prior distribution of atoms $q(\mathbf{x})$ from the knowledge of a limited set of its Fourier coefficients.

By the two theorems of Shannon quoted above, the entropy H of the source defined by $q(\mathbf{x})$ affords a quantitative measure of the extent to which the range of structures which can be generated with any likelihood has been narrowed down. Any reduction of the entropy of the source beyond that strictly necessary to accommodate the prior knowledge will be reflected by a correlative decrease in the number of 'reasonably probable' structures of prescribed size N, hence by an unnecessary commitment to a subset of all possible structures consistent with the initial data; this is equivalent to imposing extra constraints, not warranted by the data.

If the knowledge of some unitary structure factors is to be reflected in a non-uniform prior distribution of atoms, the previous reasoning leads uniquely to choosing for $q(\mathbf{x})$ the maximum-entropy distribution having the requisite Fourier coefficients, since it defines the source which reproduces the assumed prior information with minimum bias. This is a particular instance of the 'ME principle' of Jaynes (1957, 1968), which has been reviewed in detail in the context of crystallography by Wilkins, Varghese & Lehmann (1983).

The quantity which measures most directly the strength of the restrictions introduced by $q(\mathbf{x})$ is not the source entropy $H(q)$ itself, but rather the difference $H(q) - H_{\max}$, since the proportion of 'reasonably probable' N-atom structures in the ensemble of the corresponding source is $\exp\{N[H(q) - H_{\max}]\}$. This difference may be written

$$H(q) - H_{\max} = -\int_V q(\mathbf{x}) \log [q(\mathbf{x})/m(\mathbf{x})]\, \mathrm{d}^3\mathbf{x},$$

where $m(\mathbf{x}) = 1/V$ is the uniform distribution which is such that $H(m) = H_{\max}$. In some instances, extra knowledge may be available about non-uniformities in the distribution of atoms even before any structure-factor values are assumed: one then would use a non-uniform 'prior prejudice' $m(\mathbf{x})$ to define the zero of the entropy scale. For instance, if a rough molecular envelope U were known at the outset, one would use

$$m(\mathbf{x}) = \frac{\{(1-\alpha)\chi_U(\mathbf{x}) + \alpha[1-\chi_U(\mathbf{x})]\}}{[\alpha V + (1-2\alpha)U]}$$

with α small compared to 1, χ_U being the indicator function of U.

The final form of the ME criterion is thus that $q(\mathbf{x})$ should be chosen so as to maximize, under the constraints expressing the prior knowledge of some of its Fourier coefficients, its entropy

$$\mathscr{S}_m(q) = -\int_V q(\mathbf{x}) \log [q(\mathbf{x})/m(\mathbf{x})]\, \mathrm{d}^3\mathbf{x} \tag{3.3}$$

relative to the prior prejudice $m(\mathbf{x})$ which maximizes H in the absence of such knowledge.

3.3. *The maximum-entropy formalism*

Jaynes (1957) solved the maximization problem just posed in the case of general linear constraints, thus

constructing a formalism of great power and utmost elegance. In this section, the basic equations will be derived for the purpose of later comparison with the 'saddlepoint equations' in § 5.4.

3.3.1. *The maximum-entropy equations*

The most unbiased probability density $q(\mathbf{x})$, under prior prejudice $m(\mathbf{x})$, satisfying the constraint equations

$$\mathscr{C}_j(q) \equiv \int_V q(\mathbf{x}) C_j(\mathbf{x})\, \mathrm{d}^3\mathbf{x} = c_j \quad (j = 1, 2, \ldots, M), \tag{ME0}$$

where the $\mathscr{C}_j(q)$ are linear constraint functionals defined by given constraint functions $C_j(\mathbf{x})$, and the c_j are constraint values, is obtained by maximizing the entropy of q relative to m defined by (3.3). An extra constraint is the normalization condition

$$\mathscr{C}_0(q) \equiv \int_V q(\mathbf{x}) 1\, \mathrm{d}^3\mathbf{x} = 1,$$

to which it is convenient to give the label $j = 0$ so that it may be handled together with the others by putting $C_0(\mathbf{x}) \equiv 1$, $c_0 = 1$.

By a standard variational argument, in constant use in statistical mechanics, this constrained maximization is equivalent to the unconstrained maximization of the functional

$$\mathscr{S}_m(q) + \sum_{j=0}^{M} \lambda_j \mathscr{C}_j(q), \tag{3.4}$$

where the λ_j are Lagrange multipliers whose values may be determined from the constraints. This new variational problem is readily solved: if $q(\mathbf{x})$ is varied to $q(\mathbf{x}) + \delta q(\mathbf{x})$ the resulting variations in the functionals $\mathscr{S}_m$ and $\mathscr{C}_j$ will be

$$\begin{cases} \delta \mathscr{S}_m = \int_V \{-1 - \log [q(\mathbf{x})/m(\mathbf{x})]\} \delta q(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \\ \delta \mathscr{C}_j = \int_V C_j(\mathbf{x})\ \delta q(\mathbf{x})\, \mathrm{d}^3\mathbf{x}, \end{cases} \tag{3.5}$$

respectively. If the variation of functional (3.4) is to vanish for arbitrary variations $\delta q(\mathbf{x})$, the integrand in the expression for that variation from (3.5) must vanish identically. Therefore, the maximum-entropy density $q^{\mathrm{ME}}(\mathbf{x})$ satisfies the relation

$$-1 - \log [q(\mathbf{x})/m(\mathbf{x})] + \sum_{j=0}^{M} \lambda_j C_j = 0$$

and hence

$$q^{\mathrm{ME}}(\mathbf{x}) = m(\mathbf{x}) \exp(\lambda_0 - 1) \exp\left[\sum_{j=1}^{M} \lambda_j C_j(\mathbf{x})\right].$$

It is convenient now to separate the multiplier λ_0 associated with the normalization constraint by putting:

$$\lambda_0 - 1 = -\log Z,$$

where Z is a function of the other multipliers $\lambda_1, \ldots, \lambda_M$. The final expression for $q^{\mathrm{ME}}(\mathbf{x})$ is thus

$$q^{\mathrm{ME}}(\mathbf{x}) = \frac{m(\mathbf{x})}{Z(\lambda_1, \ldots, \lambda_M)} \exp\left[\sum_{j=1}^{M} \lambda_j C_j(\mathbf{x})\right]. \tag{ME1}$$

The values of Z and $\lambda_1, \ldots, \lambda_M$ may be determined by solving the initial constraint equations. The normalization condition demands that

$$Z(\lambda_1, \ldots, \lambda_M) = \int_V m(\mathbf{x}) \exp\left[\sum_{j=1}^{M} \lambda_j C_j(\mathbf{x})\right] \mathrm{d}^3\mathbf{x}. \tag{ME2}$$

The generic constraint equations determine $\lambda_1, \ldots, \lambda_M$ by the conditions that

$$\int_V (1/Z) m(\mathbf{x}) \exp\left[\sum_{l=1}^{M} \lambda_l C_l(\mathbf{x})\right] C_j(\mathbf{x})\, \mathrm{d}^3\mathbf{x} = c_j$$

for $j = 1, \ldots, M$. But, by Leibniz's rule of differentiation under the integral sign, these equations may be written in the compact form:

$$\partial/\partial\lambda_j (\log Z) = c_j \quad (j = 1, 2, \ldots, M). \tag{ME3}$$

Equations (ME1), (ME2) and (ME3) constitute the *maximum-entropy equations.*

3.3.2. *Expression of the entropy*

The maximum value attained by the entropy is readily found:

$$\begin{aligned} \mathscr{S}_m(q^{\mathrm{ME}}) &= -\int_V q^{\mathrm{ME}}(\mathbf{x}) \log [q^{\mathrm{ME}}(\mathbf{x})/m(\mathbf{x})]\, \mathrm{d}^3\mathbf{x} \\ &= -\int_V q^{\mathrm{ME}}(\mathbf{x}) \left[-\log Z + \sum_{j=1}^{M} \lambda_j C_j(\mathbf{x})\right] \mathrm{d}^3\mathbf{x}, \end{aligned}$$

i.e., using the constraint equations,

$$\mathscr{S}_m(q^{\mathrm{ME}}) = \log Z - \sum_{j=1}^{M} \lambda_j c_j. \tag{3.6}$$

The latter expression may be rewritten, by means of (ME3), as

$$\mathscr{S}_m(q^{\mathrm{ME}}) = \log Z - \sum_{j=1}^{M} \lambda_j \frac{\partial}{\partial \lambda_j} (\log Z), \tag{3.7}$$

which shows that, in their dependence on the λ's the entropy and $\log Z$ are related by a Legendre transformation (Lanczos, 1970).

3.3.3. *Generating properties of log Z*

The function $\log Z$ has useful generating properties, of which equations (ME3) are the simplest instance.

Introducing the notation $\partial_k = \partial/\partial\lambda_k$ and

$$\langle C \rangle = \int_V q^{\mathrm{ME}}(\mathbf{X}) C(\mathbf{x})\, \mathrm{d}^3\mathbf{x},$$

we have

$$\partial_k (\log Z) = (1/Z)\,\partial_k Z = \langle C_k \rangle$$
$$= c_k \quad \text{by (ME3).}$$
$$\partial^2_{jk} (\log Z) = \partial_j [(1/Z)\,\partial_k Z]$$
$$= (1/Z)\partial^2_{jk} Z - (1/Z^2)\,\partial_j Z\,\partial_k Z$$
$$= \langle C_j C_k \rangle - \langle C_j \rangle \langle C_k \rangle$$
$$\partial^3_{ijk} (\log Z) = \langle C_i C_j C_k \rangle - [\langle C_i \rangle \langle C_j C_k \rangle + \langle C_j \rangle \langle C_k C_i \rangle$$
$$+ \langle C_k \rangle \langle C_i C_j \rangle] + 2 \langle C_i \rangle \langle C_j \rangle \langle C_k \rangle.$$

One recognizes the familiar expressions for cumulants: the partial derivatives of $\log Z$ with respect to the λ's are cumulant averages of corresponding C's. This property will be further elucidated in § 5.4.

In particular, the second partial derivative matrix has elements

$$\partial^2_{jk} (\log Z) = \langle C_j C_k \rangle - \langle C_j \rangle \langle C_k \rangle$$
$$= \langle (C_j - \langle C_j \rangle)(C_k - \langle C_k \rangle) \rangle. \tag{3.8}$$

The Hessian matrix of $\log Z$ is thus a Gram matrix, hence is positive definite, so that $\log Z$ is a *convex* function of the λ's. It is clear from the expressions above that this matrix is the covariance matrix of the deviations of the C's from their maximum-entropy average values.

3.4. *The crystallographic ME formalism in P1*

The general ME formalism for linear constraints is immediately applicable to the construction of a prior distribution of atoms from the knowledge of a set of trial structure-factor values (§ 3.3): it suffices to derive and solve the corresponding ME equations. This task will be carried out in stages, and will be shown to give rise in a natural fashion to the usual algebra of phase invariants and the the Hauptman–Giacovazzo formalism. The solvability of the ME equations will be studied in § 7, using results established in §§ 5 and 6.

3.4.1. *Notation and accessory results*

The standard formulae relating the prior distribution of atoms $q(\mathbf{x})$ and its normalized structure factors $U_\mathbf{h}$ will be written

$$q(\mathbf{x}) = (1/V) \sum_\mathbf{h} U_\mathbf{h} \exp(-2\pi i \mathbf{h} \,.\, \mathbf{x})$$
$$= (1/V)[1 + 2 \sum_\mathbf{h}{}' |U_\mathbf{h}| \cos(2\pi \mathbf{h} \,.\, \mathbf{x} - \varphi_\mathbf{h})] \tag{3.9}$$
$$U_\mathbf{h} = \int_V q(\mathbf{x}) \exp(2\pi i \mathbf{h} \,.\, \mathbf{x})\, \mathrm{d}^3\mathbf{x},$$

where $\sum_\mathbf{h}$ denotes a summation over all $\mathbf{h}$, and $\sum'_\mathbf{h}$ a summation over *unique* non-origin reflexions (*i.e.* over a 'hemisphere').

A set of formulae which will be used throughout the forthcoming calculations relate to the generation of the modified Bessel functions $I_n(z)$ *via* Schlömilch's formula (Watson, 1944):

$$\exp(z \cos t) = \sum_{n=-\infty}^{+\infty} I_n(z) \exp(int)$$
$$= \sum_{n=-\infty}^{+\infty} I_n(z) \cos nt \tag{3.10}$$

since $I_{-n}(z) = I_n(z)$ or, equivalently,

$$I_n(z) = \frac{1}{2\pi} \int_0^{2\pi} \exp(z \cos t) \cos nt \, \mathrm{d}t. \tag{3.11}$$

3.4.2. *Case of one structure factor in P1*

Starting from a uniform prior prejudice $m(\mathbf{x}) = 1/V$, let the constraint be the value of $U_\mathbf{h} = |U_\mathbf{h}| \exp(i\varphi_\mathbf{h})$. The constraint equations may be written

$$\begin{cases} \int_V q(\mathbf{x}) \cos(2\pi \mathbf{h} \,.\, \mathbf{x} - \varphi_\mathbf{h})\, \mathrm{d}^3\mathbf{x} = |U_\mathbf{h}| \\ \int_V q(x) \sin(2\pi \mathbf{h} \,.\, \mathbf{x} - \varphi_\mathbf{h})\, \mathrm{d}^3\mathbf{x} = 0. \end{cases} \tag{3.12}$$

Let these two constraints be given Lagrange multipliers λ and μ, respectively. Then, by (ME2),

$$Z(\lambda, \mu) = \int_V m(\mathbf{x}) \exp[\lambda \cos(2\pi \mathbf{h} \,.\, \mathbf{x} - \varphi_\mathbf{h})$$
$$+ \mu \sin(2\pi \mathbf{h} \,.\, \mathbf{x} - \varphi_\mathbf{h})]\, \mathrm{d}^3\mathbf{x}$$
$$= (1/V) \int_V \exp\{\kappa \cos[2\pi \mathbf{h} \,.\, \mathbf{x}$$
$$- (\varphi_\mathbf{h} + \theta)]\}\, \mathrm{d}^3\mathbf{x}$$
$$= I_0(\kappa), \tag{3.13}$$

where $\lambda = \kappa \cos\theta$, $\mu = \kappa \sin\theta$. The constraint equations (ME3) give

$$\begin{cases} \dfrac{\partial}{\partial \mu} (\log Z) \equiv \dfrac{\mu}{\kappa} \dfrac{\partial}{\partial \kappa} (\log Z) = 0 \\ \dfrac{\partial}{\partial \lambda} (\log Z) \equiv \dfrac{\lambda}{\kappa} \dfrac{\partial}{\partial \kappa} (\log Z) = |U_\mathbf{h}|. \end{cases}$$

The second equation determines κ uniquely (see Fig. 1) *via* the relation

$$I_1(\kappa)/I_0(\kappa) = |U_\mathbf{h}|. \tag{3.14}$$

Finally, by (ME1),the ME distribution satisfying the constraints is

$$q^{\mathrm{ME}}(\mathbf{x}) = \frac{1}{V} \frac{1}{I_0(\kappa)} \exp[\kappa \cos(2\pi \mathbf{h} \,.\, \mathbf{x} - \varphi_\mathbf{h})]. \tag{3.15}$$

Several remarks are in order concerning this first elementary calculation.

(1) If $|U_{\mathbf{h}}|$ is small, then $\kappa \sim 2|U_{\mathbf{h}}|$ since $I_1(x)/I_0(x) \sim x/2$ as $x \sim 0$. Then, by linearizing the exponential, the above result is indistinguishable from the usual one-term Fourier series (3.9):

$$q^{\mathrm{ME}}(\mathbf{x}) \simeq (1/V)[1 + 2|U_{\mathbf{h}}| \cos(2\pi \mathbf{h} \cdot \mathbf{x} - \varphi_{\mathbf{h}})].$$

(2) If $|U_{\mathbf{h}}|$ is large, the correspondence between $|U_{\mathbf{h}}|$ and κ becomes noticeably non-linear (*cf.* Fig. 1). A full expansion of (3.15) using (3.10) gives

$$q^{\mathrm{ME}}(\mathbf{x}) = (1/V)\left\{1 + 2\sum_{n>0} \frac{I_n(\kappa)}{I_0(\kappa)} \cos[n(2\pi \mathbf{h} \cdot \mathbf{x} - \varphi_{\mathbf{h}})]\right\}, \tag{3.16}$$

which shows that the spectrum of the prior distribution has been extrapolated beyond the one datum provided by the prior knowledge. The application of the ME principle has extended the spectrum in order to suppress the familiar series-termination 'ripples'. The latter would constitute spurious detail not warranted by the given data: the rest of the spectrum is *unknown*, rather than known to be zero.

(3) If $|U_{\mathbf{h}}|$ tends to 1, its maximum possible value, the ME distribution $q^{\mathrm{ME}}(\mathbf{x})$ tends towards a limiting singular form which may be written

$$q^{\mathrm{ME}}(\mathbf{x}) = (1/V)\delta[\Phi_{\mathbf{h}}(\mathbf{x})],$$

where δ is Dirac's δ function, and $\Phi_{\mathbf{h}}(\mathbf{x}) = 2\pi \mathbf{h} \cdot \mathbf{x} - \varphi_{\mathbf{h}}$ modulo 2π. This limiting form is thus the tensor product of a regular array of δ functions along $\mathbf{h}$ and of a uniform plane distribution in the planes $\mathbf{h} \cdot \mathbf{x} \equiv \varphi_{\mathbf{h}} \bmod 2\pi$. It is a much more intelligent guess at the true nature of a distribution for which $|U_{\mathbf{h}}| \simeq 1$ than is the corresponding one-term Fourier series (which would exhibit negative regions as soon as $|U_{\mathbf{h}}| > \frac{1}{2}$).

(4) The sine constraint in (3.12) ends up playing no role ($\mu = 0$) because the ME solution satisfying the cosine constraint alone automatically satisfies the sine constraint as well, by remaining 'maximally non-committal' with regard to the choice of enantiomorph.

(5) The approximate relation $\kappa \simeq 2|U_{\mathbf{h}}|$ comes from the fact that we applied a constraint to only one reflexion of the Friedel pair $(+\mathbf{h}, -\mathbf{h})$. It is easily checked that, should we also impose the Friedel-related constraint as if it were independent, we would be led to the same answer as that defined by (3.14) and (3.15) Therefore, redundant constraints are properly weighted down in an automatic fashion. This conclusion is of general validity (Jaynes, 1968), and will be of some relevance for special reflexions in higher space groups.

(7) The relative entropy

$$\mathcal{S}_m(q) = \log I_0(\kappa) - \kappa|U_{\mathbf{h}}|$$

does not depend on the phase $\varphi_{\mathbf{h}}$, since no origin is fixed. For small $|U_{\mathbf{h}}|$, $\kappa \simeq 2|U_{\mathbf{h}}|$, $\log I_0(\kappa) \simeq (\kappa/2)^2 = |U_{\mathbf{h}}|^2$, so that $\mathcal{S}_m(q) \simeq -|U_{\mathbf{h}}|^2$ in accordance with Wilson's statistics.

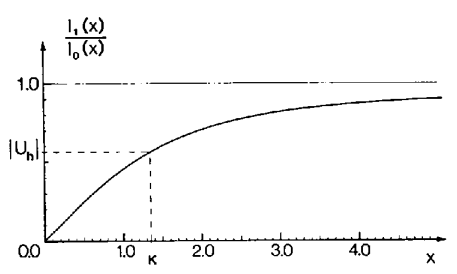


Fig. 1. The function $I_1(x)/I_0(x)$.

3.4.3. *General case of M structure factors in P*1

Let the prior knowledge of structure factors pertaining to $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_M$ be written

$$\begin{cases} \int_V q(\mathbf{x}) \cos(2\pi \mathbf{h}_j \cdot \mathbf{x} - \varphi_{\mathbf{h}_j})\, \mathrm{d}^3\mathbf{x} = |U_{\mathbf{h}_j}| \\ \int_V q(\mathbf{x}) \sin(2\pi \mathbf{h}_j \cdot \mathbf{x} - \varphi_{\mathbf{h}_j})\, \mathrm{d}^3\mathbf{x} = 0 \end{cases} \tag{3.17}$$

for $j = 1, 2, \ldots, M$, and let these constraints have multipliers $\lambda_j = \kappa_j \cos\theta_j$ and $\mu_j = \kappa_j \sin\theta_j$, respectively. Then, for a uniform prior prejudice $m(\mathbf{x}) = 1/V$, (ME2) leads to

$$\begin{aligned} &Z(\lambda_1, \mu_1, \ldots, \lambda_M, \mu_M) \\ &= \frac{1}{V}\int_V \exp\left\{\sum_{j=1}^{M} [\lambda_j \cos(2\pi \mathbf{h}_j \cdot \mathbf{x} - \varphi_{\mathbf{h}_j}) \right. \\ &\quad \left. + \mu_j \sin(2\pi \mathbf{h}_j \cdot \mathbf{x} - \varphi_{\mathbf{h}_j})]\right\} \mathrm{d}^3\mathbf{x} \\ &= \frac{1}{V}\int_V \exp\left\{\sum_{j=1}^{M} \kappa_j \cos[2\pi \mathbf{h}_j \cdot \mathbf{x} - (\varphi_{\mathbf{h}_j} + \theta_j)]\right\} \mathrm{d}^3\mathbf{x} \\ &= \frac{1}{V}\int_V \prod_{j=1}^{M} \left(\sum_{m_j=-\infty}^{+\infty} I_{m_j}(\kappa_j) \right. \\ &\quad \left. \times \exp\{im_j[2\pi \mathbf{h}_j \cdot \mathbf{x} - (\varphi_{\mathbf{h}j} + \theta_j)]\}\right) \mathrm{d}^3\mathbf{x} \\ &= \sum_{m_1=-\infty}^{+\infty} \sum_{m_2=-\infty}^{+\infty} \cdots \sum_{m_M=-\infty}^{+\infty} \\ &= I_{m_1}(\kappa_1) I_{m_2}(\kappa_2) \ldots I_{m_M}(\kappa_M). \\ &\quad \times \frac{1}{V}\int_V \exp\left\{i\left[2\pi\left(\sum_{j=1}^{M} m_j \mathbf{h}_j\right) \cdot \mathbf{x} \right.\right. \\ &\quad \left.\left. - \sum_{j=1}^{M} m_j(\varphi_{\mathbf{h}_j} + \theta_j)\right]\right\} \mathrm{d}^3\mathbf{x}. \end{aligned} \tag{3.18}$$

All integrals vanish, except those corresponding to M-tuples of integers $(m_1, m_2, \ldots, m_M)$ such that

$$\sum_{j=1}^{M} m_j \mathbf{h}_j = \mathbf{0}. \tag{3.19}$$

Therefore, since $I_{-m}(\kappa) = I_m(\kappa)$, the final expression for Z reads

$$Z = \sum_{\sum_{j=1}^{M} m_j \mathbf{h}_j = \mathbf{0}} I_{m_1}(\kappa_1) \ldots I_{m_M}(\kappa_M) \times \cos\left[\sum_{j=1}^{M} m_j(\varphi_{\mathbf{h}_j} + \theta_j)\right]. \tag{3.20}$$

The equations (ME3) by which the multipliers may be determined are now

$$\begin{cases} \dfrac{\partial}{\partial \lambda_j}(\log Z) = |U_{\mathbf{h}j}| \\ \dfrac{\partial}{\partial \mu_j}(\log Z) = 0 \end{cases}$$

or equivalently

$$\begin{cases} \dfrac{\partial}{\partial \kappa_j}(\log Z) = |U_{\mathbf{h}j}| \cos \theta_j \\ \dfrac{1}{\kappa_j}\dfrac{\partial}{\partial \theta_j}(\log Z) = -|U_{\mathbf{h}_j}| \sin \theta_j. \end{cases} \tag{3.21}$$

Finally, the ME distribution (ME1) is

$$q^{\mathrm{ME}}(\mathbf{x}) = \frac{1}{V}\frac{1}{Z} \exp\left\{\sum_{j=1}^{M} \kappa_j \cos\left[2\pi \mathbf{h}_j \cdot \mathbf{x} - (\varphi_{\mathbf{h}_j} + \theta_j)\right]\right\}. \tag{3.22}$$

These results call for a number of comments.

(1) If $M \leq 3$ and $\mathbf{h}_1, \ldots, \mathbf{h}_m$ are linearly independent, then condition (3.19) is fulfilled only for $m_1 = m_2 = m_3 = 0$, hence

$$Z = \prod_{j=1}^{M} I_0(\kappa_j).$$

This is the direct product of the Z functions (3.13) associated with individual constraints; the equations for the κ's are uncoupled and all have the form (3.14); finally, the solution $q^{\mathrm{ME}}(\mathbf{x})$ is the direct product of independent functions similar to (3.15), one for each $j = 1, \ldots, M$. Expansion of $q^{\mathrm{ME}}(\mathbf{x})$ as was done in (3.16) shows that spectrum extrapolation takes place through a 'tangent-like' formula:

$$(U_{\mathbf{h}_i + \mathbf{h}_j})^{\mathrm{ME}} = U_{\mathbf{h}_i} U_{\mathbf{h}_j}. \tag{3.23}$$

The entropy is still independent of the phases, since the latter may be chosen arbitrarily to fix the origin.

(2) Interaction occurs between the different constraints whenever a dependence relation with integral coefficients can be formed between the $\mathbf{h}_j$, as expressed by (3.19). It is no longer the case that we may solve separately for the multipliers associated with each constraint: they are now extensively coupled through numerous interaction terms. In particular, the sine constraints (with multipliers μ_j) are no longer automatically satisfied, which reflects the fact that the initial data may now contain enantiomorph-sensitive information. The entropy

$$\mathscr{S}_m(q^{\mathrm{ME}}) = \log Z - \sum_{j=1}^{M} \kappa_j |U_{\mathbf{h}_j}| \cos \theta_j$$

then depends, as does Z, on all the phase invariants that may be constructed from the phases of the initial data.

(3) The degree of interaction between reflexions may be estimated if one recalls that

$$I_n(x) \simeq (1/|n|!)(x/2)^{|n|} \quad \text{as } x \simeq 0.$$

Therefore, the term in Z associated with the dependence relation $\sum_{j=1}^{M} m_j \mathbf{h}_j = \mathbf{0}$ will be approximately of size

$$\prod_{j=1}^{M} (1/|m_j|!)|U_{\mathbf{h}_j}|^{|m_j|},$$

since for most reflexions κ_j is of the same order of magnitude as $|U_{\mathbf{h}_j}|$. In terms of $|E|$'s, which are related to the $|U|$'s by $|E_{\mathbf{h}}| = |U_{\mathbf{h}}|\sqrt{N}$ where N is the number of atoms, this quantity will be of order $(1/\sqrt{N})^{\sum_j |m_j|}$ in the usual ranking of strength (Klug, 1958). The strongest interactions will thus occur between clusters of reflexions involved in the familiar low-order relations with small coefficients (triplets, quartets, ...) for which $\sum_{j=1}^{M} |m_j|$ is lowest.

(4) The primitive concept in the present analysis of the structure of the Z function is not that of an individual phase invariant, but of an integral dependence relation between the structure-factor indices. This point of view renders obvious the existence of 'kinship relations' amongst phase invariants, since it is possible to re-group the terms of an original dependence relation so as to produce new relations between linear combinations of the initial indices. For example,

$$\mathbf{h} + \mathbf{k} + \mathbf{l} + \mathbf{m} = 0 \quad \text{(quartet on } \mathbf{h}, \mathbf{k}, \mathbf{l}, \mathbf{m})$$

may be re-read

$$(\mathbf{h} + \mathbf{k}) + \mathbf{l} + \mathbf{m} = \mathbf{0} \quad [\text{triplet on } (\mathbf{h} + \mathbf{k}), \mathbf{l}, \mathbf{m}]$$

or

$$(\mathbf{h} + \mathbf{l}) + \mathbf{k} + \mathbf{m} = \mathbf{0} \quad [\text{triplet on } (\mathbf{h} + \mathbf{l}), \mathbf{k}, \mathbf{m}]$$

or

$$(\mathbf{h} + \mathbf{m}) + \mathbf{k} + \mathbf{l} = \mathbf{0} \quad [\text{triplet on } (\mathbf{h} + \mathbf{m}), \mathbf{k}, \mathbf{l}].$$

This process corresponds exactly to that of forming 'cross vectors'. Such families of invariants are internally coupled, since the occurrence of one member

in the Z sum automatically implies that of any other member present in the initial data. Similarly, new relations can be derived from an initially given one by adding and subtracting an arbitrary vector; new cross vectors may then be generated by the above process.

It is therefore clear that the entire formalism of neighbourhoods (phasing shells) and representations (Giacovazzo, 1980) is implicit in the structure of the Z function.

(5) Finally, an abridged notation for the expression of the general Z function (3.20) will be established.

The indices $\mathbf{h}_1, \ldots, \mathbf{h}_M$ being given, let K be defined by

$$K = \left\{ \mathbf{m} \in \mathbb{Z}^M \;\middle|\; \sum_{j=1}^{M} m_j \mathbf{h}_j = \mathbf{0} \right\}. \qquad (3.24)$$

If r is the rank of the system of index vectors $(\mathbf{h}_1, \ldots, \mathbf{h}_M)$, then K is a sublattice of $\mathbb{Z}^M$ of rank $M-r$, and $\mathbb{Z}^M/K$ is a sublattice of rank r. With the notation $\psi_j = \varphi_{\mathbf{h}_j} + \theta_j$, $\Phi_j(\mathbf{x}) = 2\pi \mathbf{h}_j \cdot \mathbf{x} - \psi_j$, and the shorthand

$$I_{\mathbf{m}}(\boldsymbol{\kappa}) = \prod_{j=1}^{M} I_{m_j}(\kappa_j)$$

$$(\mathbf{m} \,.\, \boldsymbol{\Phi})(\mathbf{x}) = \sum_{j=1}^{M} m_j \Phi_j(\mathbf{x})$$

$$\mathbf{m} \,.\, \boldsymbol{\Psi} = \sum_{j=1}^{M} m_j \psi_j,$$

we may write

$$Z(\boldsymbol{\kappa}, \boldsymbol{\theta}) = \sum_{\mathbf{p} \in K} I_{\mathbf{p}}(\boldsymbol{\kappa}) \exp(i\mathbf{p} \,.\, \boldsymbol{\Psi}). \qquad (3.25)$$

3.5. *The crystallographic ME formalism for a general space group*

The presence of space-group symmetry greatly enriches the pattern of interaction between phases which ultimately determines the value of the entropy, since each reflexion acquires several 'aliases' through the operation of the point group. The generalization of the ME equations from the $P1$ case is thus by no means a trivial task, particular care being needed to ensure that special and invariant reflexions are treated correctly and efficiently.

3.5.1. *Notation and definitions*

Let G denote the space group of the crystal modulo its subgroup Λ of lattice (*i.e.* primitive) translations; G is finite, and let $|G|$ denote the number of its elements. Each element g of G acts in real space as a Euclidean motion:

$$S_g(\mathbf{x}) = \mathbf{R}_g(\mathbf{x}) + \mathbf{t}_g \quad \text{modulo } \Lambda,$$

where $\mathbf{R}_g$ is an orthogonal transformation in the point group, and $\mathbf{t}_g$ is the associated non-primitive translation. Indexing by the elements of G itself provides the notational convenience that the indices carry the group law of G directly. For example, if $uv = w$ $(u, v, w \in G)$, then the group structure of G is reflected by the identities

$$\mathbf{R}_w = \mathbf{R}_u \mathbf{R}_v$$

$$\mathbf{t}_w = \mathbf{R}_u \mathbf{t}_v + \mathbf{t}_u \quad \text{modulo } \Lambda,$$

the latter being known as the Frobenius congruences.

Considerable gains in clarity and notational ease will be afforded by the use of an algebraic construction basic to group representation theory, that of the integral group ring of a finite group. Given an arbitrary finite group G, its *integral group ring* $\mathbb{Z}G$ consists of all formal sums of elements of G with coefficients in $\mathbb{Z}$:

$$\gamma = \sum_{g \in G} m(g) g$$

between which addition and multiplication are defined by

$$\sum_{g \in G} a(g) g + \sum_{g \in G} b(g) g = \sum_{g \in G} [a(g) + b(g)] g \qquad (3.26a)$$

$$\left[\sum_{g \in G} a(g) g\right]\left[\sum_{h \in G} b(h) h\right] = \sum_{g,h \in G} a(g) b(h) gh = \sum_{k \in G} c(k) k,$$

where

$$c(k) = \sum_{k \in G} a(g) b(g^{-1} g) \qquad (3.26b)$$

and scalar multiplication is defined by

$$s\left[\sum_{g \in G} a(g) g\right] = \sum_{g \in G} s a(g) g. \qquad (3.26c)$$

An equivalent construction of $\mathbb{Z}G$ is to view it as the ring of integer-valued functions on G, with multiplication defined by convolution over G (Curtis & Reiner, 1962).

3.5.2. *The maximum-entropy equations: first version*

A temporary form of the ME equations will first be derived, without regard to special or invariant reflexions, for the purpose of examining the pattern of phase interactions brought about by the space-group symmetry.

Starting from the unitary structure-factor expression (3.9), we may express the symmetry assumption by writing that $q(\mathbf{x})$ is invariant under averaging by G:

$$q(\mathbf{x}) = (1/|G|) \sum_{g \in G} q(\mathbf{R}_g \mathbf{x} + \mathbf{t}_g). \qquad (3.27)$$

Substitution into (3.9) yields for $U_{\mathbf{h}}$ the expression

$$U_{\mathbf{h}} = \int_V q(\mathbf{x})\{(1/|G|) \times \sum_{g\in G} \exp[2\pi i \mathbf{h} \,.\, (\mathbf{R}_g\mathbf{x}+\mathbf{t}_g)]\}\, \mathrm{d}^3\mathbf{x}, \qquad (3.28)$$

where the factor $1/|G|$ is necessary since the integral is over the whole unit cell rather than the asymmetric unit.

If prior knowledge of the values of $U_{\mathbf{h}_1}, \ldots, U_{\mathbf{h}_M}$ is assumed, the constraint equations generalizing (3.17) are

$$\begin{aligned} &\int_V q(\mathbf{x}) \frac{1}{|G|} \sum_{u_j\in G} \cos[2\pi({}^t\mathbf{R}_{u_j}\mathbf{h}_j) \,.\, \mathbf{x} \\ &\qquad + 2\pi \mathbf{h}_j \,.\, \mathbf{t}_{u_j} - \varphi_{\mathbf{h}_j}]\, \mathrm{d}^3\mathbf{x} = |U_{\mathbf{h}_j}| \\ &\int_V q(\mathbf{x}) \frac{1}{|G|} \sum_{u_j\in G} \sin[2\pi({}^t\mathbf{R}_{u_j}\mathbf{h}_j) \,.\, \mathbf{x} \\ &\qquad + 2\pi \mathbf{h}_j \,.\, \mathbf{t}_{u_j} - \varphi_{\mathbf{h}_j}]\, \mathrm{d}^3\mathbf{x} = 0. \end{aligned} \qquad (3.29)$$

Assigning them multipliers $\lambda_j = \kappa_j \cos\theta_j$ and $\mu_j = \kappa_j \sin\theta_j$, respectively, and starting from a uniform prior prejudice $m(\mathbf{x}) = 1/V$, we obtain for the Z function

$$Z = \frac{1}{V}\int_V \exp\left\{\sum_{j=1}^{M} \frac{\kappa_j}{|G|} \sum_{u_j\in G} \cos[2\pi({}^t\mathbf{R}_{u_j}\mathbf{h}_j) \,.\, \mathbf{x} + 2\pi \mathbf{h}_j \,.\, \mathbf{t}_{u_j} - (\varphi_{\mathbf{h}_j} + \theta_j)]\right\} \mathrm{d}^3\mathbf{x} \qquad (3.30a)$$

$$= \prod_{j=1}^{M} \prod_{u_j\in G} \frac{1}{V}\int_V \exp\left\{\frac{\kappa_j}{|G|} \cos[2\pi({}^t\mathbf{R}_{u_j}\mathbf{h}_j) \,.\, \mathbf{x} + 2\pi \mathbf{h}_j \,.\, \mathbf{t}_{u_j} - \psi_j]\right\} \mathrm{d}^3\mathbf{x}, \qquad (3.30b)$$

where $\psi_j = \varphi_{\mathbf{h}_j} + \theta_j$. A full expansion would involve an $(M \,.\, |G|)$-fold summation over $(M \,.\, |G|)$-tuples of integers $m(j, u_j)$, with j running from 1 to M and u_j running over G for each j. At this point, the burden of notation may be greatly lightened by using the integral group ring $\mathbb{Z}G$ introduced above, defining, for each $j = 1, 2, \ldots, M$,

$$\sigma_j = \sum_{u_j\in G} m(j, u_j) u_j$$

$$\mathbf{R}(\sigma_j) = \sum_{u_j\in G} m(j, u_j) \mathbf{R}_{u_j}$$

$$\mathbf{t}(\sigma_j) = \sum_{u_j\in G} m(j, u_j) \mathbf{t}_{u_j}$$

$$\eta(\sigma_j) = \sum_{u_j\in G} m(j, u_j)$$

$$\mathcal{I}_{\sigma_j}\left(\frac{\kappa_j}{|G|}\right) = \prod_{u_j\in G} I_{m(j,u_j)}\left(\frac{\kappa_j}{|G|}\right) \times \exp[-2\pi i \mathbf{h}_j \,.\, \mathbf{t}(\sigma_j)].$$

Then, using the shorthand

$$\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_M)$$

$$\mathcal{I}_{\boldsymbol{\sigma}}\left(\frac{\boldsymbol{\kappa}}{|G|}\right) = \prod_{j=1}^{M} \mathcal{I}_{\sigma_j}\left(\frac{\kappa_j}{|G|}\right)$$

$$\boldsymbol{\eta} \,.\, \boldsymbol{\psi} = \sum_{j=1}^{M} \eta(\sigma_j)\psi_j$$

and defining

$$K = \left\{\boldsymbol{\sigma} \in (\mathbb{Z}G)^M \;\middle|\; \sum_{j=1}^{M} {}^t\mathbf{R}(\sigma_j)\mathbf{h}_j = \mathbf{0}\right\} \qquad (3.31)$$

it is straightforward to verify that

$$Z(\boldsymbol{\kappa}, \boldsymbol{\theta}) = \sum_{\boldsymbol{\sigma}\in K} \mathcal{I}_{\boldsymbol{\sigma}}\left(\frac{\boldsymbol{\kappa}}{|G|}\right) \exp(i\boldsymbol{\eta} \,.\, \boldsymbol{\psi}). \qquad (3.32)$$

Comparison with (3.24) and (3.25) shows that, formally, the structure of the Z function remains the same as it was in $P1$, provided the notion of dependence relation between the $\mathbf{h}$'s over the ring $\mathbb{Z}$ of ordinary integers is replaced by that of dependence relation over the integral group ring $\mathbb{Z}G$. This formal simplicity amply justifies the introduction of $\mathbb{Z}G$.

The ME equations themselves retain their usual form (3.21).

3.5.3. *The maximum-entropy equations: second version*

The previous formulation is valid for any set of reflexions $\mathbf{h}_1, \ldots, \mathbf{h}_M$ in any space group. Whenever special or invariant reflexions are present, however, it is not the simplest nor the most economical although remark (6) in § 3.4.2 guarantees its correctness. The optimal form of the ME equations will be derived by simplifying the structure-factor expressions (3.28) as much as symmetry allows so as to obtain the most succinct expansion possible when going from (3.30*a*) to (3.30*b*). To carry out this task in greatest generality, a few more definitions are needed for which Lang (1965) may be consulted.

A reflexion $\mathbf{h}_j \neq \mathbf{0}$ being given, let G_j be the *isotropy group* of $\mathbf{h}_j$, *i.e.* the subgroup of G consisting of those $g \in G$ such that ${}^t\mathbf{R}_g\mathbf{h}_j = \mathbf{h}_j$; let $G\mathbf{h}_j$ denote the *orbit* of $\mathbf{h}_j$ under G, *i.e.* the set of all vectors ${}^t\mathbf{R}_g\mathbf{h}_j$ as g runs through G; and let G/G_j denote the collection of *right cosets* of G_j in G. It is a consequence of Lagrange's theorem that

$$|G| = |G/G_j| \,.\, |G_j| = |G\mathbf{h}_j| \,.\, |G_j|$$

where $||$ means 'number of elements of'. Therefore,

$$\begin{aligned} &\frac{1}{|G|}\sum_{g\in G} \exp[2\pi i \mathbf{h}_j \,.\, (\mathbf{R}_g\mathbf{x}+\mathbf{t}_g)] \\ &= \frac{|G_j|}{|G|}\sum_{\gamma\in G/G_j} \exp[2\pi i \mathbf{h}_j \,.\, (\mathbf{R}_\gamma\mathbf{x}+\mathbf{t}_\gamma)] \\ &= \frac{1}{|G/G_j|}\sum_{\gamma\in G/G_j} \exp[2\pi i \mathbf{h}_j \,.\, (\mathbf{R}_\gamma\mathbf{x}+\mathbf{t}_\gamma)], \end{aligned}$$

the latter expression being independent of the choice of coset representatives γ in G/G_j, as may be checked by invoking the Frobenius congruences.

Further complication occurs if $\mathbf{h}_j$ is a *centric* reflexion, *i.e.* if $G\mathbf{h}_j$ contains $-\mathbf{h}_j$. In this case there exists $g \in G$ such that ${}^t\mathbf{R}_g\mathbf{h}_j = -\mathbf{h}_j$, and the value of the phase $\varphi(\mathbf{h}_j)$ is restricted to $\pi\mathbf{h}_j . \mathbf{t}_g$ modulo π (which only depends on the coset of g modulo G_j by virtue of the Frobenius congruences). Such an element g acts on the right on the cosets in G/G_j, and g^2 acts as the identity. Therefore, the cosets may be partitioned into two disjoint classes by picking one coset in each of the two-coset orbits of this action. Let Γ_j be one such class: then the *reduced orbit* $\Gamma_j\mathbf{h}_j$ contains once and only once the Friedel-unique half of the full orbit $G\mathbf{h}_j$. Thus, $|\Gamma_j| = \frac{1}{2}|G/G_j|$, and

$$\frac{1}{|G/G_j|}\sum_{\gamma\in G/G_j}\exp[2\pi i\mathbf{h}_j . (\mathbf{R}_\gamma\mathbf{x}+\mathbf{t}_\gamma)] = \frac{1}{|\Gamma_j|}\sum_{\gamma\in\Gamma_j}\exp[2\pi i\mathbf{h}_j . (\mathbf{R}_\gamma\mathbf{x}+\mathbf{t}_\gamma)]$$

since every term was repeated twice in the first expression.

Finally, putting $\Gamma_j = G/G_j$ if $\mathbf{h}_j$ is acentric, we may rewrite the Fourier transform relations (3.9) between $q(\mathbf{x})$ and its structure factors as

$$\begin{cases} q(\mathbf{x}) = \dfrac{1}{V}\left\{1 + 2\sum_{\mathbf{h}}'' |U_\mathbf{h}| \times \sum_{\gamma\in\Gamma(\mathbf{h})}\cos[2\pi({}^t\mathbf{R}_\gamma\mathbf{h}) . \mathbf{x} + 2\pi\mathbf{h} . \mathbf{t}_\gamma - \varphi_\mathbf{h}]\right\} \\ U_{\mathbf{h}_j} = \displaystyle\int_V q(\mathbf{x})\left[\frac{1}{|\Gamma_j|}\times\sum_{\gamma\in\Gamma_j}\exp[2\pi i\mathbf{h}_j . (\mathbf{R}_\gamma\mathbf{x}+\mathbf{x}+\mathbf{t}_\gamma)]\right] d^3\mathbf{x}, \end{cases} \tag{3.33, 3.34}$$

where $\sum''$ stands for a summation over *unique* non-origin reflexions only. Comparison of (3.34) and (3.28) shows that the optimal ME equations may be obtained by replacing, in all the definitions given in § 3.5.2,

$$\sum_{u_j\in G} \text{ or } \prod_{u_j\in G} \quad \text{by} \quad \sum_{u_j\in\Gamma_j} \text{ or } \prod_{u_j\in\Gamma_j},$$

$|G|$ within the range of j by $|\Gamma_j|$,

and

$$(\mathbb{Z}G)^M \quad \text{by} \quad \prod_{j=1}^{M}\mathbb{Z}\Gamma_j.$$

Here $\mathbb{Z}\Gamma_j$ is the set of formal sums of elements of Γ_j with the additive structure (3.26*a*); it has in general no multiplicative structure but retains the $\mathbb{Z}G$-module structure which was needed to derive (3.32).

With these modifications, formula (3.32) is optimal, and equations (3.21) incorporate the correct multiplicity factors for acentric reflexions. In the case of a centric reflexion $\mathbf{h}_j$, whose phase is restricted to ω_j modulo π, the first equation (3.21*b*) alone should be used, with κ_j now a signed magnitude in the direction defined by ω_j so that its sign cos θ_j may be removed.

The crystallographic maximum-entropy formalism is now complete.

4. Effective construction of conditional distributions of structure factors from a non-uniform prior

Having specified the statistical structure of the source of random atomic positions by the choice of $q^{\mathrm{ME}}(\mathbf{x})$, we may use it in two distinct but equivalent ways (see Fig. 2) to generate the vector $\mathbf{F}$ of structure factors corresponding to a generic N-atom structure:

the upper path is that followed in the classical formulation of direct methods, where each random atomic position $\mathbf{x}$ produced by the source gives rise to a vector $\mathbf{X}(\mathbf{x})$ of contributions to the structure factors; these vectors are accumulated in structure-factor space ($\mathbb{C}^n$) over N successive atoms to make up the vector $\mathbf{F}$ (see § 1.2);

the lower path is an equivalent one in which the N successive atoms are accumulated in their boxes in real space to produce a 'message' in the form of a random structure $\boldsymbol{\rho}$; this random structure is then subjected to a Fourier transformation, yielding the same vector $\mathbf{F}$ of structure factors.

Such a source generates not only a set but an *ensemble* of messages, *i.e.* in Wiener's terms 'a repertory of possible messages, and over this repertory a measure determining the probability of these messages' (Wiener, 1949). In the present case, this probability measure reflects the fact that different random structures – and hence different sets of structure factors – will occur with different frequencies (or statistical weights) in the ensemble: it constitutes precisely the conditional probability distribution of $\mathbf{F}$ (or $\boldsymbol{\rho}$) given the prior knowledge incorporated into $q^{\mathrm{ME}}(\mathbf{x})$.

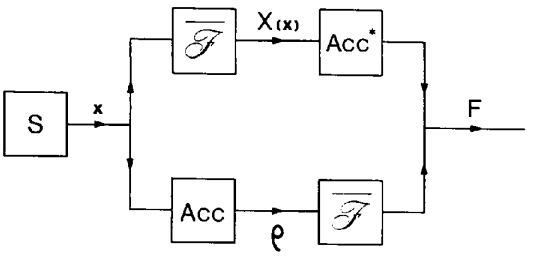


Fig. 2. The stochastic structure factor generator of direct methods. *S*: source of random atomic positions $\mathbf{x}$; Acc: accumulates atoms in real space into a random structure $\boldsymbol{\rho}$; Acc*: accumulates random vectors $\mathbf{X}(\mathbf{x})$ of atomic contributions into a random vector $\mathbf{F}$ of structure factors.

In virtue of the equivalence of the two above paths, calculating the relative frequency of a full vector of structure factors (*i.e.* the joint probability of its components) amounts precisely to evaluating the statistical weight of the corresponding atomic configuration in the ensemble of the source (Bricogne, 1982). The latter task may be carried out by means of elementary combinatorial considerations similar to those used in statistical mechanics, as will be shown in § 4.1. Alternatively, the analytical methods of § 1.2 may be employed to construct directly (§ 4.2) a local approximation to the c.p.d. of $\mathbf{U} = \mathbf{F}/N$ in the vicinity of the vector $\mathbf{U}^{\mathrm{ME}}$ of Fourier coefficients of the non-uniform prior $q^{\mathrm{ME}}(\mathbf{x})$. Since this c.p.d. has been recentred, only the leading Gaussian term of the Edgeworth asymptotic expansion (the 'central limit theorem approximation') needs to be evaluated (§ 4.2.1). By sole consideration of this Gaussian term all hitherto known formulae for estimating phase invariants can be rederived and generalized with extreme ease (§ 4.2.2). Finally, the branching problem is formulated and illustrated in § 4.3.

4.1. *Real-space combinatorial models*

Let us first discretize the source, as was done in § 1.1. The number of different sequences of N symbols resulting in a specific configuration where box i contains n_i atoms is given by the multinomial coefficient:

$$\frac{N!}{n_1!\,n_2!\ldots n_B!}, \quad \text{where } \sum_{i=1}^{B} n_i = N.$$

Hence the statistical weight (or relative frequency) of this configuration $\mathbf{n} = (n_1, n_2, \ldots, n_B)$ in the ensemble of random structures generated by the source defined by q^{ME} is

$$\mathscr{P}(\mathbf{n}) = \frac{N!}{n_1!\,n_2!\ldots n_B!}\, q_1^{n_1} q_2^{n_2} \ldots q_B^{n_B}.$$

Using Stirling's formula, we may approximate the logarithm of $\mathscr{P}(\mathbf{n})$ as

$$\log \mathscr{P}(\mathbf{n}) \simeq N \log N - N - \sum_{i=1}^{B} (n_i \log n_i - n_i) + \sum_{i=1}^{B} n_i \log q_i$$

$$= -N \sum_{i=1}^{B} (n_i/N) \log (n_i/N) + N \sum_{i=1}^{B} (n_i/N) \log q_i$$

or, finally,

$$\log \mathscr{P}(\mathbf{n}) \simeq -N \sum_{i=1}^{B} p_i \log (p_i/q_i),$$

where $p_i = n_i/N$ is the fraction of the atoms found in box i, *i.e.* the relative frequency of occurrence of symbol i in the original 'message'.

Putting $\mathbf{p} = (1/N)\mathbf{n}$, the final result reads

$$\mathscr{P}(\mathbf{p}) \sim \exp [N \mathscr{S}_{q^{\mathrm{ME}}}(\mathbf{p})] \tag{4.1}$$

where

$$\mathscr{S}_{q^{\mathrm{ME}}}(\mathbf{p}) = -\sum_{i=1}^{B} p_i \log (p_i/q_i) \tag{4.2}$$

is the entropy of distribution $\mathbf{p}$ relative to the prior q^{ME}, as already defined in § 3.2.

Generalizing to continuous distributions in the obvious way, we conclude that the logarithm of the statistical weight of a particular normalized density function or 'map' $p(\mathbf{x})$ depends on its entropy relative to the prior $q^{\mathrm{ME}}(\mathbf{x})$:

$$\mathscr{S}_{q^{\mathrm{ME}}}(p) = -\int_V p(\mathbf{x}) \log [p(\mathbf{x})/q^{\mathrm{ME}}(\mathbf{x})]\, \mathrm{d}^3\mathbf{x}. \tag{4.3}$$

If we now retrace our steps, it is clear that:

(*a*) given some prior knowledge, the conditional probability of a vector of structure factors measures its statistical weight in the ensemble of the source specified by the ME prior $q^{\mathrm{ME}}(\mathbf{x})$ incorporating that knowledge;

(*b*) the statistical weight of that vector is that of the corresponding map in the corresponding ensemble, because the Fourier transform is a 1–1 unitary (hence volume-preserving) transformation;

(*c*) the statistical weight of a map is a simple function of the entropy of that map relative to the prior $q^{\mathrm{ME}}(\mathbf{x})$.

We may therefore conclude that *the construction of conditional probability distributions of structure factors envisaged by Hauptman & Karle amounts precisely to a reciprocal-space evaluation of the entropy functional* $\mathscr{S}_{q^{\mathrm{ME}}}(p)$ (Bricogne, 1982). Entropy calculations thus afford a quantitative formulation of the qualitative geometric considerations put forward by Schenk (1981) to illustrate the intuitive basis of direct methods, since they provide an exact numerical estimate of how 'likely' various configurations of atoms are when some structure factors assume known values.

4.2. *Construction of c.p.d.'s by classical methods*

The ME prior $q^{\mathrm{ME}}(\mathbf{x})$ may be written

$$q^{\mathrm{ME}}(\mathbf{x}) = (1/V)[1 + u(\mathbf{x})]$$

with

$$u(\mathbf{x}) = \sum_{\mathbf{h} \neq \mathbf{0}} U_{\mathbf{h}}^{\mathrm{ME}} \exp [-2\pi i \mathbf{h} \,.\, \mathbf{x}]. \tag{4.4}$$

The spectrum of $q^{\mathrm{ME}}(\mathbf{x})$ is extrapolated by the process of entropy maximization beyond the data provided by the prior knowledge, as shown for example by (3.16) and (3.24). The c.p.d. of the vector $\mathbf{U}$ of yet unknown unitary structure factors will be centred around the vector $\mathbf{U}^{\mathrm{ME}}$ of extrapolated Fourier coefficients of q^{ME}.

4.2.1. *The central limit theorem approximation*

The deviation $\delta\mathbf{U} = \mathbf{U} - \mathbf{U}^{\mathrm{ME}}$, whose Fourier transform will be denoted by $\delta p(\mathbf{x})$, will have by the central limit theorem a Gaussian distribution obtained as the first term of the Edgeworth series. Using as indices a full Bragg sphere of non-origin reflexions $\mathbf{h}$, this Gaussian may be written

$$\mathscr{P}(\delta\mathbf{U}) \propto \exp\{-(1/2)N(\delta\mathbf{U})^H\mathbf{Q}^{-1}(\delta\mathbf{U})\}, \quad (4.5)$$

where $\mathbf{X}^H$ denotes the Hermitian transpose of a vector $\mathbf{X}$, and $\mathbf{Q}$ is the covariance matrix of $\delta\mathbf{U}$ (§ 3.3.3);

$$Q_{\mathbf{hk}} = U^{\mathrm{ME}}_{\mathbf{h}-\mathbf{k}} - U^{\mathrm{ME}}_{\mathbf{h}}U^{\mathrm{ME}}_{-\mathbf{k}}. \quad (4.6)$$

Thus $\mathbf{Q} = \mathbf{K} - \mathbf{W}$, where $\mathbf{K}$ is the Karle–Hauptman matrix of q^{ME} (which will be written $\mathbf{K} = \mathrm{K}H[q^{\mathrm{ME}}]$) and $\mathbf{W} = \mathbf{U}^{\mathrm{ME}}(\mathbf{U}^{\mathrm{ME}})^H$ is (up to a normalization factor) a projector.

It will now be shown that the quadratic form in the exponent of (4.5) may be simplified as follows:

$$(\delta\mathbf{U})^H\mathbf{Q}^{-1}(\delta\mathbf{U}) = (\delta\mathbf{U})^H\mathrm{K}^{-1}(\delta\mathbf{U}) \quad (4.7)$$

whenever $\delta U_{\mathbf{0}} = 0$, *i.e.* if $\delta\mathbf{U}$ does not alter the normalization of q^{ME}. For this purpose, recall that $\mathbf{K}$ is a convolution operator corresponding to pointwise multiplication by $q^{\mathrm{ME}}(\mathbf{x})$ in real space. Since $q^{\mathrm{ME}}(\mathbf{x})$ is always everywhere positive, $\mathbf{K}$ is invertible (Gohberg & Fel'dman, 1974) and

$$\mathbf{K}^{-1} = \mathrm{K}H[1/q^{\mathrm{ME}}]. \quad (4.8)$$

To prove (4.7) assume that $\mathbf{Y} = \mathbf{Q}^{-1}\mathbf{X}$, so that $\mathbf{X} = (\mathbf{K}-\mathbf{W})\mathbf{Y}$. Then

$$\mathbf{KY} = \mathbf{X} + \mathbf{WY}$$
$$\mathbf{Y} = \mathbf{K}^{-1}(\mathbf{X} + \mathbf{WY});$$

hence

$$\mathbf{Q}^{-1}\mathbf{X} = \mathbf{K}^{-1}\{\mathbf{X} + \mathbf{W}\mathbf{K}^{-1}[\mathbf{X} + \mathbf{W}\mathbf{K}^{-1}(\ldots)]\}.$$

This expansion is valid because $\mathbf{W}$ is a small enough perturbation of $\mathbf{K}$ that it does not alter its positive definite character. Therefore, putting $\boldsymbol{\Delta} = \delta\mathbf{U}$ and replacing $\mathbf{W}$ by $\mathbf{U}^{\mathrm{ME}}(\mathbf{U}^{\mathrm{ME}})^H$, we obtain

$$\begin{aligned}\boldsymbol{\Delta}^H\mathbf{Q}^{-1}\boldsymbol{\Delta} = {} & \boldsymbol{\Delta}^H\mathbf{K}^{-1}\boldsymbol{\Delta} \\ & + (\boldsymbol{\Delta}^H\mathbf{K}^{-1}\mathbf{U}^{\mathrm{ME}})((U^{\mathrm{ME}})^H\mathbf{K}^{-1}\boldsymbol{\Delta}) \\ & + (\boldsymbol{\Delta}^H\mathbf{K}^{-1}\mathbf{U}^{\mathrm{ME}})((U^{\mathrm{ME}})^H\mathbf{K}^{-1}\mathbf{U}^{\mathrm{ME}}) \\ & \times (U^{\mathrm{ME}})^H\mathbf{K}^{-1}\boldsymbol{\Delta}) + \ldots\end{aligned}$$

All terms but the first vanish if $(U^{\mathrm{ME}})^H\mathbf{K}^{-1}\boldsymbol{\Delta} = 0$, *i.e.* by Parseval's theorem if

$$\int_V q^{\mathrm{ME}}(\mathbf{x})(1/q^{\mathrm{ME}}(\mathbf{x}))\,\delta p(\mathbf{x})\,\mathrm{d}^3\mathbf{x} = 0$$

or equivalently $\delta U_{\mathbf{0}} = 0$. This proves identity (4.7), so that we may write

$$\mathscr{P}(\delta\mathbf{U}) \propto \exp\{-(\tfrac{1}{2})N(\delta\mathbf{U})^H\mathbf{K}^{-1}(\delta\mathbf{U})\}. \quad (4.9)$$

We may now use (4.8) and invoke Parseval's theorem to rewrite (4.9) as

$$\mathscr{P}(\delta p) \propto \exp\left\{-(\tfrac{1}{2})N\int_V \frac{[\delta p(\mathbf{x})]^2}{q^{\mathrm{ME}}(\mathbf{x})}\,\mathrm{d}^3\mathbf{x}\right\} \quad (4.10)$$

for any variation δp such that $\int_V \delta p(\mathbf{x})\,\mathrm{d}^3\mathbf{x} = 0$. This new expression has a very simple interpretation in real space. The role of the non-uniform prior $q^{\mathrm{ME}}(\mathbf{x})$ is to create a *differential cost* for the addition of new features at different locations in the unit cell: it is less costly, in terms of loss of conditional probability, to introduce new detail of given local root-mean-square value at places where q^{ME} is large than where it is small.

If q^{ME} results from the knowledge of some low-resolution structure factors, then carrying out phase extension by maximizing $\mathscr{P}(\delta p)$ would amount to constraining high-resolution features to appear in regions where density has already been laid out by the construction of q^{ME}. This procedure would thus perform, in a statistical fashion, the operation of 'auto-enveloping' of the electron density by its low-resolution level surfaces which has recently been investigated by Cannillo, Oberti & Ungaretti (1983).

The equivalence of (4.10) with the result previously obtained in § 4.1 by combinatorial methods [equations (4.1) and (4.2)] is readily established. The gradient of $\mathscr{S}_{q^{\mathrm{ME}}}(p)$ in the subspace of yet unspecified U's is zero at $\mathbf{U}^{\mathrm{ME}}$ by the very definition of q^{ME}, while its Hessian $\mathbf{H}(\mathscr{S})$ is defined by

$$\mathbf{H}(\mathscr{S}) = \left[\frac{\delta^2\mathscr{S}}{\delta p(\mathbf{x})\,\delta p(\mathbf{y})}\right] = [-[1/q^{\mathrm{ME}}(\mathbf{x})]\delta(\mathbf{x}-\mathbf{y})]$$

(where δ is Dirac's δ function) so that (4.10) follows from (4.1) by a local quadratic approximation of $\mathscr{S}_{q^{\mathrm{ME}}}(p)$ near $p = q^{\mathrm{ME}}$.

4.2.2. *Comparison with classical formulae*

In order to relate the expressions obtained for $\mathscr{P}$ to more standard ones in the literature, we may use the expansion

$$1/q^{\mathrm{ME}} = V[1 - u + u^2 - u^3 + \ldots] \quad (4.11)$$

valid if $|u(\mathbf{x})| < 1$ for all $\mathbf{x}$, and rewrite (4.10) as

$$\begin{aligned}\mathscr{P}(\delta\mathbf{U}) \propto \exp\Bigg\{ & -(\tfrac{1}{2})N\sum_{\mathbf{H}\neq\mathbf{0}}|\delta U_{\mathbf{H}}|^2 \\ & + (\tfrac{1}{2})N\sum_{\mathbf{H}\neq\mathbf{K}\neq\mathbf{0}}\overline{\delta U_{\mathbf{H}}}U^{\mathrm{ME}}_{\mathbf{H}-\mathbf{K}}\delta U_{\mathbf{K}} \\ & - (\tfrac{1}{2})N\sum_{\mathbf{H}\neq\mathbf{K}\neq\mathbf{L}\neq\mathbf{0}}\overline{\delta U_{\mathbf{H}}}U^{\mathrm{ME}}_{\mathbf{H}-\mathbf{K}}U^{\mathrm{ME}}_{\mathbf{K}-\mathbf{L}}\delta U_{\mathbf{L}} + \ldots\Bigg\}.\end{aligned} \quad (4.12)$$

It will now be shown that a complete theory of the estimation of invariants from their second-neighbourhood magnitudes may readily be derived by considering only the leading (Wilson) term of this expansion.

Grouping the contributions to this term from $\mathbf{H}$ and $-\mathbf{H}$ gives the expression

$$\mathscr{P}(\delta U) \propto \exp\left[-N \sum_{\mathbf{H}}' |U_{\mathbf{H}} - U_{\mathbf{H}}^{\mathrm{ME}}|^2\right] \qquad (4.13)$$

with summation over a hemisphere as in (3.9). Each factor of this product can be written (in normalized form)

$$2N|U_{\mathbf{H}}| \exp\{-N[|U_{\mathbf{H}}|^2 + |U_{\mathbf{H}}^{\mathrm{ME}}|^2 - 2|U_{\mathbf{H}}||U_{\mathbf{H}}^{\mathrm{ME}}| \cos(\varphi_{\mathbf{H}} - \varphi_{\mathbf{H}}^{\mathrm{ME}})]\} \qquad (4.14)$$

and may be treated in two different ways according as $\mathbf{H}$ is, in the terminology of § 1.2.3, a 'basis' or a 'satellite' reflexion for invariant Φ:

(1) if $\mathbf{H}$ is such that $\Phi = \varphi_{\mathbf{H}} - \varphi_{\mathbf{H}}^{\mathrm{ME}}$, (4.14) gives the c.p.d. of Φ as

$$\mathscr{P}(\Phi) = \frac{1}{2\pi I_0(X)} \exp(X \cos \Phi) \qquad (4.15)$$

with $X = 2N|U_{\mathbf{H}}||U_{\mathbf{H}}^{\mathrm{ME}}|$. This is essentially Sim's formula (Sim, 1959) since the denominator in the exponent of the latter would be $\sum_{j=1}^{N} f_j^2 = N(1/N)^2 = 1/N$;

(2) if $\mathbf{H}$ is such that $|U_{\mathbf{H}}^{\mathrm{ME}}|$ is sensitive to the value of Φ, (4.14) may be used to derive the marginal probability density for $|U_{\mathbf{H}}|$ by integration over $\varphi_{\mathbf{H}}$. This leads to a Rice distribution (Rice, 1944):

$$L(|U_{\mathbf{H}}|) = 2N|U_{\mathbf{H}}| \exp\{-N[|U_{\mathbf{H}}|^2 + |U_{\mathbf{H}}^{\mathrm{ME}}|^2]\} I_0(2N|U_{\mathbf{H}}||U_{\mathbf{H}}^{\mathrm{ME}}|) \qquad (4.16)$$

in which Φ is a parameter *via* $|U_{\mathbf{H}}^{\mathrm{ME}}|$. By a standard Bayesian argument (see Edwards, 1972), the quantity $L(|U_{\mathbf{H}}|^{\mathrm{obs}})$ affords the *a posteriori* probability or *likelihood* $\mathscr{P}_{\mathbf{H}}(\Phi)$ of Φ.

By means of these two simple devices, the distribution of any invariant Φ can be obtained in a swift and automatic fashion. The triplet case is simplest, since it involves only a Sim factor. If $U_{\mathbf{h}}$ and $U_{\mathbf{k}}$ are given, then by (3.23)

$$U_{\mathbf{h}+\mathbf{k}}^{\mathrm{ME}} = |U_{\mathbf{h}}||U_{\mathbf{k}}| \exp[i(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}})]$$

so that $\Phi = \varphi_{\mathbf{H}} - \varphi_{\mathbf{H}}^{\mathrm{ME}}$ for $\mathbf{H} = \mathbf{h} + \mathbf{k}$, and (4.15) yields the standard result. Invariants of arbitrary order $r > 3$, written $\Phi = \sum_{j=1}^{r} \varphi(\mathbf{h}_j)$ with $\sum_{j=1}^{r} \mathbf{h}_j = \mathbf{0}$, lead to more complex formulae. If $U(\mathbf{h}_1), \ldots, U(\mathbf{h}_{r-1})$ are given, then $\Phi = \varphi_{\mathbf{H}} - \varphi_{\mathbf{H}}^{\mathrm{ME}}$ for $\mathbf{H} = \mathbf{h}_r$ since

$$U(\mathbf{h}_r)^{\mathrm{ME}} = \prod_{j=1}^{r-1} |U(\mathbf{h}_j)| \exp\left[-i \sum_{j=1}^{r-1} \varphi(\mathbf{h}_j)\right].$$

By (4.15) there will be a Sim factor

$$\mathscr{P}_0(\Phi) = \frac{1}{2\pi I_0(X)} \exp(X \cos \Phi)$$

with

$$X = 2N \prod_{j=1}^{r} |U(\mathbf{h}_j)|. \qquad (4.17)$$

But the assignment of a particular phase $\varphi(\mathbf{h}_r)$ to $|U(\mathbf{h}_r)|$ will generate other sizeable contributions to (4.13) whose extrapolated magnitudes will depend on Φ, so that Rice factors (4.16) may be constructed to sharpen further the estimation of Φ. The indices $\mathbf{H}$ of these magnitudes constitute the second neighbourhood $\mathscr{N}_2(\Phi)$ of Φ (Hauptman, 1975*b*). No general rule has yet been given to identify $\mathscr{N}_2(\Phi)$, but such a rule emerges naturally from the present approach. Let J_1 and J_2 be two subsets of $\{1, 2, \ldots, r\}$ forming a *partition* of that set and containing at least two elements each, and define

$$\mathbf{H} = \sum_{j_1 \in J_1} \mathbf{h}_{j_1}, \qquad \mathbf{K} = \sum_{j_2 \in J_2} \mathbf{h}_{j_2}.$$

Then $\mathbf{H} + \mathbf{K} = \mathbf{0}$, so that

$$U_{\mathbf{H}}^{\mathrm{ME}} = \prod_{j_1 \in J_1} U(\mathbf{h}_{j_1}) + \prod_{j_2 \in J_2} U(-\mathbf{h}_{j_2}),$$

and hence the extrapolated modulus

$$|U_{\mathbf{H}}^{\mathrm{ME}}| = \left\{ \prod_{j_1 \in J_1} |U(\mathbf{h}_{j_1})|^2 + \prod_{j_2 \in J_2} |U(\mathbf{h}_{j_2})|^2 + 2 \prod_{j=1}^{r} |U(\mathbf{h}_j)| \cos \Phi \right\}^{1/2} \qquad (4.18)$$

depends on Φ and gives rise to a Rice factor

$$\mathscr{P}_{\mathbf{H}}(\Phi) \propto \exp\left\{-2N \left[\prod_{j=1}^{r} |U(\mathbf{h}_j)|\right] \cos \Phi\right\} \times I_0(2N|U_{\mathbf{H}}||U_{\mathbf{H}}^{\mathrm{ME}}|). \qquad (4.19)$$

The second neighbourhood $\mathscr{N}_2(\Phi)$ is thus in one–one correspondence with the distinct partitions (J_1, J_2), whose total number s is easily obtained:

$$\begin{aligned} s &= \frac{1}{2} \sum_{m=2}^{r-2} \binom{r}{m} \\ &= \frac{1}{2}\left[\sum_{m=0}^{r} \binom{r}{m} - \binom{r}{1} - \binom{r}{r-1} - \binom{r}{0} - \binom{r}{r}\right] \\ &= \tfrac{1}{2}[2^r - 2r - 2], \end{aligned}$$

i.e.

$$s = 2^{r-1} - r - 1. \qquad (4.20)$$

The final expression for $\mathscr{P}$ is thus

$$\mathscr{P}(\Phi) \propto \exp\left[-2(s-1)N \prod_{j=1}^{r} |U(\mathbf{h}_j)| \cos \Phi\right] \times \prod_{\mathbf{H} \in \mathscr{N}_2(\Phi)} I_0(2N|U_{\mathbf{H}}||U_{\mathbf{H}}^{\mathrm{ME}}|) \qquad (4.21)$$

each $\mathbf{H} \in \mathcal{N}_2(\Phi)$ being labelled by a partition (J_1, J_2) and the corresponding $|U_{\mathbf{H}}^{\mathrm{Me}}|$ being given by (4.18).

For $r = 4$, s is equal to 3 and (4.21) is the Hauptman quartet formula (Hauptman, 1975*a*). For $r = 5$ ($s = 10$) and $r = 6$ ($s = 25$), (4.21) is a sharper version of the results of Fortier & Hauptman (1977*a*, *b*) and Hauptman & Fortier (1977*a*, *b*). The ease with which (4.21) has been obtained here for arbitrary r may be contrasted with the exceedingly lengthy calculations found in the papers quoted for $r = 4, 5, 6$.

The mechanism used is very simple, and consists essentially of a maximum entropy extrapolation followed by a recourse to Wilson's statistics. It is straightforward to verify that the further terms in expansion (4.12) give rise respectively to the third and higher neighbourhoods of a given invariant, and that space-group symmetries could readily be incorporated by analysing the structure of $\mathbf{U}^{\mathrm{ME}}$ according to the methods of § 3.5.3. It is therefore clear that unlimited numbers of new formulae of forbidding complexity could be derived *ad libitum* by a systematic dissection of (4.12). Little would be gained, however, by pursuing these formal developments indefinitely for the purpose of estimating single invariants, in view of the remarks made in § 2.2. Rather, (4.12) should be used to derive the conditional distribution of several phase invariants simultaneously. Yet for numerical applications (§ 7) (4.10) is much more convenient: it acts as a generating function for all the families of invariants involved, and hence affords a means of manipulating them and of summing over them in closed form without having to explicitly enumerate and collect large numbers of terms.

Finally, it may be remarked that (4.11) diverges as soon as $|u(\mathbf{x})|$ exceeds 1 for some $\mathbf{x}$, *i.e.* as soon as the prior knowledge incorporated into $q^{\mathrm{ME}}(\mathbf{x})$ demands that q^{ME} should deviate from its mean value $1/V$ by more than that value. The domain of convergence of (4.12) shrinks accordingly, and the first-neighbourhood approximation (4.21) becomes poorer. This remark further illustrates the convergence problems analysed in § 2.1.

4.3. *Multimodality and branching*

The exact form (4.1) of $\mathcal{P}$ is the exponential of a relative entropy

$$\mathcal{P}(\delta p) \sim \exp\left[N\mathcal{S}_{q^{\mathrm{ME}}}(q^{\mathrm{ME}} + \delta p)\right] \qquad (4.22)$$

and it was established in § 4.2.1 that the Hessian of $\mathcal{S}_{q^{\mathrm{ME}}}$ is always negative definite. Therefore, $\mathcal{S}_{q^{\mathrm{ME}}}$ is a *globally concave* functional of its argument with a unique maximum at $\delta p = 0$, and the same holds for $\mathcal{P}$ if its argument is left unconstrained.

Once a c.p.d. $\mathcal{P}$ has been constructed, however, its exploitation usually entails its maximization under additional constraints provided by the knowledge of some yet unphased moduli. For M acentric reflexions, for instance, the geometric locus of points satisfying the moduli constraints is a product of M circles, *i.e.* an M-dimensional torus embedded in $2M$-dimensional space. Such a manifold is multiply connected, and the distance of a point to that manifold does not define a convex function: as a result, the entropy functional $\mathcal{S}_{q^{\mathrm{ME}}}$, in spite of its global concavity, will possess a multitude of local maxima when restricted to that manifold. In other words, the conditional distribution $\mathcal{P}(\delta\mathbf{U})$ becomes highly multimodal when moduli constraints on the components of $\delta\mathbf{U}$ are enforced. Under these circumstances, any phase-extension procedure based on the use of conditional distributions (§ 4.2.1.) will necessarily encounter ambiguities – a phenomenon which will be called the 'branching problem'.

Direct methods deal with this problem by means of the 'phase function' of Riche (1970, 1973), supplemented by the device of 'magic integers' (White & Woolfson, 1975), which allows a coarse global survey of the multimodality of $\mathcal{P}$ to be performed in about a dozen directions at a time. But this method uses an approximation of $\mathcal{P}$ by a product of triplet distributions (§ 2.2), which is of limited value. An algorithm for locally exploring the full conditional distribution $\mathcal{P}$ in several hundred directions simultaneously will be presented in § 7.2.

Structures resulting from phase extension along an incorrect branch (*i.e.* by picking the wrong constrained maximum of a c.p.d. $\mathcal{P}$) will be compatible with the Patterson function and hence will frequently exhibit some correct clusters of high-resolution features; in view of the real-space interpretation of a c.p.d. (§ 4.2.1) these correct fragments must lie in regions where the low-resolution prior q^{ME} is large, in order to produce a local maximum of $\mathcal{P}$, but some may be placed in a wrong region of high prior probability. Thus, the well-documented artefact of direct methods whereby correct molecular fragments are occasionally found misplaced (Karle, 1976) is a manifestation of the branching problem.

5. Maximum entropy and the saddlepoint approximation

An alternative derivation of all the results of §§ 3 and 4 will now be presented which is totally independent of the concept of entropy and hence of Jaynes's heuristics. It uses an analytical device, the 'saddle-point method', first introduced into the statistical literature by Daniels (1954), to construct directly an optimal approximation to the Fourier integral (1.10). The results obtained will be shown to be identical to that defined by (4.5) [or (4.10)] and (ME1) to (ME3), so that the two-step process described above will receive a final and rigorous justification.

The mathematical origins of the saddlepoint method go back to the work of de Laplace (1820), and its systematic study is mainly associated with the names of Riemann and Debye (see Copson, 1965; de Bruijn, 1970; Erdélyi, 1956). Its use in statistical physics by Darwin and Fowler (Fowler, 1936), later clarified by Khinchin (1949), is intimately related to the present work (§ 5.4). In its applications to probability theory, the saddlepoint method affords a technique for obtaining consistently optimal asymptotic expansions for distributions of sums of independent identically distributed random variables over a wide range of values of their arguments, and not only near the sum of their first moments. Daniels's original treatment deals with one-dimensional problems but its generalization to the multidimensional case, which is needed here, offers little difficulty.

The saddlepoint method is best understood by reference to the construction of the Edgeworth series outlined in § 1.2.2 in which the cumulant-generating function $\log M$ was expanded as a power series in the carrying variables $\mathbf{u}$ about the origin $\mathbf{u}_0 = \mathbf{0}$, leading to an asymptotic expansion for $\mathcal{P}(\mathbf{F})$ valid only near its centre $\mathbf{F}^0 = N\mathbf{U}_q$. The essence of Daniels's method is to let the choice of point $\mathbf{u}_0$ (about which $\log M$ is to be expanded) be dependent on the value of $\mathbf{F}$ for which an accurate evaluation of $\mathcal{P}(\mathbf{F})$ is being sought. The criterion used in this choice is that all substantial contributions to integral (1.10) should become maximally concentrated in a small neighbourhood of that point as $N \to \infty$. The optimal value of $\mathbf{u}_0$ turns out to be complex, so that some preliminary technical results relative to the analytical continuation of $\log M$ must first be established.

5.1. *Analytic continuation of Fourier transforms*

Let f be an absolutely integrable complex-valued function of m real variables, *i.e.*

$$f = f(\mathbf{s}), \qquad \mathbf{s} = (s_1, s_2, \ldots, s_m),$$

and let Φ be its inverse Fourier transform:

$$\Phi(\mathbf{t}) = \bar{\mathcal{F}}[f](\mathbf{t}) = \int_{\mathbb{R}^m} f(\mathbf{s}) \exp(i\mathbf{t} \cdot \mathbf{s})\, \mathrm{d}^m\mathbf{s} \qquad (5.1)$$

(note the slightly different normalization, in that there is no factor of 2π in the exponent).

Suppose f has *compact support*, that is, vanishes identically outside a closed bounded subset of $\mathbb{R}^m$. Then, by the Paley–Wiener theorem (Paley & Wiener, 1934, p. 12), Φ may be analytically continued into a function of m *complex* variables, analytic in the whole of $\mathbb{C}^m$. This theorem was generalized by Schwartz (1966, p. 272) to the case where f is a distribution (in particular, a measure) with compact support.

The result of this analytic continuation is described explicitly by

$$\begin{aligned}\Phi(\mathbf{t} - i\boldsymbol{\tau}) &= \int_{\mathbb{R}^m} f(\mathbf{s}) \exp[i(\mathbf{t} - i\boldsymbol{\tau}) \cdot \mathbf{s}]\, \mathrm{d}^m\mathbf{s} \\ &= \int_{\mathbb{R}^m} f(\mathbf{s}) \exp(\boldsymbol{\tau} \cdot \mathbf{s}) \exp(i\mathbf{t} \cdot \mathbf{s})\, \mathrm{d}^m\mathbf{s} \\ &= \bar{\mathcal{F}}[f \exp(\boldsymbol{\tau} \cdot \mathbf{s})](\mathbf{t})\end{aligned}$$

or

$$\bar{\mathcal{F}}[f](\mathbf{t} - i\boldsymbol{\tau}) = \bar{\mathcal{F}}[\exp(\boldsymbol{\tau} \cdot \mathbf{s}) f](\mathbf{t}). \qquad (5.2)$$

Thus, adding to each t_j an imaginary part $-\tau_j$ ($j = 1, \ldots, m$) is reflected by an exponential modulation of the original function f.

For every $\mathbf{t}$, $\Phi(\mathbf{t} - i\boldsymbol{\tau})$ considered as a function of $\boldsymbol{\tau}$ is of exponential type in its growth properties (Schwartz, 1966, p. 271). In particular, if the (compact) support of f spans both strictly positive and strictly negative values in every direction in $\mathbf{s}$ space (which we will abbreviate by saying that f is of class E) then $|\Phi(\mathbf{t} - i\boldsymbol{\tau})|$ grows exponentially in all directions in $\boldsymbol{\tau}$ space for any fixed $\mathbf{t}$.

The original function $f(\mathbf{s})$ may be recovered from Φ by the usual inversion formula:

$$f(\mathbf{s}) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \Phi(\mathbf{t}) \exp(-i\mathbf{t} \cdot \mathbf{s})\, \mathrm{d}^m\mathbf{t}. \qquad (5.3)$$

But the integrand may be continued analytically without singularities into $\mathbb{C}^m$, and by the m-dimensional generalization of Cauchy's theorem (Hörmander, 1973) the above integral is not changed by giving to each integration variable t_j an arbitrary imaginary part $-\tau_j$. Therefore,

$$f(\mathbf{s}) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \Phi(\mathbf{t} - i\boldsymbol{\tau}) \exp[-i(\mathbf{t} - i\boldsymbol{\tau}) \cdot \mathbf{s}]\, \mathrm{d}^m\mathbf{t} \qquad (5.4)$$

for any $\boldsymbol{\tau} \in \mathbb{R}^m$ (see for instance Rudin, 1973, p. 180).

5.2. *Families of conjugate distributions*

Let f now be a probability measure with compact support in $\mathbb{R}^m$, and let its inverse Fourier transform Φ be extended analytically by (5.2). Then the usual characteristic function $C(\mathbf{T})$ and moment-generating function $M(\boldsymbol{\tau})$ associated with f (see Klug, 1958) are defined in terms of Φ by

$$\begin{cases} C(\mathbf{T}) = \Phi(\mathbf{T}), & \mathbf{T} \in \mathbb{R}^m \qquad (5.5a) \\ M(\boldsymbol{\tau}) = \Phi(-i\boldsymbol{\tau}), & \boldsymbol{\tau} \in \mathbb{R}^m. \qquad (5.5b) \end{cases}$$

These formulae may be used to extend functions C and M themselves into functions of m complex

variables by putting

$$\begin{cases} C(\mathbf{T}+i\boldsymbol{\tau}) = \Phi(\mathbf{T}+i\boldsymbol{\tau}) & (5.6a) \\ M(\boldsymbol{\tau}+i\mathbf{T}) = \Phi(\mathbf{T}-i\boldsymbol{\tau}). & (5.6b) \end{cases}$$

The positivity of $f(\mathbf{s})$, and hence of $f(\mathbf{s})\exp(\boldsymbol{\tau}\,.\,\mathbf{s})$ for all $\boldsymbol{\tau}\in\mathbb{R}^m$, implies by (5.2) that $|\Phi(\mathbf{T}-i\boldsymbol{\tau})|$ is always a maximum for $\mathbf{T}=\mathbf{0}$ for any $\boldsymbol{\tau}$; or equivalently by (5.6*b*) that $|M(\boldsymbol{\tau}+i\mathbf{T})|$ always has a maximum on the m-fold product of real axes $\mathbf{T}=\mathbf{0}$ [M is a 'ridge function' (Lukacs, 1983)]. In this subspace, $M(\boldsymbol{\tau})$ is always real-valued and positive; furthermore, as was noted in § 5.1, $M(\boldsymbol{\tau})$ is of exponential growth if f is of class E.

Under these hypotheses we may define, for any $\boldsymbol{\tau}\in\mathbb{R}^m$,

$$f_{\boldsymbol{\tau}}(\mathbf{s}) = \frac{\exp(\boldsymbol{\tau}\,.\,\mathbf{s})f(\mathbf{s})}{\int_{\mathbb{R}^m}\exp(\boldsymbol{\tau}\,.\,\mathbf{s})f(\mathbf{s})\,\mathrm{d}^m\mathbf{s}} = \frac{1}{M(\boldsymbol{\tau})}\exp(\boldsymbol{\tau}\,.\,\mathbf{s})f(\mathbf{s}), \quad (5.7)$$

which is also a probability measure with compact support. The family $\{f_{\boldsymbol{\tau}}(\mathbf{s})\}_{\boldsymbol{\tau}\in\mathbb{R}^m}$ is called a family of *conjugate* distributions (Khinchin, 1949, p. 79). It is clear from (5.2) that the characteristic functions of the $f_{\boldsymbol{\tau}}$ are all related by analytic continuation.

Using the moment-generating properties of M, it is straightforward to calculate the first moments $\langle s_j\rangle(\boldsymbol{\tau}_0)$ and covariances $r_{jk}(\boldsymbol{\tau}_0)$ of conjugate distribution $f_{\boldsymbol{\tau}_0}(\mathbf{s})$:

$$\langle s_j\rangle(\boldsymbol{\tau}_0) = \frac{\partial}{\partial\tau_j}(\log M)|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} \quad (5.8a)$$

$$r_{jk}(\boldsymbol{\tau}_0) = \langle(s_j-\langle s_j\rangle)(s_k-\langle s_k\rangle)\rangle(\boldsymbol{\tau}_0) = \frac{\partial^2}{\partial\tau_j\,\partial\tau_k}(\log M)|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0}. \quad (5.8b)$$

Since f is of class E, it is non-degenerate (*i.e.* it is not concentrated in any strict linear subspace of $\mathbb{R}^m$), hence $[r_{jk}(\boldsymbol{\tau}_0)]$ is positive definite for all $\boldsymbol{\tau}_0$. Therefore, $\log M(\boldsymbol{\tau})$ is a strictly convex function, and consequently its gradient will assume every value in its range once and only once. By (5.8*a*) this range consists of all the first moments of the family of conjugate distributions, and it can readily be shown that these consist of all the interior points of the convex hull of the support (ICS) of f (see Rockafellar, 1970). Thus a one-to-one correspondence between the vectors $\langle\mathbf{s}\rangle\in$ ICS f and $\boldsymbol{\tau}\in\mathbb{R}^m$ is set up by the equation

$$\nabla(\log M) = \langle\mathbf{s}\rangle. \quad (5.9)$$

In other words (Khinchin, 1949), for any vector $\mathbf{s}_0\in$ ICS f, there exists a *unique* member $f_{\boldsymbol{\tau}_0}$ of the family of conjugate distributions $\{f_{\boldsymbol{\tau}}\}_{\boldsymbol{\tau}\in\mathbb{R}^m}$ which has mathematical expectation $\mathbf{s}_0$.

5.3. *The saddlepoint approximation*

In the standard probabilistic formalism, reviewed in § 1.2, the main source of difficulty resides with the step summarized in (1.10), *i.e.* calculating the Fourier transform of the Nth power of the characteristic function C of the distribution $P(\mathbf{X})$ of random vector $\mathbf{X}$.

Let $q(\mathbf{x})$ be the prior distribution of atoms. Let the set of reflexions $h_1,\ldots,h_n$ comprise m_a acentric and m_c centric reflexions, and let $m=2m_a+m_c$. Let $\mathbf{X}(\mathbf{x})$ denote the m-dimensional random vector of real components of the contributions to $F_{\mathbf{h}_1},\ldots,F_{\mathbf{h}_n}$ of a point atom of unit weight placed at $\mathbf{x}$. Then the distribution $P(\mathbf{X})$ of $\mathbf{X}$ has compact support and is of class E. If C and M denote its characteristic and moment-generating functions, respectively, then

$$C(\mathbf{t}) = M(i\mathbf{t}) = \int_{\mathbb{R}^m} P(\mathbf{X})\exp(i\mathbf{t}\,.\,\mathbf{X})\,\mathrm{d}^m\mathbf{X} = \int_V q(\mathbf{x})\exp[i\mathbf{t}\,.\,\mathbf{X}(\mathbf{x})]\,\mathrm{d}^3\mathbf{x}. \quad (5.10)$$

Equation (1.10) may therefore be written in full as

$$\begin{aligned}\mathscr{P}(\mathbf{F}) &= \mathscr{F}[C^N](\mathbf{F}) \\ &= \frac{1}{(2\pi)^m}\int_{\mathbb{R}^m} C^N(\mathbf{t})\exp(-i\mathbf{t}\,.\,\mathbf{F})\,\mathrm{d}^m\mathbf{t} \\ &= \frac{1}{(2\pi)^m}\int_{\mathbb{R}^m} M^N(i\mathbf{t})\exp(-i\mathbf{t}\,.\,\mathbf{F})\,\mathrm{d}^m\mathbf{t} \\ &= \frac{1}{(2\pi)^m}\int_{\mathbb{R}^m}\exp\{N[\log M(i\mathbf{t})-i\mathbf{U}\,.\,\mathbf{t}]\}\,\mathrm{d}^m\mathbf{t}\end{aligned} \quad (5.11)$$

where $\mathbf{U}=(1/N)\mathbf{F}=(1/N)\sum_{j=1}^N\mathbf{X}_j$ is the vector of unitary structure factors, treated as an m-dimensional real vector as far as scalar products are concerned.

The fundamental problem in the probabilistic formalism of § 1.2 is to find a good approximation of $\mathscr{P}$ for arbitrary values of $\mathbf{U}$ which may not lie in the immediate vicinity of the first-order moments $\mathbf{U}_q$ of P, which are the Fourier coefficients of $q(\mathbf{x})$. When $N\to\infty$, the main contribution to the integral is concentrated near $\mathbf{t}=\mathbf{0}$ since, as noted previously, $M(i\mathbf{t})$ has maximum modulus there. One may therefore expand $\log M(i\mathbf{t})$ as a power series near $\mathbf{t}=\mathbf{0}$ and proceed as described in § 1.2.2 to obtain the Edgeworth series of $\mathscr{P}(\mathbf{F})$. As discussed at length in § 2.2, this series is only accurate near $\mathbf{F}=N\mathbf{U}_q$.

For a given value of $\mathbf{F}$, however, we may use the extra freedom afforded by (5.4), and assign to $\mathbf{t}$ any imaginary part without changing the value of the

integral. Changing $\mathbf{t}$ to $\mathbf{t}-i\boldsymbol{\tau}$ gives

$$\mathcal{P}(\mathbf{F}) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \exp\{N[\log M(\boldsymbol{\tau}+i\mathbf{t}) - i\mathbf{U}\,.\,(\mathbf{t}-i\boldsymbol{\tau})]\}\,\mathrm{d}^m\mathbf{t}$$

$$= \exp(-\mathbf{F}\,.\,\boldsymbol{\tau}) \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \exp\{N[\log M(\boldsymbol{\tau}+i\mathbf{t}) - i\mathbf{U}\,.\,\boldsymbol{\tau}]\}\,\mathrm{d}^m\mathbf{t}. \tag{5.12}$$

For any given $\boldsymbol{\tau}$, $M(\boldsymbol{\tau}+i\mathbf{t})$ still has maximum modulus at $\mathbf{t}=\mathbf{0}$, so that an Edgeworth-type series can be constructed. But it is possible, by appropriate choice of $\boldsymbol{\tau}$, to further ensure optimal concentration to a neighbourhood of that point of all significant contributions to the integral. Assuming that $\mathbf{F}$ is feasible, *i.e.* that $\mathbf{U}$ belongs to the interior of the convex hull of the support of P, we may choose for $\boldsymbol{\tau}$ the unique value $\boldsymbol{\tau}_0$ such that:

$$\nabla(\log M)|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0,\mathbf{t}=\mathbf{0}} = \mathbf{U} \tag{5.13}$$

whose existence was proved in § 5.2 (equation 5.9). Then not only will the modulus of the integrand be maximum at that point, but its phase will be stationary since

$$\nabla[\log M(\boldsymbol{\tau}_0+i\mathbf{t}) - i\mathbf{U}\,.\,\mathbf{t}]|_{\mathbf{t}=\mathbf{0}} = \mathbf{U}-\mathbf{U} = \mathbf{0}. \tag{5.14}$$

This value $\boldsymbol{\tau}_0$ of $\boldsymbol{\tau}$ is called the *saddlepoint* of the integrand since at that point the modulus of the latter is simultaneously a maximum in the $\mathbf{t}$ subspace and a minimum in the $\boldsymbol{\tau}$ subspace. Its defining equation (5.13) is called the *saddlepoint equation.*

Expansion of $\log M$ to second order near $\boldsymbol{\tau}_0$ gives

$$\log M(\boldsymbol{\tau}_0+i\mathbf{t}) \simeq \log M(\boldsymbol{\tau}_0) + i\mathbf{U}\,.\,\mathbf{t} - \tfrac{1}{2}\sum_{j,k=1}^{m} H_{jk}(\boldsymbol{\tau}_0)t_jt_k, \tag{5.15}$$

where $\mathbf{H}(\boldsymbol{\tau}_0) = [H_{jk}(\boldsymbol{\tau}_0)]$ is the Hessian matrix of $\log M$ at $\boldsymbol{\tau}_0$. We may thus write

$$\mathcal{P}(\mathbf{F}) \simeq \exp\{N[\log M(\boldsymbol{\tau}_0) - \mathbf{U}\,.\,\boldsymbol{\tau}_0]\} \times \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \exp\left\{-(N/2)\sum_{j,k=1}^{m} H_{jk}(\boldsymbol{\tau}_0)t_jt_k\right\}\mathrm{d}^m\mathbf{t}.$$

The evaluation of the integral is standard (Cramér, 1946), so that finally

$$\mathcal{P}(\mathbf{F}) \simeq \{N/[(2\pi)^m \det \mathbf{H}(\boldsymbol{\tau}_0)]\}^{1/2} \times \exp\{N[\log M(\boldsymbol{\tau}_0) - \mathbf{U}\,.\,\boldsymbol{\tau}_0]\}. \tag{5.16}$$

This expression was called by Daniels the *saddlepoint approximation* to $\mathcal{P}(\mathbf{F})$. It is only the first term of an asymptotic expansion, but it is as accurate in the neighbourhood of $\mathbf{F} = N\mathbf{U}$ [where $\mathbf{U}$ and $\boldsymbol{\tau}_0$ are related by the saddlepoint equation (5.13)] as the first term of the Edgeworth series near its centre, the latter corresponding to the particular case where $\boldsymbol{\tau}_0 = \mathbf{0}$.

The saddlepoint approximation does therefore provide a means of constructing consistently accurate estimates of the j.p.d. of a set of structure factors in the vicinity of any prescribed feasible values of its arguments.

5.4. *Relation to the maximum entropy method*

It will now be shown that the results just obtained are exactly identical to those of the ME method presented in §§ 3 and 4.

It is clear that the principle of the saddlepoint approximation consists in using, when calculating integral (5.11), not the original distribution $P(\mathbf{X})$ itself, but the conjugate distribution

$$P_{\boldsymbol{\tau}_0}(\mathbf{X}) = [\exp(\boldsymbol{\tau}_0\,.\,\mathbf{X})/M(\boldsymbol{\tau}_0)]P(\mathbf{X}) \tag{5.17}$$

whose characteristic function is

$$M(\boldsymbol{\tau}_0+i\mathbf{t}) = \int_{\mathbb{R}^m} P_{\boldsymbol{\tau}_0}(\mathbf{X}) \exp(i\mathbf{t}\,.\,\mathbf{X})\,\mathrm{d}^m\mathbf{X}$$

$$= \frac{1}{M(\boldsymbol{\tau}_0)} \int_{\mathbb{R}^m} \exp(\boldsymbol{\tau}_0\,.\,\mathbf{X})P(\mathbf{X}) \exp(i\mathbf{t}\,.\,\mathbf{X})\,\mathrm{d}^m\mathbf{X}$$

$$= \frac{1}{M(\boldsymbol{\tau}_0)} \int_{V} \exp[\boldsymbol{\tau}_0\,.\,\mathbf{X}(\mathbf{x})]q(\mathbf{x}) \exp[i\mathbf{t}\,.\,\mathbf{X}(\mathbf{x})]\,\mathrm{d}^3\mathbf{x}.$$

This may be rewritten

$$M(\boldsymbol{\tau}_0+i\mathbf{t}) = \int_V q_{\boldsymbol{\tau}_0}(\mathbf{x}) \exp[i\mathbf{t}\,.\,\mathbf{X}(\mathbf{x})]\,\mathrm{d}^3\mathbf{x} \tag{5.18}$$

where

$$q_{\boldsymbol{\tau}_0}(\mathbf{x}) = \frac{q(\mathbf{x})}{M(\boldsymbol{\tau}_0)} \exp[\boldsymbol{\tau}_0\,.\,\mathbf{X}(\mathbf{x})]. \tag{SP1}$$

On the other hand, it will be recalled that, by (5.10),

$$M(\boldsymbol{\tau}_0) = \int_V q(\mathbf{x}) \exp[\boldsymbol{\tau}_0\,.\,\mathbf{X}(\mathbf{x})]\,\mathrm{d}^3\mathbf{x} \tag{SP2}$$

and that the value of $\boldsymbol{\tau}_0$ is determined by (5.13):

$$\frac{\partial}{\partial\tau_j}(\log M)|_{\boldsymbol{\tau}=\boldsymbol{\tau}_0} = U_j. \tag{SP3}$$

Comparison of the latter three equations with (ME1), (ME2) and (ME3) in § 3.3.1 shows complete identity, up to an obvious change in notation:

the Z function (ME2) is the M function (SP2),

which explains its generating properties pointed out in § 3.3.3;

the Lagrange multipliers defined by (ME3) are the coordinates of the saddlepoint defined by (SP3);

the saddlepoint approximation (5.16) near $\mathbf{F} = N\mathbf{U}$ amounts to a normal approximation of the j.p.d. constructed on the basis not of the initially given prior $q(\mathbf{x})$, but of a revised prior distribution (SP1) identical to the maximum entropy distribution (ME1) consistent with the assumed value of $\mathbf{U}$ under prior prejudice $q(\mathbf{x})$.

Furthermore, comparison of (5.16) with (3.6) and (4.1) shows that the saddlepoint approximation is nothing other than

$$\mathscr{P}(\mathbf{F}) \simeq \exp\left[N\mathscr{S}_q(q_{\tau_0})\right] \tag{5.19}$$

as had been derived by combinatorial arguments in § 4.1.

Finally, the conclusions reached in § 5.2 regarding the uniqueness and the condition of existence of a solution to the saddlepoint equations (5.9) settle the same question for the ME equations.

This rather striking result is not without precedent. The analytic proof of the maximum-entropy principle given here is an exact parallel of the Darwin–Fowler formulation of statistical mechanics (Fowler, 1936), which also bypasses the established route – explicit combinatorial enumerations followed by an appeal to Stirling's formula – by means of a direct saddlepoint approximation to the thermodynamic partition function. Both results can ultimately be rationalized by recalling that Stirling's formula itself can be established by a saddlepoint approximation to Euler's Γ function (de Bruijn, 1970).

5.5. *Approximation of conditional distributions*

The conclusions obtained in this section may be rephrased in the context of the problem, formulated in §§ 2.1 and 2.3, of optimally approximating c.p.d.'s of structure factors. They are embodied in the following.

Theorem: The recentring of a joint probability distribution $\mathscr{P}$ around an arbitrary feasible vector $\mathbf{F}$, a necessary prerequisite to obtaining accurate conditional distributions by specialization in the vicinity of $\mathbf{F}$, may be accomplished by using Daniels's saddlepoint approximation. The latter procedure is mathematically equivalent to updating the prior distribution of atoms to the maximum-entropy prior compatible with the assumed value of $\mathbf{F}$, then constructing a normal approximation to $\mathscr{P}$ from this non-uniform prior.

This theorem completes the proof that the two-step approximation scheme for conditional distributions, outlined in § 2.3 and developed in §§ 3 and 4, constitutes a uniquely defined optimal implementation of the basic principles of probabilistic direct methods.

6. Maximum entropy and determinantal methods

Harker & Kasper (1948) deduced from the positivity of the electron density a set of algebraic inequalities involving the structure factors, which were later shown by Karle & Hauptman (1950) to be the simplest members of a large family of determinantal inequalities. The same determinants were subsequently used by Tsoucaris (1970) to devise an alternative method of constructing approximate joint distributions of structure factors when prior knowledge of some of them is assumed: the maximum determinant method (MDM for short).

Under both guises – algebraic and probabilistic – determinants have close ties with the maximum-entropy method (MEM). It will be shown in this section that the ME equations are solvable if and only if the initial data satisfy the Karle–Hauptman inequalities (§ 6.1.2), and that the asymptotic behaviour of determinants of large order defines an entropy-like quantity which is related to the entropy of j.p.d.'s rather than maps (§ 6.3). The relative merits of the Jaynes ME method and of the MDM will be compared in § 6.4.

6.1. *Determinantal inequalities*

6.1.1. *Origin*

The positivity of $p(\mathbf{x})$ is equivalent to the positive definiteness of the Hermitian form:

$$f \mapsto T_p(f) = \int_V \overline{f(\mathbf{x})} p(\mathbf{x}) f(\mathbf{x})\, \mathrm{d}^3\mathbf{x}, \tag{6.1}$$

where f is any square-summable periodic complex-valued function with the same period lattice as p. This in turn is equivalent to the positive definiteness of the restriction of T_p to the linear span of $[\exp(2\pi i \mathbf{h}_1 \cdot \mathbf{x}), \ldots, \exp(2\pi i \mathbf{h}_n \cdot \mathbf{x})]$ for any $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$, in which T_p is represented by the matrix

$$[T_p]_{ij} = [U_{\mathbf{h}_i - \mathbf{h}_j}] \quad (i, j = 1, 2, \ldots, n). \tag{6.2}$$

Hence the determinants of all such matrices must be positive (Karle & Hauptman, 1950). Kitaigorodskii (1961) showed, by elementary operations on rows and columns, that a Karle–Hauptman (KH for short) determinant depends only on invariant combinations of the phases of its component structure factors. An alternative proof may be given directly from the definition:

$$\begin{aligned} \det[T_{ij}] &= \sum_{\sigma \in S_n} \varepsilon(\sigma) T_{1,\sigma(1)} T_{2,\sigma(2)} \cdots T_{n,\sigma(n)} \\ &= \sum_{\sigma \in S_n} \varepsilon(\sigma) U_{\mathbf{h}_1 - \mathbf{h}_{\sigma(1)}} U_{\mathbf{h}_2 - \mathbf{h}_{\sigma(2)}} \cdots U_{\mathbf{h}_n - \mathbf{h}_{\sigma(n)}}, \end{aligned} \tag{6.3}$$

where the summation is over all permutations σ of the symmetric group on n letters S_n, and $\varepsilon(\sigma)$ is the signature of permutation σ. Since any permutation is

a product of disjoint circular permutations (Carmichael, 1937), each term of this sum factors into terms corresponding to cyclic permutations, whose phases are obviously phase invariants.

6.1.2. *Solvability condition for the ME equations*

Let M structure-factor values be given, and let q^{ME} be the corresponding ME prior. Fourier analysis of q^{ME} according to (4.4) then yields extrapolated Fourier coefficients $U_{\mathbf{h}}^{\mathrm{ME}}$ which are such that $(U_{\mathbf{h}_j-\mathbf{h}_k})^{\mathrm{ME}} = U_{\mathbf{h}_j-\mathbf{h}_k}$ if $\mathbf{h}_j-\mathbf{h}_k$ is also one of the M reflexions for which a constraint value has been given. Suppose then that a subset of m reflexions may be chosen (say $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m$) so that all reflexions of the form $\mathbf{h}_j-\mathbf{h}_k$ $(j, k = 1, 2, \ldots, m)$ belong to the initial set of M constraint reflexions, up to Friedel equivalence. Then the $m \times m$ determinant

$$\delta_m = \det\left[U_{\mathbf{h}_j-\mathbf{h}_k} - U_{\mathbf{h}_j}\overline{U_{\mathbf{h}_k}}\right] \tag{6.4}$$

is necessarily positive, since by (3.8) and (4.6) it is a Gram determinant. But δ_m is equal to the $(m+1)\times(m+1)$ determinant

$$d_{m+1} = \begin{vmatrix} 1 & U(-\mathbf{h}_1) & U(-\mathbf{h}_2) & . & U(-\mathbf{h}_m) \\ U(\mathbf{h}_1) & 1 & U(\mathbf{h}_1-\mathbf{h}_2) & . & U(\mathbf{h}_1-\mathbf{h}_m) \\ U(\mathbf{h}_2) & U(-\mathbf{h}_1+\mathbf{h}_2) & 1 & . & . \\ . & . & . & . & . \\ U(\mathbf{h}_m) & U(-\mathbf{h}_1+\mathbf{h}_m) & . & . & 1 \end{vmatrix} \tag{6.5}$$

since it may be obtained from d_{m+1} by the following operations [called the 'Chio pivotal condensation process', see Eves (1966)]: subtract from the $(j+1)$th column the product by $U_{\mathbf{h}_j}$ of the first column, for $j = 1, 2, \ldots, m$; then expand d_{m+1} along its first row. Consequently, if a negative KH determinant d_{m+1} can be constructed from the initial data, the ME equations cannot be solved since the positivity of δ_m and the relation $\delta_m = d_{m+1}$ are then contradictory. The KH inequalities are thus necessary conditions for the ME equations to be solvable. They are also sufficient conditions, since they imply that the vector $\mathbf{U} = (U_{\mathbf{h}_1}, \ldots, U_{\mathbf{h}_m})$ of given first-order moments is an interior point of the convex hull of the support of $P(\mathbf{X})$ (see § 5.3), so that the saddlepoint equations (which are equivalent to the ME equations) may be solved.

6.2. *Determinants and joint distributions*

6.2.1. *The maximum determinant rule*

Tsoucaris (1970) proposed a method for constructing the conditional distributions of large families of structure factors when prior knowledge of the elements of a KH determinant is available. His procedure consists of applying the central limit theorem (CLT) after obtaining the covariance matrix of the family of structure factors by reinterpreting Sayre's equations (Sayre, 1952) as the calculation of an ensemble average:

$$\langle E_{\mathbf{l}+\mathbf{h}_j}\overline{E_{\mathbf{l}+\mathbf{h}_k}}\rangle_{\mathbf{l}} = U_{\mathbf{h}_j-\mathbf{h}_k}. \tag{6.6}$$

The covariance matrix is thus a KH matrix. If the elements of such a matrix of order m are known, then the values of $E_{\mathbf{l}+\mathbf{h}_1}, \ldots, E_{\mathbf{l}+\mathbf{h}_m}$ which maximize the CLT approximation to their joint distribution are those values which maximize the determinant of the $(m+1)\times(m+1)$ KH matrix obtained by bordering the previous matrix with $(U_{\mathbf{l}+\mathbf{h}_1}, \ldots, U_{\mathbf{l}+\mathbf{h}_m})$. Tsoucaris summarized these findings in his maximum determinant rule: 'among all the combinations of phases compatible with the condition of non-negativity of a Karle–Hauptman determinant, the most probable combination is the one that leads to the maximum value of this determinant'.

6.2.2. *Scope and limitations of the method*

The maximum-determinant method has had a number of successful applications to the solution of organic structures and to phase extension for proteins at high resolution (de Rango, Mauguen & Tsoucaris, 1975). Conceptually, it constitutes a major step towards the systematic exploitation of prior knowledge concerning some structure factors in order to sharpen probablistic relations between others. It does not achieve this goal in an optimal fashion either from a theoretical or a practical standpoint, as will now be discussed.

Tsoucaris's construction is ingenious, yet its justification by means of Sayre's equations has the character of a *deus ex machina*. Sayre's equations are not of a probabilistic nature: they are exact relations satisfied by the Fourier coefficients of any structure consisting of equal resolved spherical atoms. This logical inconsistency may be remedied by observing that, in the construction on which the MDM is based, implicit use is made throughout of the idea of a non-uniform prior distribution $q(\mathbf{x})$ which would yield the assumed values for the Fourier coefficients contained in the known determinant; one can then use the standard formula (1.11) for calculating moments, and obtain the desired result (6.6) without any need for Sayre's equations. This prior distribution $q(\mathbf{x})$, however, is never explicitly mentioned in the MDM, let alone specified with any degree of uniqueness as was done at length in § 3.3. In particular, the method of ME extrapolation by Fourier analysis of the ME prior, which affords a device for giving the most non-committal values to unspecified Fourier coefficients, has no equivalent here.

From a practical point of view, the rigid format in which the available knowledge of some structure factors must be presented – namely as a collection of elements filling exactly a KH determinant – is somewhat of a strait jacket, and makes the MDM extremely vulnerable to missing or unreliable data. Finally, the

MDM requires computations of size proportional to n^3 for n reflexions, which becomes prohibitive for large n.

By contrast, the construction of c.p.d.'s given by (4.10), which is equivalent to the form (4.5) on which the MDM is based, possesses a number of considerable advantages:

(1) all the available initial data are used both in the construction of $\mathbf{U}^{\mathrm{ME}}$ and of the covariance matrix $\mathbf{Q}$;

(2) no limitations are imposed by the fact that the data may not fill up the whole of matrix $\mathbf{K}$;

(3) the inversion of $\mathbf{K}$ is carried out by fast Fourier transform (FFT), hence at a cost proportional to $n \log n$ for n reflexions instead of n^3;

(4) the linear regression calculations used in the MDM (de Rango, Tsoucaris & Zelwer, 1974) are replaced by maximum-entropy extrapolation.

Another source of difficulties, which has so far gone undiagnosed but would preclude the unsupervised use of the MDM *ab initio*, is the branching problem. Like the entropy functional, a positive KH determinant is a strictly concave function of its arguments, but the non-convexity of the moduli constraints (§ 4.3) will cause it to have a profusion of local maxima when considered as a function of the phases alone. Thus *the maximum-determinant rule will not in general define a unique combination of phases.* No provision exists in the current implementation of the MDM for dealing with this problem.

6.3. *Determinants and entropy*

The relation between the MDM and the MEM is an intricate one, which is further complicated by a confusion of terminology owing to the possibility of defining three entropy quantities in the present context. If $q(\mathbf{x})$ is a prior distribution of atoms, $p(\mathbf{x}) = (1/F_0)\rho(\mathbf{x})$ a normalized electron density, and $\mathcal{P}(\mathbf{E})$ the joint distribution of a collection of (quasi-normalized) structure factors, then one will encounter

$$(1) \qquad \mathcal{S}_q(p) = -\int_V p(\mathbf{x}) \log [p(\mathbf{x})/q(\mathbf{x})]\, \mathrm{d}^3\mathbf{x},$$

the Shannon–Jaynes entropy of p relative to q, used in this paper;

$$(2) \qquad \mathcal{H}(\mathcal{P}) = -\int_{\mathbb{R}^{2m}} \mathcal{P}(\mathbf{E}) \log \mathcal{P}(\mathbf{E})\, \mathrm{d}^{2m}\mathbf{E},$$

the Shannon entropy of the j.p.d. of a set of m structure factors;

$$(3) \qquad \mathcal{T}(p) = \frac{1}{V}\int_V \log p(\mathbf{x})\, \mathrm{d}^3\mathbf{x},$$

the Burg entropy of p, which will be introduced in § 6.3.3.

Britten & Collins (1982) and Piro (1983) have shown the equivalence of the MDM with the MEM for entropy $\mathcal{H}(\mathcal{P})$, while Narayan & Nityananda (1982) have shown its equivalence to the MEM for entropy $\mathcal{T}(p)$. In this section, all these methods will be compared after a few mathematical results have been established.

6.3.1. *A theorem of Shannon*

It was shown by Shannon (Shannon & Weaver, 1949, pp. 89–90) that, given a positive definite $M \times M$ symmetric matrix $\mathbf{Q}$, the ME multivariate distribution having $\mathbf{Q}$ as its covariance matrix is the Gaussian distribution

$$\mathcal{G}_{\mathbf{Q}}(\mathbf{X}) = [(2\pi)^M \det \mathbf{Q}]^{-1/2} \exp(-\tfrac{1}{2}\mathbf{X}'\mathbf{Q}^{-1}\mathbf{X})$$

and that the corresponding maximum value of the entropy is

$$\mathcal{H}(\mathbf{Q}) = -\int_{\mathbb{R}^M} \mathcal{G}_{\mathbf{Q}}(\mathbf{X}) \log \mathcal{G}_{\mathbf{Q}}(\mathbf{X})\, \mathrm{d}^M\mathbf{X}$$
$$\triangleq (M/2)(1 + \log 2\pi) + \tfrac{1}{2} \log \det \mathbf{Q}. \qquad (6.7)$$

Essentially the same result holds for complex-valued random variables and a positive-definite Hermitian matrix $\mathbf{Q}$.

6.3.2. *A theorem of Szegö*

The relation discovered by Karle & Hauptman between the positivity of a periodic function and the existence of determinantal inequalities satisfied by its Fourier coefficients had been known to mathematicians since the work of Toeplitz (1911) and Carathéodory (1911). Hermitian forms such as T_p in (6.1) are referred to as *Toeplitz forms*, and an abundant mathematical literature is devoted to them (*e.g.* Grenander & Szegö, 1958). One of the central results of the theory of Toeplitz forms is a theorem of Szegö (1920), according to which, in the notation of (3.9) – with p instead of q – and (6.5),

$$\lim_{m\to\infty} (d_m)^{1/m} = \exp[(1/V)\int_V \log p(\mathbf{x})\, \mathrm{d}^3\mathbf{x}], \qquad (6.8)$$

provided the increasing families $\{U_{\mathbf{h}_j}\}_{1\le j\le m}$ eventually exhaust all the Fourier coefficients in a suitably isotropic fashion as $m \to \infty$. Szegö's original proof dealt with Fourier series in one variable only. Narayan & Nityananda (1982) provided a proof adequate for crystallographic purposes, although many proofs of a stronger result were already available in the multidimensional case (Widom, 1960, 1975; Linnik, 1975).

6.3.3. *The Burg entropy*

The quantity on the right-hand side of (6.8) is the geometric mean of the values of p, and it is a maximum when p is uniform $[p(\mathbf{x}) = 1/V]$. Furthermore, the concavity of the logarithm function implies that the

functional

$$\mathscr{T}(p) = (1/V) \int_V \log p(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \tag{6.9}$$

is strictly concave. Therefore, $\mathscr{T}(p)$ has many of the attributes of the Shannon entropy, although it lacks its combinatorial interpretation and the property of additivity for independent sources of uncertainty.

Burg (1967) proposed that the ME principle should be implemented using the $\mathscr{T}(p)$ entropy, and his work has attracted many followers (see for instance Haykin, 1979). The content of Szegö's theorem may then be phrased: if $\mathscr{P}$ is an m-dimensional Gaussian distribution whose covariance matrix is an $m \times m$ Toeplitz matrix defined from p, then maximizing the Shannon entropy of $\mathscr{P}$ is equivalent, as $m \to \infty$, to maximizing the Burg entropy of p.

6.3.4. *Entropies in terms of structure invariants*

Let $u(\mathbf{x})$ be defined by

$$p(\mathbf{x}) = (1/V)[1 + u(\mathbf{x})] \tag{6.10}$$

so that

$$u(\mathbf{x}) = \sum_{\mathbf{h} \neq \mathbf{0}} U_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}) \tag{6.11}$$

and hence

$$\int_V u(\mathbf{x})\, \mathrm{d}^3\mathbf{x} = 0. \tag{6.12}$$

If $u(\mathbf{x})$ is always strictly less than 1 in absolute value, we may use the power-series expansion

$$\log(1+u) = \sum_{n=1}^{\infty} [(-1)^{n-1}/n]\, u^n \tag{6.13}$$

and use its analytical continuation otherwise.

The Jaynes entropy $\mathscr{S}(p)$ relative to a uniform prior $q(\mathbf{x}) = 1/V$ and the Burg entropy $\mathscr{S}(p)$ are then readily obtained:

$$\begin{cases} \mathscr{S}(p) = -\dfrac{1}{2}\dfrac{1}{V}\displaystyle\int_V u^2(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \\ \qquad\qquad + \displaystyle\sum_{n=3}^{\infty} \frac{(-1)^{n-1}}{n(n-1)} \frac{1}{V} \int_V u^n(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \\ \mathscr{T}(p) = -\log V - \dfrac{1}{2}\dfrac{1}{V}\displaystyle\int_V u^2(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \\ \qquad\qquad + \displaystyle\sum_{n=3}^{\infty} \frac{(-1)^{n-1}}{n} \frac{1}{V} \int_V u^n(\mathbf{x})\, \mathrm{d}^3\mathbf{x}. \end{cases} \tag{6.14}$$

The two expressions are formally very similar, except for a different balance between successive terms, and show clearly that the change of sign in going from $-p \log p$ (for $\mathscr{S}$) to $\log p$ (for $\mathscr{T}$) is a consequence of the alternation of signs in the expansion (6.13) for $\log(1+u)$.

In terms of Fourier coefficients, the common quadratic integral is

$$-\frac{1}{2}\frac{1}{V}\int_V u^2(\mathbf{x})\, \mathrm{d}^3\mathbf{x} = -\frac{1}{2}\sum_{\mathbf{h}\neq\mathbf{0}} |U_{\mathbf{h}}|^2 \tag{6.15}$$

by Parseval's theorem, and it gives rise to Wilson's statistics. The expression of the higher-order terms,

$$\frac{1}{V}\int_V u^n(\mathbf{x})\, \mathrm{d}^3\mathbf{x} = \sum_{\mathbf{h}_1+\ldots+\mathbf{h}_n=\mathbf{0}} |U_{\mathbf{h}_1}| \ldots |U_{\mathbf{h}_n}| \times \cos(\varphi_{\mathbf{h}_1} + \ldots + \varphi_{\mathbf{h}_n}), \tag{6.16}$$

together with the alternating signs in (6.14), show that both entropy maximization procedures will endeavour to make triplet invariants positive, quartet invariants negative, *etc.*, although these will be given different relative weights. In particular, the maximization of $\mathscr{S}(p)$ gives greater relative weight to triplet invariants than does that of $\mathscr{T}(p)$.

6.3.5. *Comparison of the MDM and the ME methods*

We may recapitulate the results obtained above by their formulation in a crystallographic context:

(1) the MDM differs from the Jaynes MEM developed in §§ 3 and 4;

(2) the MDM is equivalent, by Shannon's theorem (§ 6.3.1), to maximizing the Shannon entropy of the CLT approximation to the joint distribution $\mathscr{P}(\mathbf{E})$ of the structure factors (Britten & Collins, 1982; Piro, 1983);

(3) the MDM is equivalent, by Szegö's theorem (§ 6.3.2), to maximizing the Burg entropy of $p(\mathbf{x})$ (Narayan & Nityananda, 1982).

Thus the Burg MEM constitutes the real-space counterpart of the MDM, in exactly the same fashion that the Jaynes MEM is the counterpart of the method of joint distributions (§ 4.1).

6.4. *Jaynes entropy versus Burg entropy in crystallography*

At this point, the question inevitably arises: which is the 'better' form of the maximum entropy principle? This issue is still the object of much controversy in other disciplines (see Levine & Tribus, 1979), but in crystallography Jaynes's viewpoint possesses a number of decisive advantages.

Firstly, Jaynes's entropy $\mathscr{S}(p)$ is the exact real-space counterpart of the *full* joint probability distribution $\mathscr{P}(\mathbf{E})$ of the structure factors, whereas the Burg

entropy $\mathcal{T}(p)$ is the counterpart of the *CLT approximation* to $\mathcal{P}(\mathbf{E})$. Therefore, the Burg MEM is only a second-order approximation to the Jaynes MEM. Furthermore, this CLT Gaussian approximation to $\mathcal{P}(\mathbf{E})$ is centred around $\mathbf{E}=\mathbf{0}$, so that the warnings issued in § 2.1 concerning the use of such approximations far from their centre apply to the MDM and hence to the Burg MEM. In contrast, the Jaynes MEM is equivalent to the saddlepoint approximation to $\mathcal{P}(\mathbf{E})$, which always uses the CLT in an optimal fashion around any given value of $\mathbf{E}$.

Secondly, the Jaynes MEM is analytically and computationally the more robust of the two. The integrand of $\mathcal{T}(p)$ approaches $-\infty$ wherever $p(\mathbf{x})$ approaches 0, while that of $\mathcal{S}(p)$ approaches 0 under the same circumstances. The solution of Burg's ME equations is obtained as the squared modulus of the inverse of a finite Fourier sum, while that of Jaynes's is the exponential of such a sum (§ 3.3.1) and hence affords a better safeguard against exaggerated 'peakiness'.

Thirdly, the Burg entropy of $p(\mathbf{x})$ does not possess a clearly defined probabilistic interpretation in real space. It does possess one in reciprocal space, by virtue of its relation to the Shannon entropy of $\mathcal{P}(\mathbf{E})$: in this light, the MDM appears as a prescription to maximize the uncertainty in the distribution of the structure factors $\mathbf{E}$, hence to minimize the commitment to any particular value of $\mathbf{E}$. However, there is no underlying stochastic model in reciprocal space which would lead naturally to maximizing such an entropy, so that the rationale for proceeding in this way is rather more *ad hoc* than that by which Jaynes's ME principle introduced itself in §§ 3.2 and 4.1.

Ultimately, the superiority of Jaynes's formulation lies in the fact that it affords the most natural framework in which to pursue the specific purpose of crystallographic direct methods, namely to analyse quantitatively the statistical implications of a real-space combinatorial approximation to the rules of chemistry. When direct methods reach a state of maturity where the crudeness of the present stochastic model of independent Bernoulli trials becomes the major source of limitations, this model will have to be upgraded to one taking into account the non-independence of the atoms (§ 1.1). Burg's MEM, not having explicit combinatorial foundations of this kind, would not admit as natural a generalization as Jaynes's.

7. Entropy maximization algorithms

A fundamental process in the construction of the prior q^{ME} (§ 3) and in the local survey of a c.p.d. $\mathcal{P}(\delta\mathbf{U})$ (§ 4) is the maximization of an entropy functional $\mathcal{S}$ when a constraint functional $\mathcal{C}$ is required to assume a given value. In the construction of q^{ME}, $\mathcal{C}$ will be a residual reckoned from the constraint values specified for some structure factors, while in the exploration of $\mathcal{P}$ it will be a weighted observational residual based on the measured moduli of yet unphased structure factors. Several procedures are available in the former case (§ 7.1). In the latter case, these procedures are inadequate because of the branching problem, and a new method is required. An algorithm developed by the present author, which gains complete control over branching, is presented in § 7.2. A preliminary numerical application to a small protein is described in § 7.3.

7.1. *Algorithms for the construction of a ME prior*

Computational methods suitable for this purpose fall into two categories, according to whether they rely on the analytical ME formalism (§§ 7.1.1, 7.1.2) or on purely numerical optimization procedures (§ 7.1.3).

7.1.1. *Solving the ME equations by Newton iteration*

In the case of linear constraints given by the prior knowledge of some structure factors, we may invoke Jaynes's ME formalism. It was shown previously that the ME equations have a *unique* solution (§§ 5.2 and 5.4) if the constraint values obey the Karle–Hauptman inequalities (§ 6.1.3). The crystallographic ME equations obtained in §§ 3.4 and 3.5 were derived mainly for formal comparison with standard direct methods; for computational purposes, a different writing is more appropriate.

Let $M=2m_a+m_c$ be the total number of independent real components of the m_a acentric and m_c centric structure-factor values given as data, and define the constraint functions $C_j(\mathbf{x})$ and constraint values c_j by

$$\frac{1}{|\Gamma_k|}\sum_{\gamma\in\Gamma_k}\exp\left[2\pi i\mathbf{h}_k\cdot(\mathbf{R}_\gamma\mathbf{x}+\mathbf{t}_\gamma)\right]=C_{2k-1}(\mathbf{x})+iC_{2k}(\mathbf{x})$$

and

$$U(\mathbf{h}_k)=c_{2k-1}+ic_{2k}\quad(k=1,\ldots,m_a)\qquad(7.1a)$$

for acentric reflexions,

$$\frac{1}{|\Gamma_k|}\sum_{\gamma\in\Gamma_k}\exp\left[2\pi i\mathbf{h}_{m_a+k}\cdot(\mathbf{R}_\gamma\mathbf{x}+\mathbf{t}_\gamma)\right]=\exp(i\alpha_k)C_{2m_a+k}(\mathbf{x})$$

and

$$U(\mathbf{h}_{m_a+k})=\exp(i\alpha_k)c_{2m_a+k}\quad(k=1,\ldots,m_c)\qquad(7.1b)$$

for centric reflexions, where the phase $\varphi(\mathbf{h}_{m_a+k})$ is restricted to $\alpha_k \bmod \pi$. With this notation the constraint equations (3.34) and the ME equations assume their canonical forms (ME0), (ME1), (ME2), (ME3).

Equations (ME3) may be written

$$\nabla_\lambda(\log Z)=\mathbf{c}\qquad(7.2)$$

so that they can be solved by Newton's method (Agmon, Alhassid & Levine, 1979). Let $\boldsymbol{\lambda}^{(i)}$ be the vector of trial Lagrange multipliers at iteration i, and define

$$Z^{(i)} = \int_V m(\mathbf{x}) \exp\left[\sum_{j=1}^{M} \lambda_j^{(i)} C_j(\mathbf{x})\right] \mathrm{d}^3\mathbf{x} \qquad (7.3)$$

$$q^{(i)}(\mathbf{x}) = m(\mathbf{x})/Z^{(i)} \exp\left[\sum_{j=1}^{M} \lambda_j^{(i)} C_j(\mathbf{x})\right] \qquad (7.4)$$

$$c_j^{(i)} = \int_V q^{(i)}(\mathbf{x}) C_j(\mathbf{x})\, \mathrm{d}^3\mathbf{x}. \qquad (7.5)$$

Then the new iterate $\boldsymbol{\lambda}^{(i+1)}$ will be defined by

$$\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)} + [\mathbf{H}(\log Z^{(i)})]^{-1}(\mathbf{c} - \mathbf{c}^{(i)}), \qquad (7.6)$$

where $\mathbf{H}(\log Z^{(i)})$ is the Hessian matrix of $Z^{(i)}$. It was shown in § 3.3.3 that

$$[\mathbf{H}(\log Z^{(i)})]_{jk} = \int_V q^{(i)}(\mathbf{x}) C_j(\mathbf{x}) C_k(\mathbf{x})\, \mathrm{d}^3\mathbf{x} - c_j^{(i)} c_k^{(i)}. \qquad (7.7)$$

The product $C_j(\mathbf{x})C_k(\mathbf{x})$ can be linearized by Bertaut's structure-factor algebra (Bertaut, 1955*c*, 1956, 1959*a*, *b*; Bertaut & Waser, 1957) so that the integral on the right-hand side of (7.7) may be expressed as a finite linear combination of real and imaginary parts of Fourier coefficients of $q^{(i)}$. The Newton iteration scheme for solving the crystallographic ME equations is thus completely specified.

7.1.2. *Exponential modelling*

Newton's method becomes expensive as M increases, since the cost of inverting $\mathbf{H}(\log Z^{(i)})$ grows as M^3. For very large M, the inversion technique used in § 4.2.1 provides considerable savings by making the size of computations proportional to $M \log M$. Expanding the data to $P1$, $\mathbf{H}(\log Z^{(i)})$ is simply the matrix $\mathbf{Q}^{(i)} = \mathbf{K}^{(i)} - \mathbf{W}^{(i)}$ of (4.6) constructed from the Fourier coefficients of $q^{(i)}$, restricted to the subspace spanned by the expanded set of reflexions. If M is large enough we may ignore this restriction, neglect the small perturbation $\mathbf{W}^{(i)}$ to $\mathbf{K}^{(i)}$, and use (4.8) to obtain an approximate inverse:

$$[\mathbf{H}(\log Z^{(i)}]^{-1} \simeq KH(1/q^{(i)}). \qquad (7.8)$$

Step (7.6) can then be carried out much more economically by defining

$$\omega^{(i)}(\mathbf{x}) = -\log Z^{(i)} + \sum_{j=1}^{M} \lambda_j^{(i)} C_j(\mathbf{x}) \qquad (7.9a)$$

$$q^{(i)}(\mathbf{x}) = (1/V) \exp[\omega^{(i)}(\mathbf{x})] \qquad (7.9b)$$

$$c_j^{(i)} = \int_V q^{(i)}(\mathbf{x}) C_j(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \qquad (7.9c)$$

$$\delta^{(i)}(\mathbf{x}) = (1/V)\left[1 + 2\sum_{j=1}^{M} |\Gamma_j|(c_j - c_j^{(i)}) C_j(\mathbf{x})\right] \qquad (7.9d)$$

$$\delta\lambda_j^{(i)} = \int_V [\delta^{(i)}(\mathbf{x})/q^{(i)}(\mathbf{x})] C_j(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \qquad (7.9e)$$

and replacing (7.6) by

$$\lambda_j^{(i+1)} = \lambda_j + \delta\lambda_j^{(i)}. \qquad (7.10)$$

This real-space scheme is essentially the 'exponential modelling' technique used by Collins & Mahar (1983) to impose positivity. The above analysis shows that it also yields a solution with maximum entropy under the given structure-factor constraints.

In the author's experience, this algorithm tends to be very unstable, because of the division operation in (7.9*e*), but stability can nevertheless be secured by suitably reshaping and attenuating the shifts. The large-scale solution of the ME equations can therefore be accomplished with computational complexity $M \log M$, since the steps are either pointwise operations (in real or reciprocal space) or fast Fourier transformations. A complete solution typically requires about 30 cycles involving four Fourier transformations each.

7.1.3. *The Skilling algorithm*

Entropy maximization by solving the ME equations always entails fitting the constraint values exactly, making the results vulnerable to inaccuracies which may be present in the data. Most frequently the constraint values c_j $(j = 1, \ldots, M)$ under which the entropy $\mathcal{S}$ is to be maximized are not known exactly but rather have a variance σ_j^2 attached to them. Gull & Daniell (1978) proposed that in such cases the maximum of $\mathcal{S}$ should be sought under the unique constraint that the 'reduced chi-squared' statistic (Cramér, 1946)

$$\mathscr{C} = (1/M) \sum_{j=1}^{M} [(c_j^{\mathrm{calc}} - c_j^{\mathrm{obs}})/\sigma_j]^2 \qquad (7.11)$$

be equal to 1 (not 0), and they developed a numerical method to carry out this maximization. This method was used by Collins (1982) on data from rubredoxin. Unfortunately, it is beset with severe instability problems which greatly impair its effectiveness.

Recently, Skilling and his collaborators have designed a more sophisticated algorithm which has met with considerable success in a broad variety of image-reconstruction problems (Bryan & Skilling, 1980; Burch, Gull & Skilling, 1983). This algorithm will now be described to introduce notation and concepts to be used in § 7.2.

For computational purposes, a prior distribution $q(\mathbf{x})$ will be handled as a column vector of sample values [still denoted by $q(\mathbf{x})$], or as the column vector U of its Fourier coefficients. These column vectors are the coordinates of the same vector in an abstract vector space v of dimension $\mathcal{N}$, but referred to two different bases which are related by the Fourier transform and its inverse. Scalar products between vectors

in v are defined by

$$\langle q_1, q_2\rangle = V \int_V q_1(\mathbf{x}) q_2(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \qquad (7.12a)$$

in the real-space basis and

$$\langle U_1, U_2\rangle = \sum_{\mathbf{h}} \overline{U_1(\mathbf{h})} U_2(\mathbf{h}) \qquad (7.12b)$$

in the reciprocal-space basis, and by Parseval's theorem (*i.e.* the unitarity of the Fourier transform)

$$\langle q_1, q_2\rangle = \langle U_1, U_2\rangle \qquad (7.13)$$

if q_i and U_i $(i = 1, 2)$ are related by the Fourier transform.

At a maximum of $\mathscr{S}$ under constraint (7.11) on $\mathscr{C}$, the gradients $\nabla\mathscr{S}$ and $\nabla\mathscr{C}$ are collinear. In order to move towards such a solution from a trial position, we need to represent the dependence of $\nabla\mathscr{S}$ and $\nabla\mathscr{C}$ on position in v, which involves the $\mathscr{N}\times\mathscr{N}$ Hessian matrices $\mathbf{H}(\mathscr{S})$ and $\mathbf{H}(\mathscr{C})$. The size of these matrices precludes their use in the whole of v, but they may be used in a subspace of v with smaller dimension n.

Skilling's algorithm (Burch, Gull & Skilling, 1983) defines such a search subspace by n search vectors $\{v_i(\mathbf{x}), i = 1, \ldots, n\}$, where n is 3, 4 or 6, constructed from $\nabla\mathscr{S}$, $\nabla\mathscr{C}$, and their images under the action of $\mathbf{H}(\mathscr{S})$, $\mathbf{H}(\mathscr{C})$, or both. A vector $\mathbf{X}_0$ in this subspace is parametrized by a column vector X_0 of n coordinates. In a neighbourhood of the current trial solution $q(\mathbf{x})$, the variations of $\mathscr{S}$ and $\mathscr{C}$ in this subspace may be approximated by the quadratic models

$$\begin{cases} \mathscr{S}(\mathbf{X}_0) = \mathscr{S}(\mathbf{0}) + {}^tS_0X_0 + (\frac{1}{2}){}^tX_0H_0(\mathscr{S})X_0 & (7.14a) \\ \mathscr{C}(\mathbf{X}_0) = \mathscr{C}(\mathbf{0}) + {}^tC_0X_0 + (\frac{1}{2}){}^tX_0H_0(\mathscr{C})X_0 & (7.14b) \end{cases}$$

whose coefficients are given by

$$[S_0]_i = \langle v_i, \nabla\mathscr{S}\rangle \qquad (7.15a)$$

$$[C_0]_i = \langle v_i, \nabla\mathscr{C}\rangle \qquad (7.15b)$$

$$[H_0(\mathscr{S})]_{ij} = \langle v_i, \mathbf{H}(\mathscr{S})v_j\rangle \qquad (7.15c)$$

$$[H_0(\mathscr{C})]_{ij} = \langle v_i, \mathbf{H}(\mathscr{C})v_j\rangle. \qquad (7.15d)$$

These scalar products may be evaluated either in real or in reciprocal space, by virtue of (7.13). Both $\nabla\mathscr{S}$ and $\mathbf{H}(\mathscr{S})$ are simplest in real space (§§ 3.3.1 and 4.2.1) so that (7.15*a*, *c*) are best calculated by (7.12) as

$$[S_0]_i = \int_V v_i(\mathbf{x})\{-1 - \log[q(\mathbf{x})/m(\mathbf{x})]\}\, \mathrm{d}^3\mathbf{x}$$

$$[H_0(\mathscr{S})]_{ij} = \int_V -[v_i(\mathbf{x})v_j(\mathbf{x})/q(\mathbf{x})]\, \mathrm{d}^3\mathbf{x},$$

where $q(\mathbf{x})$ is the current estimate of the solution. On the other hand, $\nabla\mathscr{C}$ and $\mathbf{H}(\mathscr{C})$ are simplest in reciprocal space and (7.15*b*, *d*) are best evaluated by (7.12*b*).

Matrix $H_0(\mathscr{S})$ is symmetric, hence can be diagonalized in an orthonormal basis of eigenvectors; as $\mathbf{H}(\mathscr{S})$ is negative definite in the whole of v, these eigenvectors can be rescaled to produce an orthogonal basis of the search subspace $\{\mathbf{e}_i, i = 1, \ldots, n\}$ in which $\mathbf{H}(\mathscr{S})$ is represented by minus the identity matrix. In terms of coordinates X_1 in this new basis, the quadratic models (7.14) become

$$\begin{cases} \mathscr{S}(\mathbf{X}_1) = \mathscr{S}(\mathbf{0}) + {}^tS_1X_1 - (\frac{1}{2}){}^tX_1X_1 & (7.16a) \\ \mathscr{C}(\mathbf{X}_1) = \mathscr{C}(\mathbf{0}) + {}^tC_1X_1 + (\frac{1}{2}){}^tX_1H_1(\mathscr{C})X_1. & (7.16b) \end{cases}$$

These models are local, and are accurate only for small enough values of their arguments. Since non-quadratic behaviour comes mainly from the entropy $\mathscr{S}$, Skilling suggested the use of the 'entropy metric'

$$\mathscr{D}(\mathbf{X}_1) = {}^tX_1X_1 \qquad (7.17)$$

to measure the size of the readjustments $\mathbf{X}_1$ which has to be kept small at each iteration.

Further simplification may be obtained by diagonalizing $H_1(\mathscr{C})$ as $R\Lambda R^{-1}$, where $R = {}^tR^{-1}$ is an orthogonal matrix and Λ is diagonal. Putting $Y = R^{-1}X_1$, $A = R^{-1}S_1$ and $B = R^{-1}C_1$, we may cast the quadratic models of $\mathscr{S}$, $\mathscr{C}$ and $\mathscr{D}$ in the final form:

$$\begin{cases} \mathscr{S}(\mathbf{Y}) = \mathscr{S}(\mathbf{0}) + (\frac{1}{2}){}^tAA \\ \qquad\qquad - (\frac{1}{2}){}^t(Y - A)(Y - A) & (7.18a) \\ \mathscr{C}(\mathbf{Y}) = \mathscr{C}(\mathbf{0}) - (\frac{1}{2}){}^tB\Lambda^{-1}B \\ \qquad\qquad + (\frac{1}{2}){}^t(Y + \Lambda^{-1}B)\Lambda(Y + \Lambda^{-1}B) & (7.18b) \\ \mathscr{D}(\mathbf{Y}) = {}^tYY. & (7.18c) \end{cases}$$

Maximizing $\mathscr{S}$ under constraints on the values of $\mathscr{C}$ and $\mathscr{D}$ leads to a condition of the form

$$\nabla_Y(\mathscr{C} - \mu\mathscr{S} + \nu/2\mathscr{D}) = \mathbf{0},$$

i.e.

$$\Lambda Y + B + \mu(Y - A) + \nu Y = 0,$$

which yields a parametric representation of the solution:

$$y_i(\mu, \nu) = (\mu a_i - b_i)/(\lambda_i + \mu + \nu) \quad (i = 1, \ldots, n) \qquad (7.19)$$

in terms of the Lagrange multipliers μ and ν (which must be positive).

To solve for μ and ν the Skilling algorithm proceeds as follows. Let $\mathscr{C}^*_{\text{aim}}$ be the target value of $\mathscr{C}$ given by (7.11), and let $d_{\max}$ be the maximum value of $\mathscr{D}$ allowed (typically 0·2). The absolute minimum value attainable by $\mathscr{C}$ in the subspace is

$$\mathscr{C}_{\min} = \mathscr{C}(-\Lambda^{-1}B) = \mathscr{C}(\mathbf{0}) - (\tfrac{1}{2}){}^tB\Lambda^{-1}B; \quad (7.20)$$

but to allow some flexibility for the maximization of $\mathscr{S}$, the target value of the current iteration is redefined as

$$\mathscr{C}_{\text{aim}} = \max[(\tfrac{2}{3})\mathscr{C}_{\min} + (\tfrac{1}{3})\mathscr{C}(\mathbf{0}), \mathscr{C}^*_{\text{aim}}].$$

The largest μ achieving this value is then determined by (7.19) and (7.18b) with $\nu = 0$. If $\mathscr{D}[Y(\mu, 0)] > d_{\max}$, ν is increased to apply a 'distance penalty' and μ is redetermined. If $\mathscr{C}_{\text{aim}}$ proves inaccessible within the distance limit imposed, $\mathscr{C}_{\text{aim}}$ is increased towards $\mathscr{C}(\mathbf{0})$ and the process is repeated until an acceptable solution $Y(\mu, \nu)$ is found. The shifts along the original search directions are then obtained by backward substitution through the various basis changes, and are used to increment $q(\mathbf{x})$.

This algorithm is very stable and is of computational complexity $\mathcal{N} \log \mathcal{N}$ for $\mathcal{N}$ reflexions.

7.2. *An algorithm for determining the branching structure of c.p.d.'s*

The Skilling algorithm was intended as a general-purpose tool for a wide range of applications involving maximum-entropy reconstructions from incomplete or noisy data. As such, it is ideally suited to the construction of a ME prior q^{ME} when the available structure-factor values are affected by normally distributed errors giving rise to a *convex* constraint function $\mathscr{C}$.

The X-ray phase problem, however, owes its unique difficulty to the non-convexity of its observational (amplitude) constraints, which gives rise to the branching problem. Skilling's algorithm cannot deal properly with such constraints: its solution method for multipliers μ and ν systematically ignores negative eigenvalues of $\mathbf{H}(\mathscr{C})$, and its search subspace is too small to allow the multimodality of $\mathscr{S}$ on level surfaces of $\mathscr{C}$ to manifest itself to any extent. The exploration of conditional distributions would therefore have to be carried out by blindly assigning random trial phases to the structure factors whose c.p.d. is to be surveyed, then solving each set of enlarged ME equations thus obtained to calculate the final entropy – a most unwiedly procedure.

An enhanced algorithm will now be described which overcomes this limitation, and effectively solves the branching problem, by enlarging the search subspace and by using a different solution method which allows the detection of multimodality under non-convex constraints.

The search subspace should always contain $\nabla\mathscr{S}$ and $\nabla\mathscr{C}$ since their eventual collinearity must be checked; but instead of supplementing them with their products by $\mathbf{H}(\mathscr{S})$ or $\mathbf{H}(\mathscr{C})$, for which the coefficients of the quadratic models are expensive to compute, we may construct extra search directions from the Fourier contributions of a selected set of reflexions. With the notation of (7.1a, b), these may be written

$$v_j(\mathbf{x}) = C_j(\mathbf{x}) \quad (j = 1, \ldots, M). \tag{7.21}$$

The only factor limiting the size of M is that the diagonalization of an $M \times M$ real symmetric matrix must remain an easy operation, so that M may be 200, 300 or more. The coefficients of the quadratic models for $\mathscr{S}$ and $\mathscr{C}$ are easily obtained by (7.15). In particular, all the components of S_0 are Fourier coefficients of $-1 - \log[q(\mathbf{x})/m(\mathbf{x})]$, and

$$[H_0(\mathscr{S})]_{ij} = -\int_V [C_i(\mathbf{x}) C_j(\mathbf{x})/q(\mathbf{x})]\, \mathrm{d}^3\mathbf{x}$$

may be expressed as a linear combination of real and imaginary parts of Fourier coefficients of $1/q(\mathbf{x})$ by Bertaut's structure-factor algebra, as was done in § 7.1.1 for (7.7). A considerable expansion of the search subspace may thus be achieved very economically by taking advantage of the specific form of the constraint functions encountered in crystallography.

Once this extended quadratic model is set up in the form (7.18), the multipliers μ and ν are determined by a method which takes into account the distance restriction $\mathscr{D}(Y) \le d_{\max}$ at the outset rather than at the end of the solution process. It is convenient to introduce the functions

$$\mathscr{C}(\mu, \nu) = \mathscr{C}[Y(\mu, \nu)]$$
$$\mathscr{D}(\mu, \nu) = \mathscr{D}[Y(\mu, \nu)]$$

whose partial derivatives are readily evaluated from (7.18) and (7.19). The solution method goes as follows.

(1) If the absolute minimum $\mathscr{C}_{\min}$ of $\mathscr{C}$ [see (7.20)] is reached within the squared distance $d_{\max}$, set $m = 0$, $\nu_1 = 0$, $\mathscr{C}^{(1)}_{\min} = \mathscr{C}_{\min}$, and go to (3).

(2) Put $\mu = 0$ and find ν such that $\mathscr{D}(0, \nu) = d_{\max}$. By (7.18$c$) and (7.19) this leads to the equation for ν:

$$D(\nu) \equiv \sum_{i=1}^{n} [b_i/(\lambda_i + \nu)]^2 = d_{\max} \tag{7.22}$$

with the condition $\nu > 0$. If the constraint function $\mathscr{C}$ is convex, all the λ_i are positive, and there is a unique positive root ν; if $\mathscr{C}$ is not convex, however, there may be many admissible roots for this equation (see Fig. 3). The latter are easily found by locating the minima of $D(\nu)$ between consecutive double poles [*i.e.* solving $D'(\nu) = 0$ for $\nu > 0$ by Newton's method] and comparing the corresponding minimum values of D to $d_{\max}$: if m of them are less than $d_{\max}$, there are m extra pairs of roots of (7.22) besides the leading root. The branches of the graph of d associated with minima of $\mathscr{C}$ may be identified by noting that along such a branch one must have $\partial\mathscr{C}(0, \nu)/\partial\nu > 0$, since the minimum reached by $\mathscr{C}$ increases if the weight ν of the antagonist constraint $\mathscr{D}$ is increased. It is readily verified that

$$\partial\mathscr{C}(0, \nu)/\partial\nu + (\nu/2)\, \partial\mathscr{D}(0, \nu)/\partial\nu = 0$$

so that ν must satisfy $D'(\nu) < 0$ and hence only the leftmost root of each pair belongs to a branch where $\mathscr{C}$ is a minimum. There are thus $m + 1$ admissible roots (or branches) ν_l ($l = 1, \ldots, m + 1$).

(3) For each admissible ν_l, calculate the minimum value of $\mathscr{C}$ attainable $\mathscr{C}^{(l)}_{\min} = \mathscr{C}(0, \nu_l)$, and set the target value

$$\mathscr{C}^{(l)}_{\text{aim}} = \max\left[(1-\alpha)\mathscr{C}(\mathbf{0}) + \alpha\mathscr{C}^{(l)}_{\min}, \mathscr{C}^{*}_{\text{aim}}\right] \quad (7.23)$$

with *e.g.* $\alpha = 0{\cdot}5$ or $0{\cdot}75$, so as to leave some freedom to maximize $\mathscr{S}$ on that patch of the hypersurface $\mathscr{C} = \mathscr{C}^{(l)}_{\text{aim}}$ contained within the region $\mathscr{D} \le d_{\max}$. Move onto this hypersurface by resetting ν to the value ν'_l defined by $\mathscr{C}(0, \nu'_l) = \mathscr{C}^{(l)}_{\text{aim}}$, which can be obtained by Newton's method.

(4) To maximize $\mathscr{S}$ under the constraints on $\mathscr{C}$ and $\mathscr{D}$, increase μ and decrease ν while keeping $\mathscr{C}[\mu, \nu(\mu)] = \mathscr{C}^{(l)}_{\text{aim}}$, until either $\mathscr{D}[\mu, \nu(\mu)] = d_{\max}$ or $\nu = 0$. Here ν is defined as a function of μ so as to keep $\mathscr{C}$ constant, hence by the implicit function theorem

$$\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)_{\mathscr{C}} = -\left(\frac{\partial\mathscr{C}}{\partial\mu}\Big/\frac{\partial\mathscr{C}}{\partial\nu}\right). \quad (7.24)$$

The final values (μ^*_l, ν^*_l) define parametrically by (7.19) the shift $Y(\mu^*_l, \nu^*_l)$, and hence the correction $\delta\mathbf{U}^{(l)}$ to be applied in order to move along the lth branch from the current solution.

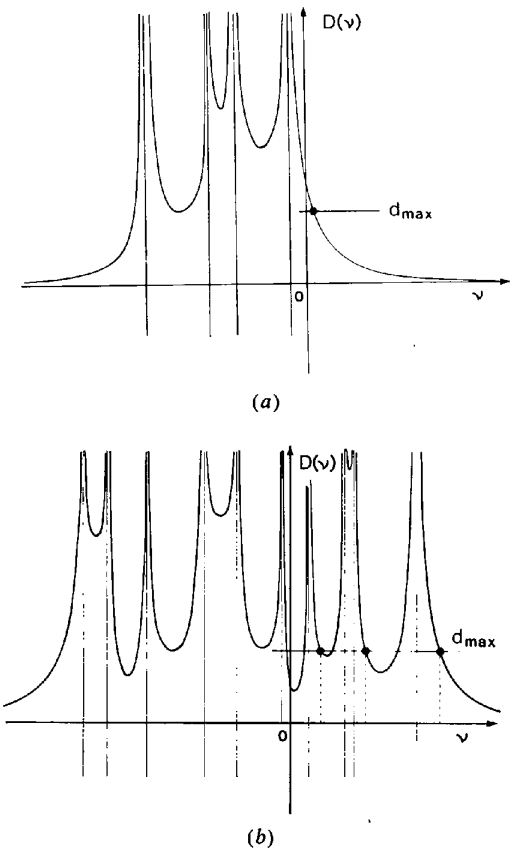


Fig. 3. Graphical solution of equation $D(\nu) = d_{\max}$. (*a*) Convex constraints: no branching. (*b*) Non-convex constraints: possible branching.

In the construction of a prior q^{ME}, the $\mathscr{C}$ function is convex ($m = 0$) and hence no branching occurs; in this mode, the algorithm is then equivalent to that of Skilling but with a more analytical solution method.

In the exploration of a c.p.d. $\mathscr{P}(\delta p)$ defined by (4.22), on the other hand, we have

$$\nabla[\mathscr{S}_{q^{\text{ME}}}(q^{\text{ME}} + \delta p)]|_{\delta p = 0} = \mathbf{0} \quad (7.25)$$

so that

$$\mathscr{D} = 2[\mathscr{S}(\mathbf{0}) - \mathscr{S}], \quad (7.26)$$

Relation (7.26) shows that the 'entropy metric' $\mathscr{D}$, introduced by Skilling on heuristic grounds, has a very natural interpretation as (twice) the decrease of $\log \mathscr{P}$. Therefore the branching structure in (7.22) is also that of $\mathscr{S}$ and hence of $\mathscr{P}$: *the present algorithm yields an explicit parametrization of the branching behaviour of conditional distributions.*

7.3. *Example of the construction of a maximum-entropy prior*

A numerical application of the algorithm just described to data from the small protein Crambin (Hendrickson & Teeter, 1981) will now be described to illustrate the novel process of construction of a ME prior. A full account of the calculations which are being carried out on Crambin will be published separately.

Measured diffraction amplitudes to 1·5 Å resolution, together with phases calculated from a refined model, were kindly provided by Dr W. A. Hendrickson. The calculated phases were associated with the observed moduli to a resolution of 3·0 Å, giving a total of 749 Fourier coefficients, and an error model was constructed for each reflexion by assigning to both radial and tangential components of the error the standard deviation of the experimental modulus. The constraint function $\mathscr{C}$ was taken to be the reduced χ^2 statistic (7.11) of this Gaussian model, for which a target value $\mathscr{C}^*_{\text{aim}} = 1{\cdot}0$ was chosen. Entropy maximization was performed iteratively by the method of § 7.2, starting from a conventional 3 Å map in which values below 0·1 were reset to 0·1, using four search directions: $\nabla\mathscr{S}$, $\nabla\mathscr{C}$, the difference map, and the current map. Convergence was obtained in 50 cycles taking 2 min of CPU time each on a VAX 11/780 computer (the progress of the calculation is summarized in Table 1).

The results are illustrated in Figs. 4(*a*), (*b*), (*c*) by displays of section 0 of the density functions; all functions are on the same scale and were contoured at the same absolute levels. Fig. 4(*a*) is a conventional Fourier synthesis at 3 Å resolution, and contains extensive negative regions due to series-termination effects. Fig. 4(*b*) is the 3 Å ME prior distribution of atoms compatible with the same data: it is everywhere positive, and is much sharper since the series-

termination effects have been suppressed (deconvoluted) by extrapolating the spectrum. This sharpening is meaningful, as may be ascertained by detailed comparison with the conventional map at 1·5 Å resolution (Fig. 4(*c*)). The two most noticeable changes occur at the bottom of the maps, at a proline residue (Pro 5, to the left) and an aromatic ring seen sideways (Phe 3, to the right).

Fourier analysis of the 3 Å ME prior to 1·5 Å resolution gave values of $\mathbf{U}^{ME}$ between 3·0 and 1·5 Å which were compared to the known values of these structure factors. As shown in Table 2, about 1200 reflexions (*i.e.* a quarter of the reflexions in this range) with $|U_{\mathbf{h}}| \geq 0{\cdot}01$ had their phases predicted with a mean absolute error of 45°: the extrapolated values $\mathbf{U}^{ME}$ around which all conditional distributions of high-resolution structure factors would be recentred are thus already a fairly good estimate of the correct values. The Gaussian c.p.d. (4.22) of the deviations from $\mathbf{U}^{ME}$ has not yet been completely surveyed, but it is nevertheless clear that its stringency is much greater than that of the c.p.d. which would result from an appeal to Sim's formula (which is often used in phase extension). Indeed, the final value of the entropy (Table 1) implies that, for $N = 440$ atoms, this c.p.d. contains about 360 bits more information than that derived from Sim's formula (*i.e.* with $q^{ME} = 1/V$); this amounts to specifying an extra 360 phases with a mean absolute error of 45°. It ought to be stressed that this phase prediction has been accomplished *without in any way consulting the amplitude information available beyond* 3 Å *resolution.*

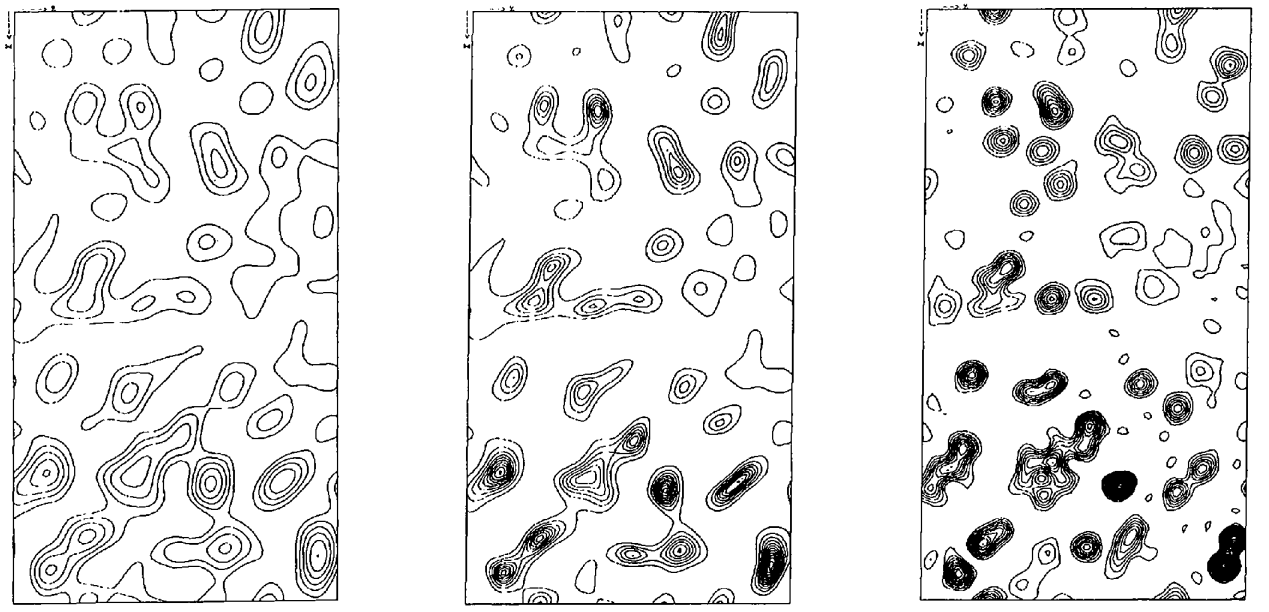
Table 1. *Summary of entropy maximization at* 3·0 Å

Cycle	χ^2	Entropy	$\cos(\nabla\mathcal{S}, \nabla\mathscr{C})$
0	518·47	−0·37841	0·61052
5	123·44	−0·46712	0·83445
10	45·67	−0·51370	0·83316
15	21·52	−0·53762	0·82436
20	11·45	−0·55234	0·81684
25	6·69	−0·56176	0·82152
30	4·03	−0·56915	0·84931
35	2·48	−0·57445	0·88163
40	1·61	−0·57755	0·88834
45	1·13	−0·57714	0·95896
50	1·07	−0·57528	0·96451

Work is in progress to explore systematically the conditional distribution of high-resolution structure factors, and in particular the branching behaviour of the phase-extension process (§ 4.3) as described by (7.22).

8. A recursive *ab initio* phasing strategy for macromolecules

Ab initio phase determination by direct methods may be conceived as a succession of phase extension steps (§ 4.3), with explicit book-keeping of the possible branching behaviour throughout the calculation. The algorithm proposed in § 7.2 provides the tactical device required to carry out each step; the final task is to design a strategy by which these individual steps can be organized globally.



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

Fig. 4. Section 0 of Crambin maps. (*a*) Conventional Fourier synthesis at 3 Å resolution. (*b*) Maximum-entropy prior from 3 Å data. (*c*) Conventional Fourier synthesis at 1·5 Å resolution.

Table 2. *Comparison of extrapolated and true phases*

Resolution shells are of equal thickness in $(d^*)^2$ by steps of 1/36.

Shell	Number of reflexions	Mean absolute error (°)
1	78	2·56
2	135	2·57
3	179	2·58
4	177	2·61
5	187	38·70
6	172	45·94
7	166	48·70
8	178	48·53
9	155	49·45
10	128	45·97
11	80	48·16
12	58	42·07
13	51	37·23
14	47	33·66
15	26	50·61
16	20	38·55

8.1. *Multisolution algorithm*

The analysis carried out in § 2 showed the necessity for a multisolution algorithm (§ 2.4) in any calculation involving the construction and exploitation of j.p.d.'s and c.p.d.'s of structure factors, but it left undefined the mechanism by which successive sets of trial phases were to be generated. The study of the branching problem and its solution in § 7.2 provides such a precise rule: each trial phase set need only be enlarged by moving along the branches leading from the corresponding ME prior to the local maxima of a conditional distribution derived from that prior. This completely specifies, in a recursive fashion, a general procedure for *ab initio* phase determination which accommodates branching explicitly, and is thus free from the risk of becoming uncontrollably trapped around false solutions.

The management of such a recursive computation is a complex task, which is best accomplished by using as a book-keeping device a *multisolution tree* representing the various phase choices made and the parentage relations between them. Well-established programming techniques are available to carry out these constructions (Nilsson, 1971), and an example of their implementation may be found in the well known chess-playing program of Spracklen & Spracklen (1978).

The root node of the tree consists of the origin-fixing phases, and its first ramification occurs with the choice of enantiomorph. The subsequent growth of the tree is governed by four fundamental processes taking place at each node:

(1) updating the prior distribution of atoms $q(\mathbf{x})$ to the ME distribution compatible with the phase choices made up to the current node, using any of the methods described in § 7;

(2) constructing the c.p.d. of unphased structure factors on the basis of that ME prior as a quadratic model in a large search subspace (§ 7.2);

(3) locating the maxima of this (usually multimodal) c.p.d. when moduli constraints are activated, as indicated in § 7.2, equation (7.22);

(4) expanding the current node by creating a branch leading to a new tip node for each of these maxima.

Such a tree would quickly grow to an unwiedly size if all the alternatives were allowed to develop at the same pace. It is thus desirable to supervise the growth of the tree so as to maximize the chance of finding the correct set of phases without having to develop too many of its branches . For this purpose, a mechanism akin to natural selection may be used: at each stage, the tip nodes of the current tree are assigned a score depending

(*a*) on the currently achieved value of the entropy, which (as was shown in § 3) measures the size of the population of still reasonably probable structures attached to that node;

(*b*) on its ability to confer a high likelihood to the observed moduli in the second neighbourhood of the basis reflexions by means of the Rice likelihood functions (4.16) or of their centric equivalents.

This score then determines the priority with which each node will be allowed to 'ramify' in step (4).

Any phase information available from multiple isomorphous replacement or anomalous scattering can be incorporated into the definition of the constraint function $\mathscr{C}$ further to help direct the growth of the tree. Non-crystallographic symmetry may be exploited by using as extra search directions not the functions (7.21) themselves, but combinations of them into which the non-crystallographic symmetry has been built beforehand; the algorithm of § 7.2 will then also point out the ambiguities inherent to the process of imposing such symmetries.

8.2. *Estimates of strength reconsidered*

The algorithm just described will accumulate ever-increasing amounts of prior information along each path down the multisolution tree, and will construct at every node *much more accurate* conditional distributions than would be obtained by specializing structure factors to their assumed values in a joint distribution still based on a uniform prior distribution of atoms, since asymptotic expansions are always used with small arguments (§ 2.3.1). But these new distributions might only indicate accurately that the phase relations are no stronger than those obtained by the classical formalism. This is not the case: the new distributions will also be *much more stringent.*

It was shown earlier (§ 4.2.2) that the second-neighbourhood formulae for estimating phase invariants are equivalent to applying Sim's formula, using as a 'known part' the ME prior q^{ME}, but deriving the c.p.d. of the deviations $\delta\mathbf{U} = \mathbf{U} - \mathbf{U}^{\mathrm{ME}}$ on the basis of a *uniform* prior $m(\mathbf{x}) = 1/V$ [which leads to the Wilson

term in (4.12)]. By contrast, the full c.p.d. constructed here uses q^{ME} both as its known part (for recentring) *and* as its prior distribution of atoms to obtain all terms of (4.12) in closed form.

The resulting increase in strength may thus be accurately monitored by consideration of the entropy $\mathcal{S}_m(q^{ME})$ of the ME prior q^{ME} relative to a uniform prior prejudice $m(\mathbf{x}) = 1/V$. This quantity (see § 3.2) measures the extent to which the current prior knowledge imposes further restrictions on the range of still reasonably probable structures compared to the state of maximum ignorance expressed by m: for N atoms, this range is restricted by a further factor $\exp[N\mathcal{S}_m(q^{ME})]$. A numerical estimate of this factor given in § 7.3 (2^{360} for a ME prior determined from 749 reflexions) indicates that this gain in strength is considerable.

Previous estimates of the potential usefulness of high-order probabilistic relations between phases (Klug, 1958) must therefore be revised, and much of the ensuing pessimism can be mitigated. Individual high-order relations do remain weak, but the above algorithm provides a way of organizing great numbers of them around 'seeds' of prior choices capable of bringing about their mutual reinforcement.

It is noteworthy that this new algorithm will be minimally impaired by large inaccuracies in the amplitude measurements. In the standard approach, such inaccuracies would make individual phase relations unreliable, and no method exists to remedy this situation. By contrast, large standard deviations would allow $\mathcal{S}_m(q^{ME})$ to reach less negative a value under constraint (7.11), so that only the overall strength of phase relations would decrease. The new method is thus 'robust', which is of some relevance if it is to be applicable to very large structures yielding rather noisy data.

8.3. *Hierarchy of struture in macromolecules*

In spite of the gain in strength provided by this new multisolution strategy, it would still seem very difficult to deal with a protein structure of (say) 10 000 atoms without an enormous amount of data and great computational expenditure. But the situation appears worse than it actually is, because the crude stochastic approximation to chemistry used throughout direct methods (§ 1.1) leads to conceiving such a structure as a 'perfect chaos' of 10 000 totally independent atoms. In reality, biological macromolecules incorporate a high degree of structural hierarchy (Schulz & Schirmer, 1979) in order to be able to fold reproducibly. This is reflected in reciprocal space by considerable modulation, at low resolution, of the Gaussian shape of the radial intensity distribution which a 'chaotic' structure would closely follow. As shown in Fig. 5, there are marked intensity peaks corresponding successively to:

(*a*) the packing of individual molecules in the crystal lattice (~30 Å);

(*b*) the tertiary structure of each molecule (~10–12 Å);

(*c*) the secondary structure of each tertiary structure element (~5·5 Å for α-helices, ~4·7 Å for β-sheets).

Similar phenomena may be observed in crystals of nucleic acids. These peaks correspond to fluctuations from uniformity which would only occur with any likelihood for a much smaller number of independent elements than the total number N of atoms. This violation of Wilson's statistics was first investigated by Luzzati (1955) by means of combinatorial arguments similar to those used in this work.

Well-marked troughs are found in resolution ranges which separate scales of structural elements corresponding to successive levels of structure. These troughs define privileged intermediate stages in the multisolution process. At each of these resolutions, the structure may be considered as constructed from pseudo-atoms or group scatterers whose number N_{eff} (the 'effective N') is much smaller than N, and for which a non-uniform prior distribution is available by virtue of the choices made for the strong reflexions of the previous intensity peaks (*i.e.* of the previous levels of structure). These two factors will cause the strength of the phase relations used in determining a protein structure to be much greater than would normally be expected if no advantage were taken of their structural hierarchy.

Conclusion

The critical assessment of the current methodology of direct phase determination presented at the beginning of this article brought to light two main weaknesses of the traditional approach which undermined

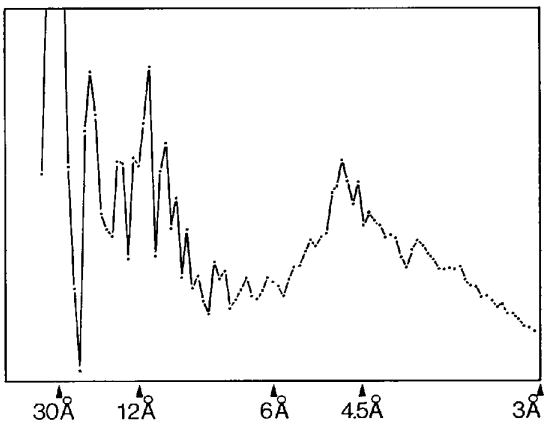


Fig. 5. Modulation of the radial intensity distribution associated with the structural hierarchy of proteins. The data shown are from a pea lectin, and were kindly made available by Professor F. L. Suddath. The peak at 4·7 Å resolution reflects the high proportion of β-sheet present in this structure.

both the accuracy and the strength of the phase relationships which were being derived: the lack of an adequate treatment of information regarding the non-uniformity of the prior distribution of atoms, and the approximation of joint distributions by products of marginal distributions.

The theoretical foundations of direct methods were enlarged by examining the real-space counterpart of the classical formulation in reciprocal space. Jaynes's maximum-entropy principle was then shown to capture precisely the hitherto missing concept required to design a procedure for the proper handling of prior information assumed or acquired in the course of phase determination: interactions among phases are optimally generated and propagated by constantly updating the prior distribution of atoms so as to reflect all the information available at each stage.

The mathematical techniques by which these principles are put to use have been considerably strengthened. The effective construction of conditional probability distributions of very large numbers of structure factors, which are the fundamental tools of direct phase determination, no longer appears as an impossible task. It can be carried out accurately and efficiently within the framework of a multisolution strategy. The conditional distributions produced are stronger and more accurate than those derived from any previously available method, and can be obtained and exploited with computations of size $\mathcal{N} \log \mathcal{N}$ for $\mathcal{N}$ reflexions. Explicit description and control of the branching problem have been achieved, and the multisolution algorithm proposed here can take advantage of the structural hierarchy of macromolecules. The concept of entropy played a ubiquitous role at all stages of the design and of the implementation of this scheme, although an autonomous formulation of it was given within the more traditional language of classical direct methods.

These theoretical developments, together with the preliminary numerical results obtained so far, seem to warrant a much more optimistic attitude than has hitherto prevailed towards the feasibility of phase determination *ab initio* for macromolecules.

I am most indebted to Trinity College, Cambridge, for an extended Research Fellowship during the tenure of which this work was undertaken.

I wish to thank Dr. A. Klug and Dr D. Sayre for stimulating discussions at various stages of this research.

References

AGMON, N., ALHASSID, Y. & LEVINE, R. D. (1979). In *The Maximum Entropy Formalism*, edited by R. D. LEVINE & M TRIBUS, pp. 207–209. Cambridge: MIT Press.

AHLORS, L. V. (1966). *Complex Analysis*. New York: McGraw-Hill.

BERTAUT, E. F. (1955*a*). *Acta Cryst.* **8**, 537–543.

BERTAUT, E. F. (1955*b*). *Acta Cryst.* **8**, 544–548.

BERTAUT, E. F. (1955*c*). *Acta Cryst.* **8**, 823–832.

BERTAUT, E. F (1956). *Acta Cryst.* **9**, 322.

BERTAUT, E. F. (1959*a*). *Acta Cryst.* **12**, 541–549.

BERTAUT, E. F. (1959*b*). *Acta Cryst.* **12**, 570–574.

BERTAUT, E. F. & WASER, J. (1957). *Acta Cryst.* **10**, 606–607.

BRICOGNE, G. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 258–264. New York: Oxford Univ. Press.

BRITTEN, P. L. & COLLINS, D. M. (1982). *Acta Cryst.* A**38**, 129–132.

BRUIJN, N. G. DE (1970). *Asymptotic Methods in Analysis*, 3rd ed. Amsterdam: North-Holland.

BRYAN, R. K. & SKILLING, J. (1980). *Mon. Not. R. Astron. Soc.* **191**, 69–79.

BURCH, S. F., GULL, S. F. & SKILLING, J. (1983). *Computer Vision, Graphics and Image Processing*. In the press.

BURG, J. P. (1967). Annual Meeting of the International Society of Exploration Geophysics, Oklahoma City. Reprinted in *Modern Spectrum Analysis* (1978), edited by D. G. CHILDERS, pp. 34–41. New York: IEEE Press.

CANILLO, E., OBERTI, R. & UNGARETTI, L. (1983). *Acta Cryst.* A**39**, 68–74.

CARATHÉODORY, C. (1911). *Rend. Circ. Mat. Palermo*, **32**, 193–217.

CARMICHAEL, R. D. (1937). *Introduction to the Theory of Groups of Finite Order*. New York: Dover Publications.

COLLINS, D. M. (1982). *Nature (London)*, **298**, 49–51.

COLLINS, D. M. & MAHAR, M. C. (1983). *Acta Cryst.* A**39**, 252–256.

COPSON, E. T. (1965). *Asymptotic Expansions*. Cambridge Univ. Press.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

CURTIS, C. W. & REINER, I. (1962). *Representation Theory of Finite Groups and Associative Algebras*. New York: Wiley–Interscience.

DANIELS, H. E. (1954). *Ann. Math. Stat.* **25**, 631–650.

EDWARDS, A. W. F. (1972). *Likelihood*. Cambridge Univ. Press.

ERDÉLYI, A. (1956). *Asymptotic Expansions*. New York: Dover Publications.

EVES, H. (1966). *Elementary Matrix Theory*. Boston: Allyn and Bacon. Reprinted by Dover Publications, 1980.

FORTIER, S. & HAUPTMAN, H. (1977*a*). *Acta Cryst.* A**33**, 572–575.

FORTIER, S. & HAUPTMAN, H. (1977*b*). *Acta Cryst.* A**33**, 694–696.

FOWLER, R. H. (1936). *Statistical Mechanics*, 2nd ed. Cambridge Univ. Press.

GERMAIN, G. & WOOLFSON, M. M. (1968). *Acta Cryst.* B**24**, 91–96.

GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*. London: Academic Press.

GOHBERG, I. C. & FEL'DMAN, I. A. (1974). *Convolution Equations and Projection Methods for their Solution. Transl. Math. Monogr.* No. 41. Providence: American Mathematical Society.

GRENANDER, U. SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Berkeley: Univ. of California Press.

GULL, S. F. & DANIELL, G. J. (1978). *Nature* (London), **272**, 686–690.

HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70–75.

HAUPTMAN, H. (1975*a*). *Acta Cryst.* A**30**, 472–476.

HAUPTMAN, H. (1975*b*). *Acta Cryst.* A**30**, 822–829.

HAUPTMAN, H. & FORTIER, S. (1977*a*). *Acta Cryst.* A**33**, 575–580.

HAUPTMAN, H. & FORTIER, S. (1977*b*). *Acta Cryst.* A**33**, 697–701.

HAUPTMAN, H. & KARLE, J. (1953). *The Solution of the Phase Problem*: I. *The Centrosymmetric Crystal. Am. Crystallogr. Assoc. Monogr.* No. 3. Pittsburgh: Polycrystal Book Service.

HAUPTMAN, H. & KARLE, J. (1954). *Acta Cryst.* **7**, 452–453.

HAYKIN, S. (Editor) (1979). *Nonlinear Methods of Spectral Analysis. Topics in Applied Physics*, Vol. 34. New York: Springer-Verlag.

HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature* (London), **290**, 107–113.

HÖRMANDER, L. (1973). *An Introduction to Complex Analysis in Several Variables*, 2nd ed. Amsterdam: North-Holland.

JAYNES, E. T. (1957). *Phys. Rev.* **106**, 620–630.

JAYNES, E. T. (1978). *IEEE Trans.* SSC-**4**, 227–241.

KARLE, J. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML & B. SEDLACEK, pp. 155–164. Copenhagen: Munksgaard.
KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181–187.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.
KHINCHIN, A. I. (1949). *Mathematical Foundations of Statistical Mechanics.* New York: Dover Publications.
KITAIGORODSKII, A. I. (1961). *The Theory of Crystal Structure Analysis.* New York: Consultants Bureau.
KLUG, A. (1958). *Acta Cryst.* **11**, 515–543.
LANCZOS, C. (1970). *The Variational Principles of Mechanics*, 4th ed. Univ. of Toronto Press.
LANG, S. (1965). *Algebra.* Reading, MA: Addison Wesley.
LAPLACE, P. S. DE (1820). *Théorie Analytique des Probabilités.* Paris: Gauthier–Villars.
LEVINE, R. D. & TRIBUS, M. (Editors) (1979). *The Maximum Entropy Formalism.* Cambridge: MIT Press.
LINNIK, I. JU. (1975). *Math. USSR-Izv.* **9**, 1323–1332.
LUKACS, E. (1983). *Developments in Characteristic Function Theory.* London: Griffin.
LUZZATI, V. (1955). *Acta Cryst.* **8**, 795–806.
NARAYAN, R. & NITYANANDA, R. (1982). *Acta Cryst.* **A38**, 122–128.
NAYA, S., NITTA, I. & ODA, T. (1965). *Acta Cryst.* **19**, 734–747.
NILSSON, N. J. (1971). *Problem-Solving Methods in Artificial Intelligence.* New York: McGraw-Hill.
PALEY, R. E. A. C. & WIENER, N. (1934). *Fourier Transforms in the Complex Domain. Am. Math. Soc. Colloquium Publications*, Vol. 19. Providence: American Mathematical Society.
PIRO, O. E. (1983). *Acta Cryst.* **A39**, 61–68.
RANGO, C. DE, MAUGUEN, Y. & TSOUCARIS, G. (1975). *Acta Cryst.* **A31**, 227–233.
RANGO, C. DE, TSOUCARIS, G. & ZELWER, C. (1974). *Acta Cryst.* **A30**, 342–353.
RICE, S. O. (1944). *Bell System Tech. J.* **23**, 1–162. Reprinted in *Selected Papers on Noise and Stochastic Processes* (1954), edited by N. WAX, pp. 133–294. New York: Dover Publications.
RICHE, C. (1970). *C. R. Acad. Sci. Paris*, **271**, 396–398.
RICHE, C. (1973). *Acta Cryst.* **A29**, 133–137.
ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton Univ. Press.
RUDIN, W. (1973). *Functional Analysis.* New York: McGraw-Hill.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
SCHENK, H. (1981). *Acta Cryst.* **A37**, 573–578.
SCHULZ, G. E. & SCHIRMER, R. H. (1979). *Principles of Protein Structure.* New York: Springer-Verlag.
SCHWARTZ, L. (1966). *Théorie des Distributions.* Paris: Hermann.
SHANNON, C. E. (1949). *Proc. Inst. Radio Eng. NY*, **37**, 10–21.
SHANNON, C. E. & WEAVER, W. (1949). *The Mathematical Theory of Communication.* Urbana: Univ. of Illinois Press.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SPRACKLEN, D. & SPRACKLEN, K. (1978). *SARGON: A computer chess program.* Rochelle Park: Hayden Book Company.
SZEGÖ, G. (1920). *Math. Z.* **6**, 167–202.
TOEPLITZ, O. (1911). *Rend. Circ. Mat. Palermo*, **32**, 191–192.
TSOUCARIS, G. (1970). *Acta Cryst.* **A26**, 492–499.
WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions*, 2nd ed. Cambridge Univ. Press.
WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* **A31**, 53–56.
WHITTAKER, E. T. & WATSON, G. N. (1927). *A Course of Modern Analysis.* Cambridge Univ. Press.
WIDOM, H. (1960). *Trans. Am. Math. Soc.* **94**, 170–180.
WIDOM, H. (1975). *Publ. Math. IHES*, **44**, 191–240.
WIENER, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* Cambridge: MIT Press.
WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* **A39**, 47–60.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1984). **A40**, 445–450

Multiple Laue Rocking Curves

BY D. PETRASCHECK

Institut für Theoretische Physik, Universität Linz, 4040 *Linz, Austria*

AND H. RAUCH

Atominstitut der Österreichischen Universitäten, 1020 *Wien, Austria*

(*Received* 12 *August* 1983; *accepted* 7 *February* 1984)

Abstract

Multiple Laue rocking curves of perfect crystals show a narrow central peak with a width of some 10^{-3} s arc and an additional oscillatory structure. The finite structure of these curves is analyzed for two- and three-crystal Laue arrangements. These profiles can be used for precise determinations of structure factors and for an extension of small-angle scattering experiments to the extreme small-angle regime where large objects and long-range particle correlations become visible. An extremely high angular resolution can be achieved without significant reduction of the intensity, owing to a decoupling of the angular resolution from angular width of the beam. The analytically calculated rocking curves are compared to numerical results and to experimental results and show good agreement with both.

Introduction

The dynamical diffraction of X-rays and neutrons on multiple-perfect-crystal arrangements has been studied extensively during the past years. Monolithic and polylithic, plane and bent, static and vibrating systems have been discussed. Dynamical focusing

0108-7673/84/040445-06$01.50