

TOPICAL REVIEW

Acta Cryst. (1996). **A52**, 561–589

Maximum Entropy and Bayesian Statistics in Crystallography: a Review of Practical Applications

CHRISTOPHER J. GILMORE

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

(Received 8 June 1995; accepted 8 March 1996)

You should call it 'entropy' and for two reasons: first, the function is already in use in thermodynamics under that name; second, and more importantly, most people don't know what entropy really is, and if you use the word 'entropy' you will win every time!

von Neumann's advice to Shannon as to what to call the function $-\sum_i p_i \log p_i$

Abstract

A survey of practical applications of the maximum entropy method to problems in crystallography is presented along with the related field of Bayesian statistical inference. Included in this review are problems of the processing of intensity data, the crystallographic phase problem, accurate electron-density studies and algorithms for the constrained maximization of entropy. The discussion of the phase problem reviews the solution of structures from powder and single-crystal diffraction data sets, as well as applications to electron and macromolecular crystallography. It is believed that the methodology has a secure future and a sound statistical basis, but it is still not a technique that is regularly employed by crystallographers and the reasons for this are also examined.

1. Introduction

Aspects of the maximum entropy (ME) formalism and Bayesian statistics have permeated the crystallographic literature for about 20 years, yet many crystallographers are quite justifiably puzzled by them. Practical applications are reported in the fields of crystallographic data processing, solving various facets of the phase problem, deriving accurate charge densities and

deconvoluting powder diffraction overlaps, as well as numerous theoretical papers, often with very diverse approaches. Yet the methodology has not fully entered main-stream crystallography and it makes only fleeting appearances, it at all, in current textbooks, despite considerable, and often justifiable, claims as to its powers and its statistical and mathematical correctness.

Why is this? What aspects of the ME formalism and Bayesian methods can be of value in a routine way to crystallographers? What algorithms are of most utility in a crystallographic environment? This review attempts to answer these questions in a practical way. Like everything associated with the method, it is bound to arouse some controversy and it is not the intention to avoid areas of dispute. At the very least, we hope to provoke some thought on the best fields in which to apply the technique and how best to use it in a *practical* way to solve *practical* problems.

ME theory is an integral part of a much wider topic – Bayesian statistics and, in particular, Bayesian inference, the latter being defined as a method of plausible scientific reasoning. Indeed, it makes most sense when considered as part of this field, and so the scope of this review includes Bayesian methods also, even when they are applied without recourse to maximum entropy, although the latter, with its much more extensive crystallographic literature, will naturally dominate. Following an introduction to the methodology in §2, the areas to be investigated in detail are:

(i) *Processing intensity data.* §3 discusses the extraction of intensities from single-crystal diffraction data. This is then extended to the more complicated problem of the deconvolution of overlapped reflections in powder diffraction and related topics.

(ii) *Solving the phase problem.* §4 is the longest and deals with a wide variety of phasing problems that lend themselves so well to ME techniques. Such problems

Chris Gilmore received his doctorate from Bristol University in 1971 having worked with Peter Woodward. For the past 25 years, he has worked on direct methods techniques, and is the author of the MITHRIL computer program. Since 1986, he has focused on maximum entropy methods, applying them to structures that produce incomplete or inaccurate data; these include powder and electron diffraction data sets, as well as macromolecules and membranes. He is a member of the BCA Council. Other passions include mountains and music.

are largely associated with data sets that are incomplete in the way they sample reciprocal space and/or have data at less than atomic resolution, so that we discuss in detail powder and fibre diffraction, electron diffraction and aspects of macromolecular crystallography.

(iii) *Accurate charge densities.* The ME method has been used to derive accurate electron densities using both single-crystal and powder diffraction data. How reliable are these results, especially when compared with multipole refinement methods? Is this a viable method for producing accurate electron densities? Is the technique suitable for spin densities? All this is discussed in §5.

(iv) *Small-angle scattering and related techniques.* §6 surveys the current literature.

(v) *Maximum-entropy algorithms.* In §7, there is a survey of how entropy is best subjected to constrained maximization and which algorithms work best in given circumstances.

(vi) *Finally, what of the future?* In §8, there is a summary and a brief look at the future.

These topics cannot always be considered independently from each other, so that various sections will inevitably overlap but at the very least it is hoped to present a non-specialized overview of an exciting and fundamentally significant area of crystallography in action. Maximum likelihood methods are not unrelated to the statistical methods being surveyed here, but to include them would extend the scope of this review far beyond its original boundaries; they are, therefore, excluded.

2. The theoretical background

2.1. The maximum entropy principle

Consider first a discrete probability distribution for which we know a set of N normalized probabilities $P = \{p_1, p_2, \dots, p_N\}$. Such a distribution has an entropy, S , or information content I given by

$$S = -I = -\sum_{j=1}^N p_j \log p_j. \quad (1)$$

The logarithm in (1) can be to any base; often used bases are 2 (in which case the entropy is measured in bits) and e (natural units). Equation (1) was first defined and used in a non-thermodynamic context by Shannon when working with problems of the capacity of communication channels, the transmission of signals down noisy lines and coding processes (Shannon & Weaver, 1949), so creating the branch of science known as information theory. In image processing, (1) is called a regularizing function.

Where does the entropy function $-\sum_i p_i \log p_i$ come from? Shannon uses very simple probability arguments. More familiar, from statistical mechanics, may be a derivation from combinatorial models and, before

proceeding to a simple example of entropy in action, it will be useful to reproduce it here. It stems from work by Boltzmann. Consider an observed image divided into pixels. This image can be in any number of dimensions: a simple spectrum is one-dimensional, a photograph is two-dimensional *etc.* Now suppose we have a team of monkeys (why is it always monkeys?) generating trial images by strewing dots at random, and in such a trial N_i such dots land in the i th pixel. Let there be n pixels in total and the total number of dots be $\sum_{i=1}^n N_i = N$. The image has an associated multiplicity, W :

$$W = N!/N_1!N_2! \dots N_n!, \quad (2)$$

which is simply the number of possible ways that the N dots can be distributed in the n pixels. If the resulting image disagrees with the data it is rejected; if it agrees then it is saved. In this way, a set of images that agree with the observed data is generated. Using Stirlings's formula for large N , (2) becomes

$$\begin{aligned} \log W &= N \log N - \sum_{i=1}^n N_i \log N_i \\ &= -N \sum_{i=1}^n (N_i/N) \log N_i/N \\ &= -N \sum_{i=1}^n p_i \log p_i, \end{aligned} \quad (3)$$

where $p_i = N_i/N$, the fraction of dots found in pixel i . The entropy function can thus be seen as a multiplicity measure. In addition, in the language of classical ME theory, we expect that virtually all the feasible images generated by the monkeys will be close to that of maximum entropy. As we shall see, this classical approach has only limited applicability in the crystallographic context.

Jaynes extended Shannon's ideas to create a general formalism of great power and elegance, which he applied to a wide variety of problems in physics [see, for example, Jaynes (1957, 1983, 1986, 1989) – in fact anything written by Jaynes is well worth reading.] He also put the ME formalism into an historical perspective in the line of Bernoulli, Laplace, Jeffreys and Cox along with Maxwell, Boltzmann and Gibbs (Jaynes, 1979).

Jaynes overcame dimensionality problems in (1) by defining a relative or cross entropy H :

$$H = -\sum_{j=1}^N p_j \log(p_j/m_j), \quad (4)$$

where the m_j are normalized probabilities representing our prior knowledge of an event or events. Where we have no prior information, these probabilities are taken as uniform and H now measures a gain (or loss) of information in moving from a prior probability m_j to an *a posteriori* probability p_j . In the continuous case, (4) becomes

$$H = - \int p(x) \ln[p(x)/m(x)] dx, \quad (5)$$

where $m(x)$ refers to a prior probability distribution, often taken as uniform. As we shall see, the ability to incorporate prior knowledge into the calculations is a major benefit of the ME method.

Suppose we have m constraints, expressed as expectation values, $\langle A_r \rangle$, which represent our knowledge derived from experiment or, alternatively, are in the form of hypotheses. The constraints can be written

$$\sum_{i=1}^N A_{ri} p_i = \langle A_r \rangle, \quad r = 1, 2, \dots, m, \quad (6)$$

and from this we wish to derive the set of probabilities $\{P\}$. Often, the situation under study is a mathematically indeterminate one in which $N \gg m$. Herein is another appeal of the method in many situations, but it should be remembered that many other regularizing functions have this same property: it is not unique to entropy maximization. The ME approach consists of maximizing the entropy (2) or minimizing the information content, subject to the constraints (6), to determine $\{P\}$. The method is clearly being used as a method of inference, *i.e.* we are inferring the probabilities $\{p_1, p_2, \dots, p_n\}$ in an optimally unbiased way with the constraints providing the ability to incorporate prior beliefs. Probabilities will be used here in a general way often simply related to map or image pixels through normalization as in §2.2.

It is interesting to note that Shannon himself considered this extension of the method to fields outside communication theory to have no fundamental significance (Tribus, 1979). For a relatively simple discussion of the use of entropy maximization in this way and in a crystallographic environment, see Collins (1985).

To make the situation clearer in a practical situation, the next section discusses a simple but important example – the Brandeis dice problem.

2.2. The Brandeis dice problem

This classical problem (and something of a *cause célèbre*) may help to illuminate this formalism [Jaynes (1979), but see also Prince, Sjölin & Alenljung (1988), who discuss it as an introduction to ME theory, but solve the equations differently using dual functions – see §7]. Consider a dice. When it is thrown, there are six possible results i where $1 \leq i \leq 6$. If it is thrown N times, an ‘honest’ dice will give all $p_i = 1/6$, and an average throw of 3.5 when thrown repeatedly. Suppose, however, we have a dishonest or biased one for which the mean is 4.5, *i.e.*

$$\sum_{i=1}^6 i p_i = 4.5. \quad (7)$$

Given this information *and nothing else*, what probability $p_{i,i=1,6}$ should we assign to the i spots on the next throw?

Clearly, the problem is indeterminate and it is possible to devise an infinity of solutions. (For example, the die only gives $i = 5$ or $i = 4$ with equal probabilities.) The ME principle, however, generates a unique solution, which, in the jargon of the subject, is maximally unbiased, has a minimum information content and is declared maximally non-committal with respect to missing information.

There are two constraints: equation (7) and the normalization condition

$$\sum_{i=1}^6 p_i = 1. \quad (8)$$

Assuming uniform prior probabilities, we maximize S subject to the constraints (7) and (8). This is a standard variation problem solved by the use of Lagrangian multipliers. Let λ be the Lagrangian associated with constraint (7), then

$$p_i = [1/Z(\lambda)] \exp \left[-\lambda \sum_{i=1}^6 i p_i \right]. \quad (9)$$

The Z function is a normalizing constant that subsumes constraint (8):

$$Z(\lambda) = \sum_{i=1}^6 \exp \left[-\lambda \sum_{j=1}^6 j p_j \right]. \quad (10)$$

A numerical solution using standard variation methods gives $\{p_1, \dots, p_6\} = \{0.05435, 0.07877, 0.11416, 0.16545, 0.23977, 0.34749\}$, with an entropy of 1.61358 natural units.

Now what does this result mean? To quote Jaynes, ‘... our result as it stands is only a means of describing a state of knowledge about the outcome of a single trial. It represents a state of knowledge in which one has only (i) the enumeration of the six possibilities, and (ii) the mean value constraint ... *and no other information*. The distribution is maximally noncommittal with respect to all other matters; it is as uniform ... as it can get without violating the given constraint.’

Controversy still rages around this simple result. For example, a referee of this paper states that ‘... the ME solution is not really a plausible solution to the problem as stated by Jaynes is perhaps a source of the general scepticism toward ME’. There is also a hostile critique by Rowlinson (1970), staunchly attacked by Jaynes (1979). A critical appraisal would take us too far afield, but this example is at least a simple illustration of the formalism at work. Obviously, the dice problem is a very marginal situation and, to quote the same referee again, ‘... one would be very foolish to place any bets on the basis of so little information’. This is certainly true but proponents of the ME formalism will argue:

What do you do in a situation like this – give up or extract what little information you can?

In the dice problem, we have fitted the constraint (our knowledge of the average throw of the die) exactly. In most experimental contexts, however, this may not be appropriate since the constraints are subject to experimental error. Often a reduced χ^2 statistic is used as a measure of fit between observed and calculated constraints:

$$\chi^2 = [1/(N - r)] \sum_{j=1}^r (1/\sigma_j)^2 [(A_j) - D_j]^2, \quad (11)$$

where D_j is the calculated value of the corresponding constraint (A_j) and σ_j is the corresponding standard deviation. A target value of $\chi^2 = 1.0$ is often used. This can increase the computational problem since one is compelled to find the entropy maximum on a contour of constant χ^2 but modern algorithms can address this quite readily for the situation where the constraints are convex; it is more problematic when this is not the case. It should also be noted that this use of (11) assumes that the errors are either normally distributed or close to it, which may not always be the case.

2.3. More on the ME principle

As we have already indicated, this methodology can be controversial [for a further flavour of this, see Jaynes (1987), or the discussion section in Donoho, Johnstone, Hoch & Stern (1992)] and, indeed, it is sometimes applied in very inappropriate ways. However, to place the procedure on a secure statistical basis, Shore & Johnson (1980*a,b*) have used the theory of statistical inference to show that, indeed, constrained entropy maximization is a procedure of optimal consistent inference. In this case, the entropy function $-\sum_i p_i \log p_i$ is derived from a set of postulates, some of which are the basic axioms of probability theory and others are based on heuristic arguments:

(i) The result of any statistical inference should be unique.

(ii) It should also be invariant with respect to any coordinate system or transformation.

(iii) For the situation where information is independent, the results of an inference procedure should give the same answer if the information is used independently or jointly.

(iv) Item (iii) should also hold for independent subsystems.

Tikochinsky, Tishby & Levine (1984) discuss reproducible experiments and consistency conditions to show that under linear constraints the ME principle is the only consistent algorithm for inferring discrete probability distributions with given constraints, although there is some dissension here as well as in Shore & Johnson's work (see, for example, Titterton, 1984).

Bricogne (1984, 1988*a*) has also derived the maximum entropy equations in a crystallographic phasing environment using the Daniels saddlepoint method (Daniels, 1954). [There is a good review of the saddlepoint method, a formalism that most crystallographers seem to ignore even when it could be of considerable value to them, by Reid (1988).] Here, there is no recourse to the terminology or mathematics of information theory at all, and yet Bricogne (1984, 1988*a*) has obtained equations that are identical to those derived from the ME approach. This can be viewed as an independent justification for the method, provided we are working with probabilities. For some researchers, the method is not used in its strict probabilistic mode but only as a smoothing or regularizing function for the improvement of images of which electron-density maps are a three-dimensional example. The underlying probability rationale can be lost in this situation, as sometimes happens when working with accurate electron densities.

The relationship of the ME formalism with thermodynamics takes us far away from the focus of this review, but Tribus (1961) has shown how all the laws of classical thermodynamics can be derived using the ME principle. In a later review (Tribus, 1979), he states: 'The debates engendered by this approach have been extensive and, on occasion, bitter. I have concluded from them that thermodynamics is as much a branch of theology as it is a branch of science!'

From this very brief description, maximum entropy may seem an isolated statistical technique but in fact it is best perceived as a part of Bayesian statistics, and this topic now needs to be discussed.

2.4. Bayesian statistics

Bayes's theorem, on which Bayesian statistics is based, appears in many guises. It was first derived by the 18th century English clergyman and talented mathematician Thomas Bayes, and found in his papers on his death. [See Molina (1963) for a reprint of this paper and a critical discussion.] It was rediscovered by Laplace, who used it to solve a wide variety of outstanding problems in astronomy. At its simplest, the theorem is uncontroversial. Let us suppose that we are carrying out an experiment to gain information on a parameter θ . Before commencing the experiment, we have prior belief of the possible values of θ , which we express as an unconditional probability distribution $p(\theta)$. If the experiment yields data represented by the vector \mathbf{x} , then Bayes's theorem states

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta). \quad (12)$$

The constant of proportionality may be determined by normalization. The term $p(\theta|\mathbf{x})$ is called the posterior distribution (or simply posterior); the function $p(\mathbf{x}|\theta)$ is

the likelihood, which consults the data from the experiment, and $p(\theta)$ is the prior probability.

As written, this is an uncontroversial equation simply derived from the axioms of probability theory and used by all statisticians to invert conditional probabilities, *i.e.* to obtain $p(\mathbf{x}|\theta)$ from $p(\theta|\mathbf{x})$ and *vice versa*. However, Bayesian statisticians take the topic a whole step further by using (12) as a model of scientific inference. In addition, probabilities are defined not in the narrow frequentist way but in the broader, and much more controversial, sense of degree of belief, which is ideally suited to science where measurements can be very difficult to relate to frequencies without the use of very convoluted argument, and a counter-intuitive use of language. Furthermore, (12) is used iteratively: the posterior becomes an updated prior and the whole cyclic process repeated until convergence is reached. From these departures from classical statistical methods stem a great deal of sometimes very acrimonious dissent. (See, for example, Jaynes 1979.) It is not possible or desirable to pursue the intricacies of the Bayesian method but here we can recommend two books by O'Hagen (1988, 1994). The classic text on likelihood is by Edwards (1972), although this is quite strongly critical of entropy techniques. In the crystallographic context, there is a very clear exposition by French (1978), who describes a three-stage model in crystallography from a Bayesian viewpoint, and an elegant paper concerning the crystallographic phase problem in a Bayesian context by Bricogne (1988*a*). Finally, there is a clear account of the formalism in the context of the deconvolution of overlapping powder diffraction peaks by Sivia, David, Knight & Gull (1993).

How does the ME principle fit into the Bayesian framework? A common objection to the use of Bayes's theorem is that a systematic error or bias in the prior $p(\theta)$ will give rise to similar errors in the posterior $p(\theta|\mathbf{x})$. The use of the likelihood, *i.e.* the intervention of the experimental data, makes this less of a problem than may appear at first but the ME formalism becomes a very powerful tool here: we chose a prior *via* maximum entropy, which has a minimum bias and is maximally non-committal towards missing data. This is then an optimal prior for use with Bayes's theorem. This turns out to be

$$p(\theta) = \exp[\text{entropy}(\theta)]. \quad (13)$$

A prior with a large entropy will have a minimum of artefacts or structure when compared with one derived under similar constraints that has a small entropy.

For example, in the Brandeis dice problem, once the ME prior is set up; the die is then thrown. This experimental result is used to modify the ME prior *via* Bayes's theorem, and this process continues as data accumulate. As a general text on the use of the ME method as a tool for processing data in both Bayesian and non-Bayesian contexts, we recommend Buck &

Macaulay (1991), although it presents very diverse approaches to the methodology.

In summary:

(i) Bayes's theorem tells us how to modulate prior probabilities as extra evidence accrues.

(ii) ME theory tells us how to assign these prior probabilities in the first place. It readily admits the use of prior knowledge and also works in mathematically indeterminate situations.

We are now ready to survey the literature for applications of Bayesian methods and the ME formalism to practical problems in crystallography. We will begin with the processing of intensity data.

3. Data processing

3.1. Single-crystal data

A frequent problem that arises in the collection and measurement of intensity data concerns the weak reflections. Their true intensities are always non-negative but experimentally this may not be the case because of counting errors. It is traditional to treat negative intensities as zero; indeed, at first sight, there seems little else to do with them, but in one of the earliest papers in the crystallographic literature to employ the Bayesian methodology, French & Wilson (1978) have described an alternative approach using Wilson statistics (Wilson, 1949) as a source of prior information, and the method has been utilized in a diffractometer profile processing program (Oatley & French, 1982).

What prior information do we have about the intensity, I , of a reflection? In addition to the simple requirement of non-negativity, we know that, taken as a whole, any moderate or large data set satisfies Wilson statistics. Thus, for acentric reflections, our prior knowledge is

$$\begin{aligned} p(I) &= \Sigma^{-1} \exp(-I/\Sigma), & I \geq 0, \\ p(I) &= 0, & I < 0, \end{aligned} \quad (14)$$

while, for the centric case,

$$\begin{aligned} p(I) &= (2\pi\Sigma I)^{-1} \exp(-I/2\Sigma), & I \geq 0, \\ p(I) &= 0, & I < 0, \end{aligned} \quad (15)$$

where Σ is the mean intensity of the relevant shell in reciprocal space, which is currently unknown because the individual intensities are unknown. It is important to remember that Wilson statistics apply to a random and uniform distribution of atoms in the unit cell; this works well enough for medium-sized structures but is quite inappropriate for proteins because of solvent regions, very small structures, situations where there are dominant scatterers, as well as any structure that displays a high degree of atomic regularity.

We can now apply Bayes's theorem to this situation: all the intensities, I , whether positive, zero or negative, are used to estimate Σ ; from this, we obtain a posterior distribution estimate of the intensity J , $p(J|I)$, from which we can estimate the expectation value of the intensity given I , $E(J|I)$:

$$E(J|I) = \int_0^{\infty} Jp(J|I) dJ \quad (16)$$

and its associated variance $\text{var}(J|I)$:

$$\text{var}(J|I) = \int_0^{\infty} [I - E(J|I)]^2 p(J|I) dJ. \quad (17)$$

It seems to us that this is a secure basis on which to process data and one that should be used routinely, although one must always bear in mind the caveat of the applicability of Wilson statistics. Why is it not used in this way? It seems to be absent in many crystallographic computing packages, but *CCP4* (1993) and *XTAL* (Hall & Stewart, 1992) are two that do provide it. Blessing (1987, 1989) reports that the method is helpful with data sets that have a large fraction of weak data. However, Watkin (private communication) expresses the experience that the method often overestimates the weak reflections; perhaps this is a consequence of using Wilson statistics or a lack of accuracy in the measurement of weak reflections, but obviously the method is extendable beyond such restrictions. For an example of the modified use of the Bayesian approach that does not use Wilson statistics as a prior, see Lenstra *et al.* (1991). In their case, working on $C_{11}H_{18}N_2O_2$, they succeeded in incorporating an extra 1363 weak reflections into the least-squares refinement using Bayesian analysis of the weak reflections; the R factor naturally increased but the variances of the atomic coordinates fell by a factor of two. It is also worth mentioning here that the learned-profile method of diffractometer data collection is, in reality, a Bayesian method in which knowledge of peak profiles is accumulated and used during data collection (Clegg, 1981). Related to this is a technique devised by Wilkinson, Khamis, Stansfield & McIntyre (1988), in which weak reflection intensities are extracted *via* a peak profile prior derived from strong reflections that are nearby in reciprocal space. It seems to be successful when applied to neutron area detector data.

Another problem that arises in data processing is the scaling of intersecting data sets where a given reflection h has its intensity I measured in several data sets (multiple films or frames from an area detector) i . Denote such an intensity as I_{hi} and the corresponding average intensity as J_h and let G_i be the scale factor that puts J_h on the scale of I_{hi} . Collins (1984) (see also Prince & Collins, 1992) has proposed the use of the entropy functional

$$S = - \sum_{h,i} (G_i J_h)' \log[(G_i J_h)' / I_{hi}], \quad (18)$$

where the prime denotes normalization. Q is maximized by setting

$$\partial S / \partial G_i = 0 \quad (19)$$

and the procedure is used iteratively to give an optimal value for J_h . Although there are some problems with weights, the scheme converges well even with difficult data and the method offers a robust and error-resistant method of scaling, at least with the data used as a test case, although the purist can question the use of $(G_i J_h)$, even when normalized, as a probability measure in the entropy expression (18).

There is also a second problem here that recurs throughout this review (it is by no means unique to this problem): There are a number of papers proposing techniques employing a ME or Bayesian formalism that seem eminently worthwhile but lead nowhere in the sense that they are not taken up and used as part of the crystallographic armory. This makes a critical assessment very difficult. One cure for this problem is quite simple – program the ideas, provide a good interface to existing packages and distribute the code freely!

When collecting data on an area detector, the 2D diffraction pattern is measured as N pixels, usually of equal size, and the result is a set of accumulated counts in each of the pixels $\{p_{ij}\}$. A problem arises as to the determination of the positions of the Bragg peaks in this set. Just as for obtaining intensities, the strong reflections really pose no problems, weak reflections do. Lehmann, Robinson & Wilkins (1986) carry out a constrained entropy maximization to obtain a peak boundary and then use the raw data to obtain a peak intensity. Each pixel is normalized by dividing by the total counts. It is now treated as a probability measure. Three constraints are imposed:

(i) The total number of counts is constant.

(ii) Let p_{ij} be the calculated count in pixel i, j and let o_{ij} be the recorded normalized count for the same pixel. The reduced χ^2 statistic defined by

$$\chi^2 = N^{-1} \sum_{ij} w_{ij} (p_{ij} - o_{ij})^2 \quad (20)$$

is set to unity and the w_{ij} are weights approximated by

$$w_{ij} = N / \sum_{ij} o_{ij}. \quad (21)$$

(iii) We know that the Bragg peak is smooth. The entropy functional imposes smoothness but additional local smoothness is imposed by coupling intensities of adjacent peaks.

With simulated data, the results are convincing, although this method or any related one also does not seem to have found common acceptance or usage.

3.2. Powder diffraction data

In powder diffraction, a three-dimensional diffraction pattern is collapsed into one dimension giving rise to the familiar problem of peak overlap, which exacerbates all the problems found in the single-crystal case. The overlaps are of two sorts: those that arise accidentally as a result of lattice parameters and Bragg angle and those that are a consequence of point-group symmetry, and for which the overlap is exact. Clearly, the latter are a much more difficult problem than the former but both serve to limit the applicability of powder diffraction as a routine tool for determining and refining crystal structures from powders. The synchrotron has made a profound difference to the problems of structure solution, but the intrinsic overlap arising from line broadening and the existence of exact peak overlap in high-symmetry space groups still impose very severe limitations on what problems can and cannot be addressed. For a comprehensive and readable review of all these difficulties, see Shirley (1984). We will discuss the problems of *ab initio* structure solution from such data sets in §4.5, but it is relevant to survey here the Bayesian and ME methods that have been used for overlap deconvolution at the point of data processing. For a brief, but useful, survey of all the other non-Bayesian methods that have been proposed, see Estermann, McCusker & Baerlocher (1992).

In practice, the field is not large. The most noteworthy is a recent paper by Sivia & David (1995), which extends the French & Wilson approach to include not only profile fitting but peak deconvolution as well. It follows an earlier approach to the problem by the same authors (Sivia, David, Knight & Gull, 1993) using polynomial deconvolution. They propose two possible priors: (1) the Wilson and (2) the Jeffreys prior. The latter (Jeffreys, 1939) takes the form

$$p(\theta) \propto 1/\theta. \quad (22)$$

It is equivalent to a probability density function that is uniform with respect to $\log(\theta)$ and arises when there is complete ignorance of the absolute scale of the intensity data. Fig. 1 shows this methodology in action for benzene. The ability to deconvolute almost completely overlapped reflections and to deal with the problem of negative intensities is noteworthy.

It is too early to cast a definitive judgement on the method proposed but it does look exciting, and is extendable to non-Wilson statistics. Earlier, David (1987, 1990) had proposed maximizing the entropy of a Patterson map as a method of resolving even exact symmetry-imposed overlaps. All the intensities, whether overlapped or not, are used to generate a Patterson map; its entropy is maximized subject to intensity constraints and this is followed by an inverse Fourier transform, which yields estimates for individual intensities. The method worked well for the test case of

TiO₂ in space group $P4_2/mnm$ in resolving even overlaps arising from point-group symmetry and is capable of working with neutrons without resorting to multichannel formalisms (Bricogne, 1988a), but it is limited in the complexity of the problem that can be

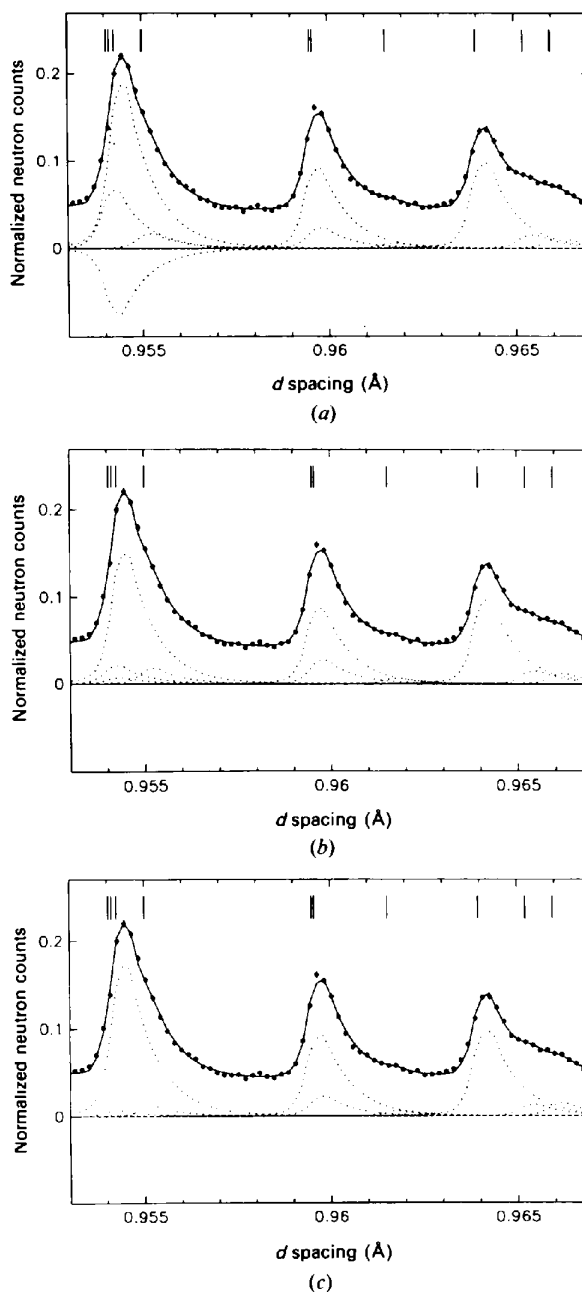


Fig. 1. The observed and calculated powder diffraction pattern of benzene between 0.953 and 0.967 Å. (a) The results of least-squares fitting using the Pawley method. (b) The square of individual inferred structure-factor amplitudes from the Bayesian analysis. (c) The calculated intensities given by the refinement of the crystal structure. (From Sivia & David, 1995.)

tackled. David estimates that it could be of value in cases where the cell volume is less than 1000 \AA^3 but this may be rather optimistic (TiO_2 has a unit-cell volume of only 62.4 \AA^3).

Working purely in a maximum entropy environment, Steenstrup & Wilkins (1989) have also demonstrated its use in the processing of energy-dispersive powder diffraction patterns from synchrotron radiation involving $\text{EuBa}_2\text{Cu}_3\text{O}_{9-\delta}$. They obtain better looking spectra in which it is easier to identify diffraction peaks and fluorescent lines.

It is worth noting also that the ubiquitous Rietveld method (Rietveld, 1969) is also a Bayesian non-ME method of peak deconvolution for powder diffraction data, not unrelated to that of Wilkinson *et al.* (1988) as described in §3.1 in which weak peaks are extracted from profile parameters derived from strong reflections.

In conclusion, Bayesian approaches have much to offer here especially with weak data sets. In subjects other than crystallography, ME methods have often demonstrated considerable power in extracting weak signals from noisy backgrounds – it is something they can do very well and it seems to us that this is a field with much to offer, but there is too little in the way of available software at present.

4. The phase problem

4.1. Introduction

We now come to the most difficult and fundamental problem in crystallography – that of the phase problem. Direct methods are now routine for all light-atom structures with up to *ca* 150 atoms in the asymmetric unit [see Woolfson (1987) for a historical review of this subject and Woolfson & Fan (1995) for an excellent new monograph], and even small proteins such as crambin and rubredoxin have also been solved in this way (Hauptman, Weeks, Smith, Teeter & Miller, 1993), whilst the heavy-atom method is an almost routine technique for solving most biological macromolecular structures. However, there is still a great need for a more powerful and consistent formalism for approaching structure solution. In particular, there are a number of areas involving both small and large molecules where existing methods either fail completely or are very laborious. For small molecules, these are cases where the data are either incomplete and/or extend to less than atomic resolution, *i.e.* less than *ca* 1.1 \AA . This means that structures from powder or fibre diffraction and electron diffraction can be very difficult to solve, especially if there are no heavy atoms present. For example, how many *unknown* all-light-atom structures (say nothing heavier than oxygen) with more than 30 non-H atoms in the asymmetric unit have been solved *ab initio* from powder diffraction data? The answer is none. In protein crystallography, of

course, the ultimate challenge is the *ab initio* solution of a native protein without derivative data. Is such a calculation possible?

We will now address these problems individually, starting with a short theory section, and assess what has been achieved so far by ME and Bayesian approaches. The author does, however, have a problem here: a lot of the results to be presented are partly his own work, and this presents difficulties when writing a review. Nonetheless, a good measure of objectivity will be attempted!

4.2. Some necessary theory and notation

We start with a set of unitary structure factors $|U_{\mathbf{h}}|^{\text{obs}}$, which are derived from the intensity data by normalization procedures. (For N point atoms of unit weight, $|U_{\mathbf{h}}|^{\text{obs}} = 1/N |F_{\mathbf{h}}|^{\text{obs}}$.) The notation U will be used to define an arbitrary vector of n U 's, $U = (U_{\mathbf{h}_1}, U_{\mathbf{h}_2}, U_{\mathbf{h}_3}, \dots, U_{\mathbf{h}_n})$. We also have a set of phase angles $\varphi_{\mathbf{h}}$, most of which are either unknown (the *ab initio* case) or only approximately determined. For *ab initio* structure solution, because of rules of origin and enantiomorph definition, some phases can usually, but not always (it depends on the space group), be assigned subject to certain well known rules, whereas, in the case of macromolecular crystallography, there will be a large set of such phases, albeit rather inaccurately known at times, derived from isomorphous replacement methods. In electron diffraction, the Fourier transform of the corresponding image may give useful phase information.

Now, traditional direct methods attempt to determine the unknown phase angles from probabilistic considerations working in a reciprocal-space formalism in which triplets and negative quartets are generated as relationships between phases and these relationships used in tangent-like formulae to refine and extend phase angles. Only when sufficient phases are determined do we switch to real space with the generation of trial electron-density maps. What are the underlying mathematical principles here? Bricogne (1988*b*, 1991*b*) is recommended for a clear relatively non-mathematical description and we shall follow his arguments closely in this and the next section.

First: a model in which random atoms are uniformly distributed in the unit cell (the Wilson model) has been used. This is not really problematic for most small molecules but is significantly in error for macromolecules, where the structure comprises disordered solvent regions and ordered protein (see §3.1).

Second: we have constructed joint probability distributions of small collections of structure factors, usually three (triplets), $P(U_{\mathbf{h}}U_{\mathbf{k}}U_{-\mathbf{h}-\mathbf{k}})$, or four (quartets), $P(U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{l}}U_{-\mathbf{h}-\mathbf{k}-\mathbf{l}})$.

These are marginal distributions into which we have substituted the known structure-factor amplitudes to get

joint conditional distributions which act as indicators of the most probable combinations of the associated phases, and these are the relationships that get used in the tangent formula.

Problems arise from this approach:

(i) To make the mathematics of deriving the joint probabilities tractable, severe approximations are used (either the Edgeworth series or related Gram-Charlier methods), which make the probability distributions accurate only for very small U magnitudes, whereas, in practice, we need to phase the large ones!

(ii) The triplets and quartets are marginal distributions with every unwanted phase or magnitude either absent or integrated out. What we really need is to develop phases *via* a probability distribution of the form $P(\text{new phase}|\text{everything we currently know})$ without resorting to marginal distributions at all. But, to quote Bricogne (1988b): 'there is no tractable unique expression for $P(\mathbf{F})$ into which we could substitute the measured values of the large moduli to get the conditional distribution of the corresponding phases...'

With this in mind, let us now investigate the ME approach to phasing.

We need to extend our nomenclature: those reflections that are phased comprise the basis set $\{\mathbf{H}\}$; the disjoint set, the non-basis set, of unphased amplitudes is $\{\mathbf{K}\}$; the phase problem is one of phasing $|U_{\mathbf{h}\in\mathbf{K}}|^{\text{obs}}$ from $U_{\mathbf{h}\in\mathbf{H}}$. At first sight, solving a crystal structure seems straightforward: a set of reflections is chosen to form the basis set. For a small molecule, for example, this will be the origin-defining ones that obey the required rules. These reflections are used as constraints in an entropy maximization to generate a map $q^{\text{ME}}(\mathbf{x})$. Such a calculation is a far cry from the Brandeis dice problem of §2.2, but there is a link *via* the use of the ME method in handling indeterminate problems where only certain expectation values are known; in this case, the expectation values are the basis set U 's – both phase and magnitude. Thus, $q^{\text{ME}}(\mathbf{x})$ satisfies the following conditions:

(i) It is optimally unbiased, *i.e.* it has maximum entropy.

(ii) Its Fourier transform reproduces the constraints to within experimental error.

(iii) The Fourier transform of $q^{\text{ME}}(\mathbf{x})$ generates estimates of amplitudes and phases for reflections in $\{\mathbf{K}\}$. This process is called *extrapolation*.

This is shown diagrammatically in Fig. 2.

At this stage, for a small basis set, the extrapolation is rather weak and the structure is certainly not visible in any map one cares to generate. So what can be done? The most obvious answer is to examine the most strongly extrapolated reflections and to use the phase information from them coupled with the observed U magnitudes and to pass these into the basis set. As we will see in §4.3, this usually leads to disaster: the ME solution gets trapped in a local maximum in phase space, whole subsets of reflections become wrongly phased and the structure remains unsolved. This is a manifestation of the *branching problem*: phases are being selected without exploring the relevant U space in sufficient detail so that what appears to be an unambiguous choice of new phase is, in reality, no such thing. So what can be done? Extra unphased reflections must be added to the starting set with permuted phases giving rise to a multisolution environment just as in conventional direct methods and these reflections should be those that are very weakly extrapolated so that their inclusion in a new basis set offers maximum surprise to the calculations. How then do you select the correct permutation set? Can you choose those that give maps of maximum entropy? The answer to this is no and it is discussed in detail in §4.10, but the Bricogne formalism has the answer.

4.3. The Bricogne formalism

Bricogne has outlined a methodology of solving crystal structures based on the ME or saddlepoint method coupled with likelihood evaluation, which directly addresses the branching problem. This he has developed in a series of papers (Bricogne, 1982, 1984, 1988a,b, 1991a,b,c, 1992, 1993). We will now examine this formalism, continuing to follow the arguments in Bricogne (1988b, 1991a).

What have we done when we have defined a basis set and performed the entropy maximization? We have explored $P(\mathbf{U})$ around the basis set phase subregion in U space. Why do we need maximum entropy? We only

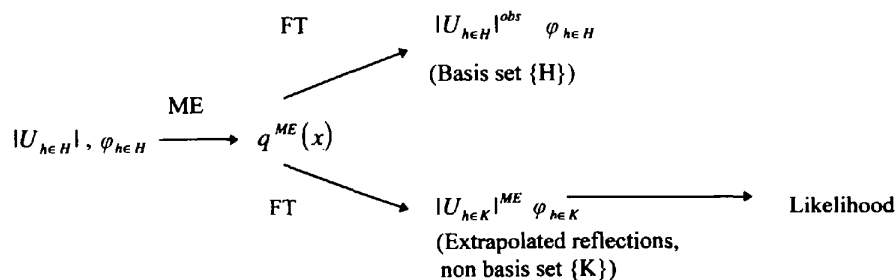


Fig. 2. Phasing *via* maximum entropy.

require $P(\mathbf{U})$ to reproduce those U magnitudes that are members of the basis set. It makes no direct reference to non-basis-set reflections and so it is highly indeterminate. The ME method is an obvious way of overcoming this: for such a phase choice maximizes the entropy of the distribution subject to the constraint that it must reproduce the basis set amplitudes and phase to generate a non-uniform distribution of atoms $q^{\text{ME}}(\mathbf{x})$. $P(\mathbf{U})$ is thus multimodal and this is the source of the branching problem.

Bricogne (1984, 1988*a,b*, 1991*a*) has shown a second way of getting to this same result without invoking the heuristic arguments of ME theory. This involves the saddlepoint approximation (Daniels, 1954; Reid, 1988), which is invoked as an alternative to the ME principle and shown to generate not only the same equations but also more accurate modifications of those predicted by the ME method alone (Bricogne, 1988*a*).

There is a remarkable benefit of the ME or saddlepoint method here: the approximations to $P(\mathbf{U})$ are now accurate for large U magnitudes and we have recentred the usual Edgeworth-series expansion for $P(\mathbf{U})$ away from $\mathbf{U} = 0$. In real space, this is equivalent to leaving behind the Wilson uniform distribution of atoms to a non-uniform one – one of the long-term goals of direct methods.

So far, all ME phasing methods perform in this way whether their authors realize it or not, but what can we do next? How do we stop ourselves locking into an incorrect solution? Let us accept that we cannot construct a tractable unique expression for $P(\mathbf{U})$ into which we could substitute the measured values of the moduli to get the conditional distribution of new phases, instead let us explore new subregions $P(\mathbf{U})$ where we test various new phase combinations to overcome the multimodality or branching problem. In practice, we take the current basis set and add to it a few reflections with large associated U magnitudes, which optimally enlarge the second neighbourhood of the basis set. The second neighbourhood is defined by reflections $\mathbf{h}_1 \pm \mathbf{R}_g \mathbf{h}_2$ for $\mathbf{h}_1, \mathbf{h}_2 \in \mathbf{H}$, where \mathbf{R}_g is the transpose of a rotation matrix obtained from the crystal space group. For acentric reflections, phase choices can be $\pm\pi/4, \pm 3\pi/4$, *i.e.* quadrant permutation; for centrics, both possible values are used, *e.g.* $0, \pi$ or $\pm\pi/2$. We now have several possible basis sets and we now explore $P(\mathbf{U})$ around each of these phase subregions by performing a constrained entropy maximization for each phase permutation and so derive associated phase and amplitude extrapolation. However, we have made no judgement as to which subregion is more likely to be the correct one. In the Bricogne formalism, this decision is made using likelihood. For every subregion, $P(\mathbf{U})$ is approximated to a multivariate Gaussian centred around the vector \mathbf{U}^{ME} corresponding to $q^{\text{ME}}(\mathbf{x})$. It is possible to express this in an analytical form. Several approximations exist

but we will concentrate here on the simplest – the diagonal form in which the extrapolates in \mathbf{U} are decoupled and treated as independent, *i.e.* the covariances between them are zero. This gives rise to readily tractable likelihood functions.

For each acentric extrapolated non basis set reflection \mathbf{k} , the likelihood measure, in its diagonal approximation, can be written (Bricogne, 1984, 1988*a,b*; Bricogne & Gilmore, 1990)

$$\begin{aligned} \Lambda_{\mathbf{k}} &= (|U_{\mathbf{k}}|^{\text{obs}} / \varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2) \\ &\times \exp\left\{-\frac{1}{2} \left[(|U_{\mathbf{k}}|^{\text{obs}})^2 + |U_{\mathbf{k}}^{\text{ME}}|^2 \right] / (\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2) \right\} \\ &\times I_0 \left[|U_{\mathbf{k}}|^{\text{obs}} |U_{\mathbf{k}}^{\text{ME}}| / (\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2) \right], \end{aligned} \quad (23)$$

where $\varepsilon_{\mathbf{h}}$ is the statistical weight of \mathbf{h} , $\sigma_{\mathbf{k}}^2$ the variance of $|U_{\mathbf{k}}|^{\text{obs}}$ and Σ a refinable measure of unit-cell contents, $\Sigma = 1/(2N)$ for N point atoms in the unit cell. The distribution (23) is a Rice distribution comprising a Gaussian (the exponential term) with an offset represented by the Bessel-function term (I_0). Note that this expression is a measure of agreement between $|U_{\mathbf{k}}|^{\text{obs}}$ and $|U_{\mathbf{k}}^{\text{ME}}|$; indeed, it has a maximum where $|U_{\mathbf{k}}|^{\text{obs}} = |U_{\mathbf{k}}^{\text{ME}}|$. For the centric case, the Bessel function is replaced by a cosh term:

$$\begin{aligned} \Lambda_{\mathbf{k}} &= [2|U_{\mathbf{k}}|^{\text{obs}} / \pi (2\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2)] \\ &\times \exp\left\{-\frac{1}{2} \left[(|U_{\mathbf{k}}|^{\text{obs}})^2 + |U_{\mathbf{k}}^{\text{ME}}|^2 \right] / (2\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2) \right\} \\ &\times \cosh \left[|U_{\mathbf{k}}|^{\text{obs}} |U_{\mathbf{k}}^{\text{ME}}| / (2\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2) \right]. \end{aligned} \quad (24)$$

In the spirit of traditional likelihood analysis, we can define a corresponding null hypothesis for the situation of null extrapolation, $|U_{\mathbf{k}}^{\text{ME}}| = 0$, which gives the Gaussian distribution of Wilson statistics. For acentric reflections,

$$\begin{aligned} \Lambda_{\mathbf{k}}^0 &= [|U_{\mathbf{k}}|^{\text{obs}} / (\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2)] \\ &\times \exp\left\{-\frac{1}{2} (|U_{\mathbf{k}}|^{\text{obs}})^2 / (\varepsilon_{\mathbf{k}} \Sigma + \sigma_{\mathbf{k}}^2) \right\}. \end{aligned} \quad (25)$$

The extension to centric reflections is obvious. Define

$$L_{\mathbf{k}} = \log(\Lambda_{\mathbf{k}} / \Lambda_{\mathbf{k}}^0). \quad (26)$$

Then the global log-likelihood gain (LLG) is

$$\text{LLG} = \sum_{\mathbf{k}} L_{\mathbf{k}}. \quad (27)$$

The LLG will be largest when the phase assumptions for the basis set lead to predictions of deviations from the Wilson distribution in the unphased reflections, and in this context it is used as a powerful figure of merit. However, rather than just choose those phase sets with high associated LLG, which is a somewhat subjective process, the Student t test is used (Shankland, Gilmore, Bricogne & Hashizume, 1993). The LLGs are analysed for phase indications using the t test, which defines the level of significance of the difference between two means. The simplest example involves the detection of the main effect associated with the sign of a single

centric phase. The LLG average, μ^+ , and its associated variance V^+ are computed for those sets in which the sign of this permuted phase under test is +. The calculation is then repeated for those sets in which the same sign is - to give the corresponding μ^- and variance V^- . The t statistic is then

$$t = |\mu^+ - \mu^-| / (V^+ + V^-)^{1/2}. \quad (28)$$

The use of the t test enables a sign choice to be derived with an associate significance level. This calculation is repeated for all the single-phase indications and is then extended to combinations of two and three phases. An extension to acentric phases is straightforward by employing two signs to define the phase quadrant both in permutation and in the subsequent analysis. In general, only relationships with associated significance levels <2% are used, but this is sometimes relaxed with sparse data sets.

Only those solutions that are consistent with the t -test results are kept further reduced if necessary to leave 8–16 remaining. New reflections are then permuted and a corresponding set of ME solutions is generated. In this way, we build a phasing tree in which each phase choice is represented as a node and has a score, or figure of merit, based on its log-likelihood gain (LLG). The root node of the tree is defined by the origin-defining reflections. The first set of phase permutations defines the second level. Those that do not pass the analysis of likelihood are discarded, then further phase permutations are used to generate the third level and this continues until a recognizable structure or structural fragment appears. Fig. 3 shows a simple outline of a three-level phasing tree. There is a nice example of this method in action using idealized one-dimensional data by Carter (1994). Note that likelihood is not used in this approach in the Bayesian way of multiplying the prior, instead it becomes a very powerful figure of merit which controls the branching problem by selecting the correct nodes. However, there is a Bayesian connection in the way in which the prior, $q^{\text{ME}}(\mathbf{x})$, is continually updated from level to level in the phasing tree by adding new phase information. We shall see later that the power of likelihood is such that it can be used to make other decisions concerning envelopes for macromolecules, parameters for non-crystallographic symmetry and to measure the effective number of atoms in the unit cell at low resolution.

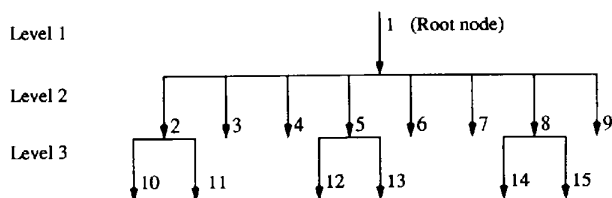


Fig. 3. A three-level phasing tree with 15 nodes.

Now, $q^{\text{ME}}(\mathbf{x})$ is a probability distribution and not a map in the traditional sense (although its peaks do correspond to atom positions), and thus it needs conversion to a more conventional one. The trial electron density maps generated in this approach are called centroid maps and can be visualized as Sim-filtered maximum entropy maps. For \mathbf{k} acentric, the Fourier coefficients are

$$|U_{\mathbf{k}}|^{\text{obs}} [I_1(X_{\mathbf{k}})/I_0(X_{\mathbf{k}})] \exp(i\varphi_{\mathbf{k}}^{\text{ME}}), \quad (29)$$

where

$$X_{\mathbf{k}} = (2N/\varepsilon_{\mathbf{k}}) |U_{\mathbf{k}}|^{\text{obs}} |U_{\mathbf{k}}^{\text{ME}}|. \quad (30)$$

For \mathbf{k} centric, these coefficients become

$$|U_{\mathbf{k}}|^{\text{obs}} \tanh(X_{\mathbf{k}}) \exp(i\varphi_{\mathbf{k}}^{\text{ME}}) \quad (31)$$

with

$$X_{\mathbf{k}} = (N/\varepsilon_{\mathbf{k}}) |U_{\mathbf{k}}|^{\text{obs}} |U_{\mathbf{k}}^{\text{ME}}|. \quad (32)$$

The practical differences between centroid and maximum entropy maps can be seen with reference to membrane data in Gilmore, Shankland & Fryer (1993).

One final practical point needs to be made here concerning the tightness with which one fits the constraints in the entropy maximization. This has been discussed briefly in §2.2 but can be very important in a phasing environment. If the fit between $|U_{\mathbf{k}}|^{\text{obs}}$ and $|U_{\mathbf{k}}^{\text{ME}}|$ is very slack, the ME extrapolation is weaker than it can be, and the phasing power and the discriminating power of likelihood is correspondingly reduced. If, however, there is overfitting, spurious details (often looking like small stones) appear in $q^{\text{ME}}(\mathbf{x})$, which give false phase indications. The latter situation can often be detected by the use of likelihood: if the LLG is monitored through the iterative cycles of entropy maximization, a maximum is often reached and this can be used as a place to stop. Alternatively, the χ^2 statistic can be used with a default choice of unity as a place to terminate. In practice, for small basis sets either method works well, but for large basis sets maximum LLG is preferred.

The *MICE* computer program is a practical implementation of part of the Bricogne formalism (Bricogne & Gilmore, 1990; Gilmore, Bricogne & Bannister, 1990; Gilmore, 1993), as is *BUSTER* (Bricogne, 1993). Surveys of the method and its practical applications can be found in Gilmore & Bricogne (1992), Gilmore (1993) and Gilmore, Shankland & Bricogne (1993). The main components of *MICE* are:

(i) A robust iterative entropy-maximization algorithm in which phased intensities, non-crystallographic symmetry and the contributions from known structural fragments correctly placed in the unit cell can be used as constraints. Envelopes can be used as non-uniform priors. The entropy maximization can stop at maximum LLG or at a given χ^2 .

(ii) A tree-directed search through a space of trial phase sets using entropy maximization for each phase permutation.

(iii) Estimation of LLG for each phase set and the use of the t test to prune unwanted nodes.

(iv) The generation of centroid maps and associated graphics.

From the very beginning it was recognized that the ME formalism was not subject to the same limitations as traditional direct methods and, in particular, the technique is stable irrespective of data resolution or the sampling of reciprocal space, and is also robust with respect to errors in the intensity data, which themselves could be modelled in the calculations. It also abandons the Wilson random-atom model and allows the inclusion of molecular envelopes and other constraints. Thus it should, at least in principle, be capable of tackling difficult problems in powder and electron diffraction as well as macromolecular crystallography.

4.4. Single-crystal X-ray diffraction

Normalization is the first stage of a structure solution when direct phasing is employed. To do this, we need expectation values for intensities. As we have already explained, this nearly always invokes Wilson statistics. Castleden & Fortier (1994) have examined a ME alternative. It is possible to use an exact series representation for the required expectation values but this poses a number of numerical problems and the ME alternative is shown to be a little less accurate but nonetheless a very workable alternative. Given the critical nature of normalization, more work should be done in this area.

The first application to a single-crystal structure solution came from Narayan, Nityananda & Vani (1983) using arguments based on preferred maps as discussed in §2.1. This is a pioneering piece of work that has been somewhat underestimated but employing a methodology that has not been used since. Their method was to refine initial random phases by maximizing the entropy with respect to the phase itself. After demonstrating its applicability in two dimensions, they used random phases to solve the structure of (*2E*, *6E*)-2,6-octadiene-1,8-dioic acid in space group *P1*, a 12-atom problem, and had some success with a 19-atom structure. Unfortunately, they were limited in accessible problems by available computer power. There is an interesting quotation near the end of their paper: '... we feel that the maximum entropy method is not just the direct methods in a new disguise, but is distinctly different. As such it merits further study in more complicated situations.' In fact, their method was unlikely to succeed on large problems because of an underlying assumption that entropy can be used as a measure of phase correctness.

Around the mid 1980's, there was a flood of optimism from ME specialists who were outside the

crystallographic community that the phase problem should be readily solvable by available ME algorithms. Their approach is typified by Livesey & Skilling (1985) and is again combinatorially based: an electron-density map is generated from a set of known phases and subjected to constrained entropy maximization. In the language of this method, this map is the preferred map. Where the initial phases were to come from was not discussed but a subsequent paper (Gull, Livesey & Sivia, 1987) produced an *ab initio* solution of a small centrosymmetric structure, $C_{10}H_{14}Cl_7O_4Sn$. They claim that the resulting electron-density maps are 'of a very high quality, comparable or even superior to the conventional maps calculated from the refined phases'. As discussed in the previous section, the method they are using is quite inapplicable to most real problems because it ignores, and indeed cannot process, the branching problem. Gull *et al.* ignored the problem altogether. They looked at the strongest extrapolates ($|U_{h\mathbf{e}K}|$) and add these to the basis set in a stepwise fashion, which locks the calculation into a local entropy maximum from which it cannot escape (Gilmore, Bricogne & Bannister, 1990). There is a detailed account of what happens when you do this by Sudo, Hashizume & Carvalho (1995) when attempting to solve Mg_3BN_3 from powder data in this way.

Harrison (1989) overcame the branching problem to a certain extent (although it is always present) by creating a large basis set using conventional direct methods, and these were extended and refined by a ME algorithm. Two structures were solved here: one a small molecule, thiolysine, which is readily accessible to conventional direct methods, and the other an unsolved 15-base oligomer of DNA having 610 atoms in the asymmetric unit and diffracting to only 3 Å. This is an impressive achievement but it does require initially reasonable phase estimates, which in this case come from direct methods and may often be difficult to obtain. It raises the question, however, as to why it has not been used subsequently as a general crystallographic technique.

The first *ab initio* applications of the Bricogne formalism were in the *MICE* program, where two small organic molecules (both readily solvable by direct methods) were solved (Gilmore, Bricogne & Bannister, 1990). It produced maps that were almost free of noise and that even resolved atom type in situations where oxygen was the heaviest atom, but it is no competition for conventional direct methods largely because of the computer time involved. However, when the data cease to satisfy the criteria of completeness and atomic resolution, the situation changes and this is the topic of the next three sections.

4.5. Powder diffraction

In powder diffraction, a three-dimensional diffraction pattern is collapsed into one dimension by spherical

averaging and, as a result, reflections that would otherwise be separately measured overlap. The degree of overlap increases with Bragg angle and unit-cell dimensions, thus often reducing the effective resolution of the data set to 1.5 Å or less – in many cases this is closer to 2 Å. For reviews of the problems of structure solution from powders, see Shirley (1984) and Giacovazzo (1993). There have, however, been many successes in solving crystal structures *ab initio* from such data, especially when heavy atoms are present. [See, for example, McCusker (1988) and Morris, Harrison, Nicol, Wilkinson & Cheetham (1992). The latter deals with a 29-atom problem, albeit with Ga and P atoms present.] Synchrotron radiation with its high intensity and a highly collimated incident beam can be of great assistance in resolving overlaps experimentally, but many samples exhibit a large intrinsic line broadening or non-accidental overlaps can be present, so there is still a need for a systematic approach to this problem that is capable of utilizing overlap information in an active way.

The ME approach to this problem is surprisingly quiet. Indeed, we can only refer to aspects of our own work. A suitable adaptation of the single-crystal approach has been formulated by Bricogne (1991c) and incorporated into the *MICE* program (Gilmore, K. Henderson & Bricogne, 1991; Shankland, Gilmore, Bricogne & Hashizume, 1993). In particular, *all* the intensity data are used in the normalization procedure and the log-likelihood gain is calculated using both overlapped and non-overlapped reflections. In addition, the latter can be included in the basis set as constraints for entropy maximization using the concept of hyperoctant phase permutation. The overlapped reflections can also be used in the final centroid maps so they play an active and often essential role in the entire phasing procedure. The methodology was initially applied successfully to an inorganic salt, KAIP_2O_7 , and a clathrasil, Sigma-2 (Gilmore, K. Henderson & Bricogne, 1991), then subsequently to unknown structures LiCF_3SO_3 (Tremayne, Lightfoot, Glidewell, Mehta, Bruce, Harris, Shankland, Gilmore & Bricogne, 1992) and formylurea (Tremayne, Lightfoot, Harris, Shankland, Gilmore, Bricogne & Bruce, 1992), an inorganic salt, Mg_3BN_3 (Shankland, Gilmore, Bricogne & Hashizume, 1993) and the organic molecule 1,3,4,6-tetrathiapentalene-2,5-dione (Lightfoot, Tremayne, Harris, Glidewell, Shankland, Gilmore & Bruce, 1993). The latter two structures use good-quality laboratory diffraction data. For a more detailed review, see Shankland (1996). Recently, in my own laboratory, we have solved five zeolite structures quite routinely using this method (Gilmore, unpublished results).

Finally, Takata *et al.* (1995) have studied the structure of the metallofullerene Y@C_{82} with synchrotron powder diffraction. The results confirm that the yttrium atom is indeed in the C_{82} cage, but only partially ordered, and displaced from the cage centre. To

confirm this, a ME map was generated using phased reflections from the model as constraints. This is not, of course, a model-free confirmation of the yttrium positions, since the phases were derived from a Rietveld refinement that included them but, nonetheless, the final maps were much easier to interpret using this technique. There is, we believe, a lot of mileage in using the ME method in conjunction with incomplete structural fragments.

4.6. Electron diffraction

Electron crystallography is a term used to describe crystal structure determination using electron diffraction intensities either alone or in conjunction with the Fourier transform of the corresponding high-resolution electron-microscope images. In terms of intensity, electrons are scattered by matter *ca* 10^4 times more efficiently than X-rays (it is rather like having your own synchrotron!) and the technique has the potential to give structural data from crystals whose maximum dimension is only 50 nm, but there are problems with dynamical scattering, diffraction incoherence, radiation damage, sample preparation and the sparseness of the data sets produced. This has led some authors to question the whole viability of *ab initio* structure determination using electron crystallography. Vainshtein's book is the classic work on the subject (Vainshtein, 1964), but it is now superseded by an excellent new book by Dorset (1995) that is essential reading for anyone interested in the subject. There are also recent reviews by Vainshtein, Zvyagin & Avilov (1992) and Fryer & Gilmore (1992). After being something of the Cinderella of structure-determination methods, the methodology is currently enjoying a surge of activity thanks to Dorset's pioneering work (see, for example, Dorset, 1991*a,b,c*, 1993; Dorset & McCourt, 1994) and research carried out by Fan and co-workers (see, for example, Fan, Zhong, Zheng & Li, 1985; Han, Fan & Li, 1986; Liu, Fan & Zheng, 1988). Dorset applies the techniques of traditional direct methods to the problems of *ab initio* structure determination, often using symbolic addition techniques or Sayre's equation. They prove surprisingly successful with this data, partly because the structures are small and partly because of the high degree of control of the process that his techniques provide. Fan, Zhong, Zheng & Li (1985) also employ traditional direct methods, including the use of Sayre's equation, which is found to exhibit the required stability (Liu, Fan & Zheng, 1988). However, the problem is ideal for the application of the ME methodology since, as we have seen, it is robust with respect to limited sampling of reciprocal space and to errors in the diffraction intensities.

Instead of proceeding initially from a basis set that defines the origin (and enantiomorph, if relevant), it is possible to use phases obtained from the Fourier

transform of a suitable image instead. These are usually of low resolution and the images can be difficult to obtain but, nonetheless, their transform can often provide very useful additional phase information. In the case of macromolecular electron diffraction, these phases are, in general, essential for any successful phasing.

4.6.1. *The electron diffraction of small molecules.* Again, the only ME applications in the literature come from our own research. The first is that of perchlorocoronene, $C_{24}Cl_{12}$, using only projection data down the c axis (Dong, Baird, Fryer, Gilmore, MacNicol, Bricogne, Smith, O'Keefe & Hovmöller, 1992). Experimental images were obtained at 3.2 \AA together with a diffraction pattern extending to 1.0 \AA . The Fourier transform of the image yielded four phases, which were used as a basis set. Routine phase extension using ME procedures in *MICE* produced a map in which all the diffraction intensities are phased and whose resolution exceeds 1 \AA . As a check, there is a single-crystal X-ray diffraction study in the literature (Baird, Gall, MacNicol, Mallinson & Mitchie, 1988) and the electron diffraction result, although completely independent of the X-ray structure, produced a solution that was identical to it, at least in projection.

Three-dimensional electron diffraction intensity data from diketopiperazine (2,5-piperazinedione, $C_4H_6N_2O_2$) were obtained from Vainshtein (1955). It is one of the most complete three-dimensional data sets ever collected from an organic crystal with 318 unique measured diffraction intensities. The space group is $P2_1/a$ with $Z = 2$. The electrostatic potential maps produced by Vainshtein were based on these intensities, but he used the phases from an X-ray crystallographic structure determination to reveal the complete structure. These results were controversial, since it is well known that both electron-density and electrostatic potential maps are most sensitive to phase and not amplitude errors. (Indeed, even maps computed from unit structure factors with the correct phases will, in general, reveal the crystal structure.) However, Dorset has solved the structure *ab initio* using symbolic addition and found the positions of all the non-H atoms without recourse to the X-ray result. The ME method also solved the structure without difficulty (Gilmore, Shankland & Bricogne, 1993; Gilmore & Nicholson, 1996). The solution from a phasing tree with the highest likelihood gave an electrostatic potential which was virtually noise free with all the non-H atoms clearly resolved. Fig. 4(a) shows the likelihood-preferred map.

A contrasting data set is provided by the electron diffraction data for $CuCl_2 \cdot 3Cu(OH)_2$ (Voronova & Vainshtein, 1958). This crystallizes in the symmorphic space group $P2/m$ with $Z = 1$. The Cu atoms lie at the unit-cell vertices so there is a predominance of zero phase angles, which makes it difficult to locate the

remaining Cu, Cl and O atoms. In addition, the sampling of reciprocal space is much less complete than for diketopiperazine with only 124 measured intensities available. The structure was originally solved using Patterson methods. With *MICE*, a simple calculation gave an electrostatic potential of remarkable clarity: all the non-H atoms are clearly visible despite the dominant features of the Cu atoms. Only the Cl atoms are rather weakly indicated (see Fig. 4b).

Voigt-Martin, Yan, Gilmore, Shankland & Bricogne (1994) have used *MICE* to solve the unknown structure of 4-[4-(*N,N*-dimethylamino)benzylidene]pyrazolidine-3,5-dione at 1.4 \AA in projection using 42 reflections. There is an excellent agreement between the solution and independent model-building and high-resolution electron-microscopy studies. In a very similar way, Voigt-Martin, Yan, Yakimansky, Schollmeyer, Gilmore & Bricogne (1995) have solved the structure of [9,9'-bianthryl]-10-carbonitrile using 150 unique diffraction intensities, and independently verified the result with model building and image simulation techniques.

We have also studied other structures with the same level of success, which leads us to concur with Dorset that the technique of *ab initio* structure solution using electron diffraction data has a general viability even

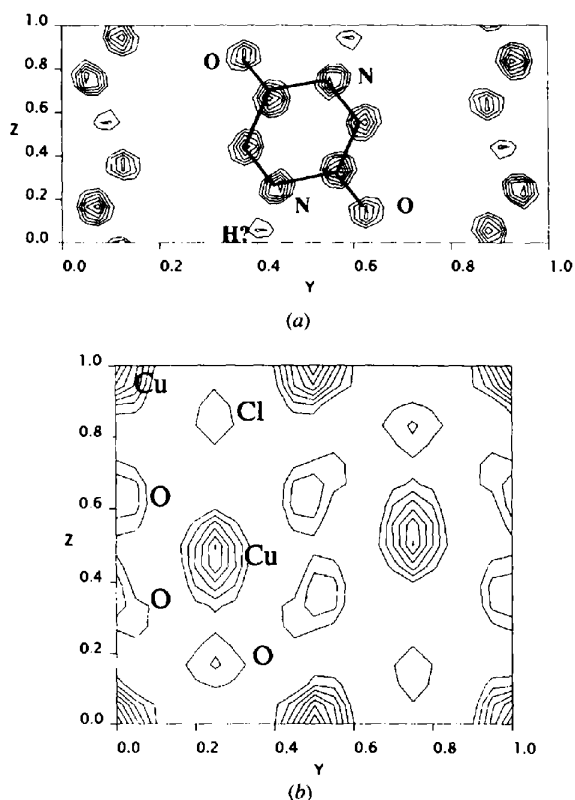


Fig. 4. Potential maps derived from the *MICE* computer program for (a) diketopiperazine, $C_4H_6N_2O_2$, and (b) $CuCl_2 \cdot 3Cu(OH)_2$. (From Gilmore, Shankland & Bricogne, 1993.)

when heavy atoms are present and offers a direct challenge to powder diffraction methods. We also quote Zandbergen & Cava (1995): 'We anticipate that within a few years quantitative electron diffraction ... will provide a quick way to determine crystal structures from small crystallites'. The method had come of age, and the maximum entropy method, properly formulated, has a major role to play.

4.6.2. *Electron microscopy – improving image data.* Our main interest in this review is diffraction. However, as we have already observed, an electron-microscope image can be a useful source of phase information via its Fourier transform. It follows that any image improvement will translate into a corresponding enhancement of the phase information. Furthermore, since much of the problem here is concerned with extracting a signal from a noisy background, the problem should be accessible to traditional ME methods. The literature, however, is surprisingly sparse.

Frieden (1988) has applied ME image reconstruction techniques to electron-microscopy images. He concludes: 'Anyone can produce a notable gain in resolution simply by inverse-filtering the image. The trick is to somehow avoid the excessive oscillations that ... naive approaches will produce ... Choices are, by their nature, personal, but I have not seen any other algorithms achieve the quality of output that these [ME and median window] produce.' He also considers that ME prefers point and line sources and gives some nice examples, including a deliberately misapplied ME image restoration. Non-periodic images are also studied by Farrow & Ottensmeyer (1988); the ME image-restoration method suppresses artefacts without removing high-frequency components in the image but the situation is rendered somewhat artificial by the use of model Gaussian noise in the trial data sets.

Anderson, Martin & Thomas (1989) exploit the periodic nature of the image (both Frieden and Farrow do not) using ME methods and avoid conventional Fourier filtering. They have applied it to a number of small-molecule films including poly{(benzo[1,2-*d*:4,5-*d'*]bisthiazol-2,6-diyl)-1,4-phenylene}. In all cases, the periodicities are clarified dramatically and show a considerable improvement over conventional Fourier filtering.

There is much exciting work to do in this field; advances could greatly increase the power of the ME method when used as a source of phase information for the *ab initio* solution of crystal structures whether small or large.

4.7. *Fibre diffraction*

Once again, the literature is sparse. This is a pity; one can view fibre diffraction as lying midway between powder and single-crystal diffraction in terms of the

information it gives. For a good review of the method, see Millane (1988). Given the successes of the ME formalism in other areas, it seems obvious that this could be a fruitful area for further development. The combination of entropy maximization and likelihood evaluation is clearly applicable here and Bricogne (1991c) has derived much of the necessary theory, but it has not been explored further in a practical way.

One outstanding achievement, however, must be described: the Pfl Inovirus calculations by Marvin, Bryan & Nave (1987), which follow an earlier report by Bryan, Bansal, Folkhard, Nave & Marvin (1983). They used an ME algorithm to calculate electron density distributions on the native virus data extending to 3 Å and on a single isomorphous derivative using iodine with a resolution of 5 Å. The initial structure of the virion was determined at low resolution by direct interpretation of the intensity distribution of the fibre diffraction coupled with model building. Phase extension was then carried out by ME methods. No phasing trees were built, so presumably the branching problem was considered unimportant. The authors recognize this, since there is an enantiomorph definition which takes place implicitly, and perhaps by accident, in the procedure. Nonetheless, the final results are convincing and, given the nature of the data, something of a triumph: no other statistical method could have processed this data. In addition, Gonzales, Nave & Marvin (1995) have used ME to a much lesser extent in a study of the Pfl filamentous bacteriophage fibre data at 3.3 Å resolution. They calculated a ME map using the native data and a model derived from constrained simulated annealing, then used this with iodine-derivative data to generate another ME map, which revealed the iodine positions. Given these successes, we must therefore ask why ME methods are not more used in fibre diffraction?

4.8. *Protein X-ray crystallography*

There has been much more activity here than with fibres, which is a reflection of the global activity of the field and the huge potential gains to be made by developing new phasing techniques for macromolecules. Indeed, some of the earliest applications of the ME method were made in the area of protein crystallography. For a review of recent activity in protein phasing methods, see Gilmore (1992).

Collins's early work (Collins, 1982) studied phase extension from 2 Å MIR phases to 1.5 Å for rubredoxin A, a small protein. Two years later, Bricogne (1984) demonstrated phase extrapolation in the small protein crambin using experimental data coupled with correct phases starting at 3 Å and extending to 1.5 Å. Both these instances are problems that cannot be tackled with conventional direct methods and both authors demonstrated the process was stable at these resolutions. It was

already clear in this early stage of its development that the methodology was of some potential importance.

There are two surveys by Bryan (1988, 1989) that discuss possible strategies for SIR, unknown heavy-atom positions and phase extension in macromolecular crystallography. These papers give a good account of the problem coupled with applications mostly using model data. Most interesting, however, is an attempt with a small molecule to incorporate structural information into the phasing at a much earlier stage than usual, using triple correlations of atomic coordinates. The whole question of incorporation of such prior knowledge is of fundamental importance – finding a practical way of doing this would be a major and possibly essential advance in the techniques of solving large structures *ab initio*. The SIR problem is further discussed by Bryan & Banner (1987) but they use entropy as a criterion of phase choice (see §4.10), which must limit its applicability.

As we discussed in §2, one of the benefits of ME is its ability to incorporate many different constraints, not just those of amplitude and phase, and this is vital in macromolecular crystallography where the most obvious additional constraint is that of the envelope or mask in conjunction with solvent flattening. Prince, Sjölin & Alenljung (1988) described an algorithm for doing this with a test on native ribonuclease A and an application to two previously unknown structures, bovine prothrombin, in which initial phases were obtained from MIR data, and fragment TR2C from bull testis calodulin. There is a more detailed description of the latter in Sjölin, Svensson, Prince & Sundell (1990). In the latter case, there are two molecules in the asymmetric unit; the first was found by molecular replacement and the second located using the ME-envelope solvent-flattening combination. *MICE* has also been adapted by Xiang, Carter, Bricogne & Gilmore (1993) to use an envelope and solvent flattening to bring about substantial model-free map improvement at 3.1 Å resolution for cytidine deaminase. Fig. 5 is a graphic demonstration of the improvement in the mean phase error that was achieved when compared with conventional methods.

There are major differences in these two approaches: Prince and co-workers maximized the entropy of the maps after conventional solvent flattening by fitting the entire data sets exactly (*i.e.* giving a reduced $\chi^2 = 0.0$). Carter *et al.*, however, included the envelope as a prior in the entropy maximization from the very beginning, as well as monitoring the behaviour of the LLG as the entropy maximization proceeded, stopping when it reached a maximum, and hence optimum, value. This prevents the building up of spurious detail in the maps that will inevitably occur by overfitting the basis set data when they are used as constraints. Carter's conclusion is that the use of ME in this way is 'a superior way to

improve isomorphous replacement electron-density maps before model building.'

There is a growing body of evidence to support this. Recently, for example, Laphorn *et al.* (1994) used *MICE* in a non-tree-building mode, but with the inclusion of the envelope constraint, to improve maps for the hCG structure; the improvement made a significant contribution to the modelling of some of the loops in the protein. The fact that this is a model-free method is important: it minimizes the accumulation of error in the whole phasing process.

All the above have assumed that there is no significant branching problem, and therefore no need to carry out any phase permutation. Carter and co-workers have examined a problem where this is not the case and used an adapted version of *MICE* to solve the previously unknown TrpRS structure (Doublé, Xiang, Gilmore, Bricogne & Carter, 1994; Doublé, Bricogne, Gilmore & Carter, 1995). In this case, there were serious problems with a lack of isomorphism, and a poorly known molecular envelope exacerbating all the usual problems. These were overcome in a novel way:

(i) Strong reflections for which there were only weak or contradictory phase indications were given permuted phases.

(ii) Different envelopes were tried and analysis of the LLG used to decide on the correct mask.

(iii) In a structure of this size, phase permutation can lead to huge phasing trees and a corresponding computational explosion. To overcome this, incomplete factorial designs (i.f.d.'s) were used [see Carter & Yin (1994) for a recent discussion of i.f.d.'s in the context of crystallization], which greatly reduce the number of nodes needed for a given permutation experiment but still allow the selection of the correct phases by an appropriate use of significance testing. The result was a successful structure solution of an unknown protein and probably the most significant demonstration of the power of the formalism to date.

Another benefit of the ME method is the relatively easy way in which other constraints such as non-crystallographic symmetry (n.c.s.), partial structures, isomorphous structures *etc.* can be built into the calculations (Bricogne, 1988*a,b*, 1992); some applications of n.c.s. are described in §4.9.

Attempts to solve protein structures *ab initio* take us into a whole new world, but before we look at the few results that there are, we should make it clear what *ab initio* means in this context, since some authors have shown a tendency to re-define the problem into one that is much more tractable. *Ab initio* macromolecular structure determination is precisely what it is in the small-molecule case – the solution of a crystal structure from its native data with no prior knowledge other than the chemical formula. Thus, structures solved using phases derived from SIR or MIR experiments are not true *ab initio* structure solutions.

The most ambitious result here comes from Sjölin, Prince, Svensson & Gilliland (1991), who claim to have solved the 323-residue protein bovine chymosin, previously solved by standard techniques. They employed a fractional factorial design to permute phases (a method not unlike the use of codes or incomplete factorial designs, but less efficient in its covering of phase space). Centric phases were used initially and the phase indications were extracted using Yates's algorithm with entropy employed as a measure of phase correctness and to control the branching. Several layers of a phasing tree were carried out in this way and the resulting maps look very convincing, but here are some questions concerning this approach:

(i) Entropy cannot be used as an indicator of phase correctness (see §4.10).

(ii) The test of significance quoted in this paper (Table 1 of Sjölin *et al.*, 1991) do not give any phase

indications with a significance level lower than 20%, and of these over half are incorrectly indicated (Carter, unpublished results; Bricogne, unpublished results).

The use of designed experiments, *i.e.* the fractional factorial design, is a significant advance, however, and this was the first published application of such a method in a ME environment.

Avian pancreatic polypeptide (App) is a small protein (Glover, Haneef, Pitts, Wood, Moss, Tickle & Blundell, 1983) made somewhat unusual by the availability of good 1 Å data. It is used as a test structure by direct methods practitioners interested in developing techniques of phasing protein structures. *MICE* has been used to identify best phase sets for App from those generated by the *SAYTAN* direct methods computer program working at only 2 Å resolution (Gilmore, A. N. Henderson & Bricogne, 1991). In this study, sets of phases from *SAYTAN* derived from

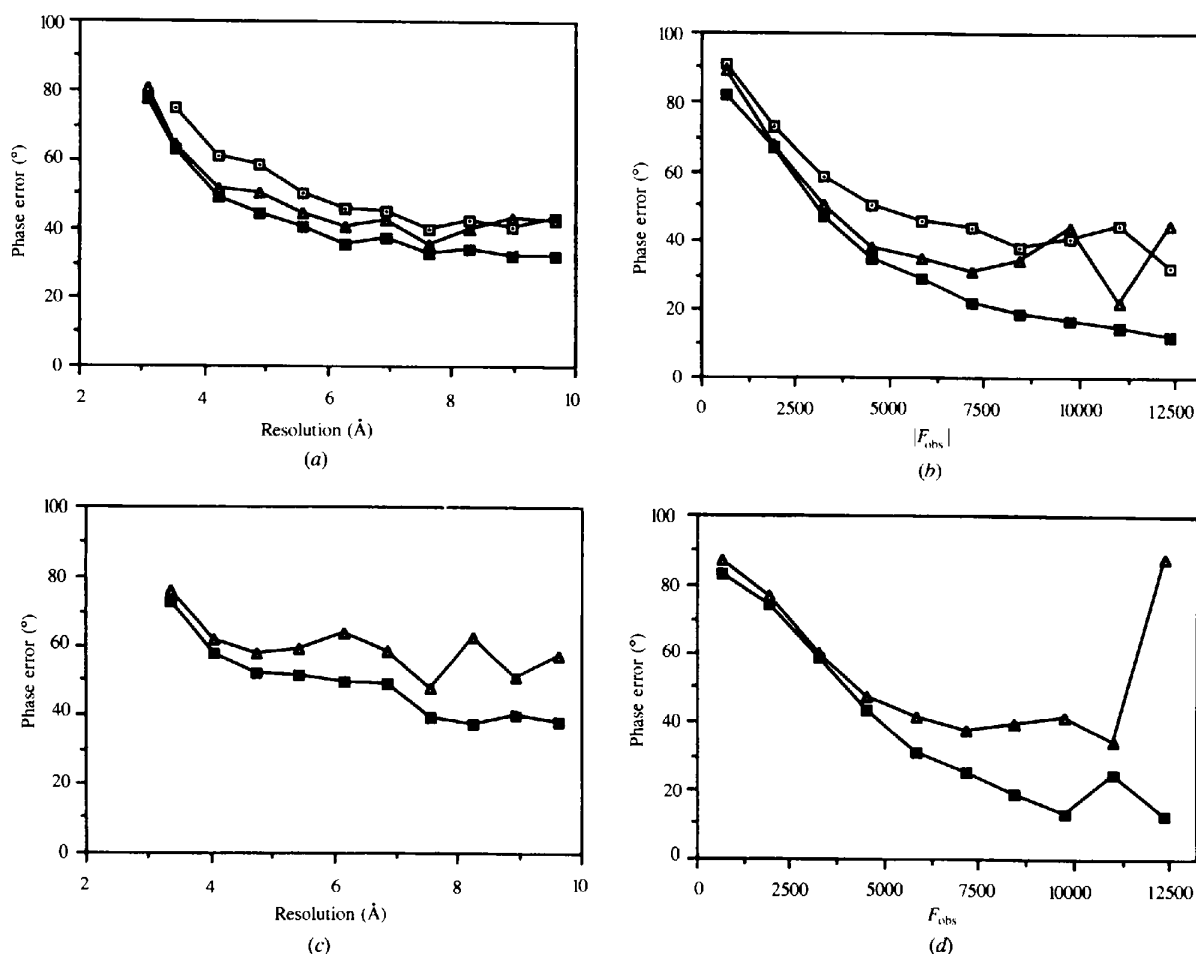


Fig. 5. Mean-phase-error distributions for MIRAS, conventionally flattened and centroid. (a) Overall mean phase error as a function of resolution. (b) Overall mean phase error as a function of $|F_{obs}|$, on an arbitrary scale. (c) Mean phase errors for extrapolated reflections (reflections in $\{K\}$) as a function of resolution. (d) Mean phase errors for extrapolated reflections as a function of $|F_{obs}|$. Symbols: \square MIRAS phase sets; \blacksquare centroid phase set from maximum entropy solvent flattening; \triangle solvent flattened phase set. (From Xiang, Carter, Bricogne & Gilmore, 1993.)

1 Å data could not be ranked by any conventional figure of merit, but they were passed to *MICE* using only the 2 Å data. Each phase set was subjected to constrained entropy maximization coupled with likelihood evaluation; the sets with minimum phase error (around 45°) were clearly indicated by the LLG estimate. No other figure of merit could work so well with this level of complexity at this resolution, confirming in a crystallographic environment a famous theorem by Neyman & Pearson (1933*a,b*) (see also Edwards, 1972, ch. 9) that establishes likelihood as an optimum criterion in statistical inference more powerful than any other.

App contains one Zn atom per asymmetric unit, the position of which is not particularly difficult to locate at 1 Å, but Gilmore & Nicholson (1996), using only the 3 Å data, have shown the Zn atoms can be readily located to within 0.5 Å of their refined positions *ab initio* using a combination of entropy maximization, likelihood and coding. For a structure of this complexity, there is a large increase in the number of reflections that need to be phased to produce an interpretable map; this, in turn, leads to a proportional increase in the size of the phasing tree to the point where it can become computationally unrealistic. However, Bricogne (1993) has shown that this problem can be overcome by the use of error-correcting codes. It is not possible or appropriate to discuss this vast subject here. Hill's book (Hill, 1991) provides a simple introduction and two classic but more advanced volumes are those by MacWilliams & Sloane (1977) and Conway & Sloane (1993). Their use enabled the Zn atoms to be located *ab initio* at 3 Å in a routine way, involving a three-level phasing tree with over 8000 nodes, although the calculations took less than 24 h on a network of modest Unix workstations. The use of codes is likely to become a key advance in the use of ME in protein phasing.

Finally, Bricogne has published some interesting test results in an attempt to solve crambin *ab initio* at 1.5 Å resolution using the *BUSTER* program (Bricogne, 1993). Both Roth (1991) and Schluenzen, Volkmann, Thygesen, Hansen, Harms, Bennett & Yonath (1994) have reported work on a large ribosomal subunit from *Thermus thermophilus* at low resolution using *MICE*.

4.9. Protein electron crystallography

Electron diffraction techniques can also be applied to biological macromolecules. In this case, an *ab initio* phase determination is not practical because of the structural complexity of the molecules. However, instead of proceeding from a basis set that defines only the origin and enantiomorph, the phases of the basis set reflections can be obtained from the Fourier transform of a suitable image. Such phase information is usually of relatively low resolution and the practical problems of obtaining such image data for macromolecules are considerable. The electron diffraction

data are, however, more easily obtained and have a much higher resolution than the images. The problem then arises as to how to phase the diffraction data starting from the lower-resolution image transform phases when processing macromolecule data. The ME method is ideally suited to this problem but, again, the only published work in this field is our own.

We have presented trial applications to the two-dimensional purple membrane data from *Halobacterium halobium* (Henderson, Baldwin, Downing, Lepault & Zemlin, 1986; Baldwin, Henderson, Beckman & Zemlin, 1988), showing that it is possible to produce good phase extrapolation starting with basis sets having a resolution as low as 15 Å (Gilmore, Shankland & Fryer, 1992, 1993).

Cholera toxin is the protein responsible for the clinical symptoms of cholera. It consists of five *B* subunits, each of which has molecular weight 10 600 Da, and one *A* subunit composed of two polypeptides *A*₁ and *A*₂ of molecular weight 23 500 and 5 500 Da, respectively. The five *B* subunits are arranged in a pentameric ring around the central *A* core, exhibiting fivefold non-crystallographic symmetry (Brisson & Moser, 1991). We have been working with a two-dimensional data set from a sample crystallized in two dimensions on lipid layers in which 56 unique image-derived phases were available at 8.8 Å resolution plus 1417 diffraction intensities extending to 4 Å. The problem is to phase the 4 Å data from the 56 known phases imposing fivefold non-crystallographic symmetry (n.c.s.) and solvent flattening, while working wholly with projection data.

Hitherto, we have been using likelihood as an indicator of phase choices and envelope definition, but it is of course a general technique of statistical inference. Accordingly, in the cholera toxin case we have successfully tested the likelihood criterion as an accurate and reliable predictor of: (i) the effective number of atoms in the unit cell (at 4 Å this is not equal to the number of atoms in the cell); (ii) the centroid coordinates for the fivefold non-crystallographic axes; (iii) the envelope radius; in the presence of known phase information from the image data. We have used these results to carry out phasing extension on cholera toxin. The first results are very encouraging (Gilmore & Nicholson, 1996).

4.10. Is entropy a viable indicator of phase correctness?

The short answer is no, but further amplification is obviously needed:

(i) Naively, there is no reason why entropy should be such an indicator. Entropy can be considered a measure of map flatness; there is no reason why the correct map among many possible maps each with different phases but the same amplitudes should be the flattest.

(ii) Lemaréchal & Navaza (1991), in a critique of Sjölin, Prince, Svensson & Gilliland (1991), give arguments concerning the phases assigned to the Lagrangian multipliers used in the entropy maximization (see §7 for further details of the use of Lagrangian multipliers) to demonstrate that entropy cannot be used as a reliable figure of merit. There is an additional discussion by Navaza (1991).

(iii) Gilmore, Bricogne & Bannister (1990), Gilmore, K. Henderson & Bricogne (1991), Gilmore, A. N. Henderson & Bricogne (1991), Shankland, Gilmore, Bricogne & Hashizume (1993) and Sudo, Hashizume & Carvalho (1995) all give examples where the use of entropy in this way is disastrous. McLachlan (1993) also cites examples using data from the membrane protein bacteriorhodopsin.

(iv) Sato (1992), however, disagrees and claims that entropy is a valid indicator for data sets of resolution greater than 1.5 Å, but in this case he is working with very large basis sets of accurate phases extracted from a refined small-molecule structure, C₁₀H₁₃N₅O₃.

Is it a question of resolution: can entropy work as a figure of merit at low resolution where one expects maps free of large peaks? Entropy is a measure of map flatness, so should perhaps be able to select the 'best' ones. The trouble here is that density builds quite quickly even at low resolution. To demonstrate this, Table 1 presents a simple one-level phasing tree for halorhodopsin using electron diffraction data from Havelka, Henderson, Heymann & Oesterhelt (1993). At 15 Å, it can be seen that LLG is an excellent indicator of phase correctness whereas entropy is simply incorrect.

Likelihood, in our experience is the optimum figure of merit to use in this environment. Entropy can be included by using $NS+LLG$ (S is the entropy), but there are problems here with the definition of N , which cannot be simply equated with the number of atoms in the unit cell. There is a summary of a discussion on this matter at a conference on Direct Methods of Phasing in Macromolecular Crystallography in Florida, USA, in 1992 by Sayre (1993), which defines the controversies on these points quite succinctly.

4.11. The phase problem: some conclusions

We believe the ME and Bayesian methods have the power and potential to transform whole areas of the phase problem in crystallography. Properly used, they provide a wholly consistent statistical formalism for dealing with all aspects of phasing. Practical progress may seem very slow but few people are working in this area and there is a huge body of well established theory still waiting to be programmed and tested. Remember how long it took

Table 1. *Is entropy a good figure of merit at low (15 Å) resolution?*

A simple 1-level phasing tree for halorhodopsin data using a 15 Å basis set. It can be seen that the LLG is the optimum indicator of phase error in the basis set, and that the maximum entropy solutions have high mean phase errors.

Node no.	LLG	Entropy	Mean phase error (°)
1	0.787	-0.027	18.8
2	-0.095	-0.023	69.3
3	0.000	-0.026	58.5
4	0.472	-0.023	127.3
5	-0.034	-0.025	103.9
6	0.050	-0.025	99.5
7	0.076	-0.025	88.7
8	-0.014	-0.024	139.2
9	-0.042	-0.026	43.6
10	0.061	-0.024	94.1
11	0.065	-0.025	83.2
12	-0.001	-0.024	133.8
13	0.775	-0.027	73.8
14	-0.087	-0.024	124.3
15	-0.012	-0.026	113.4
16	0.494	-0.023	164.0
17	0.810	-0.027	0.0
18	-0.035	-0.023	50.5
19	-0.066	-0.026	39.7
20	0.501	-0.023	90.2
21	0.058	-0.026	30.2
22	0.080	-0.024	80.7
23	0.043	-0.025	69.9
24	-0.063	-0.024	120.4
25	0.050	-0.025	24.8
26	0.092	-0.024	75.3
27	0.033	-0.025	64.5
28	0.033	-0.025	115.0
29	0.796	-0.027	55.0
30	-0.026	-0.023	105.5
31	-0.079	-0.026	94.7
32	0.513	-0.023	145.2

direct methods to become established from the initial theory papers. In the past year, applications to unknown protein structures have begun to demonstrate significant success, which should inspire confidence and further development. A major advance would be the ability to incorporate known structural information into the calculations from the very beginning. The Bayesian formalism is ideal for this, but how can it be done? There is a growing body of feeling that this is a necessary minimal prerequisite for the *ab initio* solution of even relatively small protein structures - solving larger structures in this way is, as yet, an unattainable goal.

In macromolecular electron microscopy, we pose a question: Given the problems of obtaining high-resolution phase information from image data, would it not be better to concentrate instead on lower-resolution images, then use the high-resolution diffraction intensities, phasing them *via* ME phase-extension procedures?

5. Accurate charge and spin densities

5.1. Single-crystal diffraction data

We are concerned here with the determination of accurate charge densities. For a review of current conventional techniques in the field, see Koritsansky (1993). Given a solved structure with high-quality data techniques such as high-order refinement, the aspherical atom formalism, static deformation and deformation densities are used to obtain bonding densities. It is an obvious extension of the previous sections of this review to ask if maximum entropy or Bayesian methods can be of value here given the inherent lack of bias in the ME maps, the imposition of positivity of density, smoothing and the ability to extend resolution beyond that of the measured data.

Sakata & Sato (1990) have studied the charge density in Si in this way, using a full set of phased data as constraints in an entropy maximization, with careful determination of the data scaling factor using the *Pendellösung* method (Saka & Kato, 1986). Following Collins methodology (Collins, 1982), they computed a ME map including all 30 phased reflections but excluding the forbidden reflections, which in Si arise from the asphericity of the electron density, which is itself a consequence of bonding electron density and anharmonic thermal motion of the Si atoms. They also carried out a calculation in which one reflection was used to define the origin, as required by the space group, and the remaining unphased reflections were included as constraints using only their amplitudes. This is, of course, a model-free reconstruction of the electron density. In both cases, the final maps were deemed identical and showed electron density in the expected bonding regions. They conclude that the ME method: 'will be very useful in accurate structure analysis, because it is a model free analysis, it can yield very high resolution on a density map provided that data are sufficiently accurate, and it may work without phase information.' However, it is only 'model free' in the simplest of cases; for most structures, the phase problem will need to be solved first. Schotte, Schotte, Bleif & Papoular (1995) have examined KOH, NaOH and NaOD *via* neutron diffraction looking at the distribution of density around the H atoms, and conclude that a naive approach to using ME in this way is unsatisfactory, and that a suitable non-uniform prior is required. This seems to the author to be a logical requirement of the use of entropy in this field, and is discussed further in §5.4 with reference to spin densities. Kumazawa, Kubota, Takata & Sakata (1993) have produced a computer program, *MEED*, for electron density calculations that follows the formalism described by Saka & Kato.

Jauch & Palmer (1993) have cast grave doubts about this whole approach. They have applied the ME method to accurate γ -ray diffraction data for MnF₂

and NiF₂ as well as the Si *Pendellösung* data of Saka & Kato. Their conclusions may be summarized as follows:

(i) A constraint function based on χ^2 can be very misleading. In MnF₂, 45% of the constraint function arose from one low-order reflection, so there is a highly non-random contribution from individual reflections in the entropy maximization.

(ii) ME emphasizes any sharp features.

(iii) Smooth features in low-density regions are contaminated with spurious artefacts.

They conclude: 'Obviously a ME analysis does not provide a suitable alternative to least squares refinements relying on elaborate physical models.'

Jausch (1994) has pursued this matter further, pointing out, quite correctly, that ME is being used not as a tool of statistics (where it surely belongs) but as a map smoothing, or regularizing, function. The smoothing is, of course, always present in the method but is only a by-product and not the basic rationale. One consequence of this is that ME maps are susceptible to exhibiting spurious detail similar to those inherent in Fourier inversion except that positivity is ensured. There is also a further criticism by de Vries, Briels & Feil (1994). They note that the ME property of producing the flattest electron density consistent with the data causes the calculated values of the strong reflections to deviate systematically as much as possible from their measured values. They also point out that the method does deal with series termination errors quite well and they introduce a weighting scheme that minimizes artefact production but, nonetheless, as they clearly demonstrate, spurious detail is still present and peaks at atomic centres are much too sharp.

From both our experience with ME maps and the same theoretical considerations, we agree with all these reservations. We would also emphasize a further point: all the reported calculations in this field use a uniform prior. This allows any unfitted electron density to be distributed as uniformly as possible in the unit cell by the ME calculations. We know that, physically, this is not true – the unfitted density should accumulate in the relevant bonding regions. At the very least, a non-uniform prior that will do this is required, but this cannot be properly defined without knowledge of the answer. However, the ability of the ME method to carry out limited super-resolution, the positivity and smoothness it imposes and its inherent lack of bias if correctly used make it an obvious tool for electron density studies. What is needed are some rigorous studies and comparisons with accurate multipole refinement, with the realization that the ME formalism is no substitute for accurate measurements and detailed careful refinements, nor can it compensate in full for series termination effects: it de-emphasises but does not remove them.

5.2. Powder diffraction data

The problem becomes even more acute when powder data are considered. Sakata, Uno, Takata & Mori (1992) have applied the ME method to rutile (TiO_2) using accurate long-scan powder diffraction data. This structure had previously been studied by single-crystal deformation-density methods by Restori, Schwarzenbach & Schneider (1987). One of the attractions of using powders in this context is that extinction ceases to be a serious problem. Sakata *et al.* report bonding density in the Ti—O bonds but none in the Ti—Ti bonds, which is in partial agreement with single-crystal studies. Kumazawa, Takata & Sakata (1995*a,b*) have further studied thermal motion in rutile starting with a model obtained from neutron powder diffraction. From the nuclear density distribution, they obtained an anharmonic thermal model parameterized to fourth order. The starting model here comes from a maximum entropy nuclear density map derived from the experimental data.

Nakahigashi & Higashimine (1995) have studied α -tin using two overlapped sets and 11 non-overlapped reflections and claim that the bonding density between the adjacent Sn atoms in the (110) plane can be clearly seen. However, if the single-crystal ME method is open to criticism, the problem must be exacerbated in the powder case, where data by their very nature are likely to be inferior, with the usual problem of peak-profile fitting and overlaps, although for very simple situations this may not be too important.

5.3. Surface densities

If powder diffraction is controversial in the context of electron densities, then surface diffraction poses even greater problems: the data are difficult to obtain, are incomplete to avoid reflections where the bulk solid contributes and there may well be errors in the model. Not deterred, Carvalho, Hashizume, Stevenson & Robinson (1966) have produced electron density maps for the Si(111) 7×7 surface using both X-ray and electron diffraction. They find no electron density between the atoms because of problems with data quality. It would be a major advance to see surface densities and one can therefore expect more research in this field to appear, perhaps for simpler surfaces.

5.4. Spin densities

Some significant and very interesting work has been carried out in the area of spin densities by Papoular and co-workers (Papoular & Gillon, 1990; Papoular & Delapalme, 1994; Papoular, Zheludev, Ressouche & Schweizer, 1995). In this method, one uses polarized neutrons to derive ME reconstructions of spin densities. Experimentally, a periodic spin density is induced in a single-crystal sample by applying a strong magnetic

field via a cryomagnet at low temperature. Experimental problems arise from the use of a cryomagnet in which the applied field is in the vertical (z) direction. Since the diffracted-beam detector is also in the horizontal plane, it may be lifted up from the plane, but only by a small amount because of the equipment construction and design. The net effect is that reciprocal space is sampled only in a few layers; thus, if the crystal is mounted with the c axis vertical then only $hk0$, $hk1$, $hk2$ and $hk3$ flipping ratios may be measured. Resolution along the h and k directions is not so restricted. To reconstruct spin densities to high resolution in any projection is then the problem. It is not unlike the sampling problem in electron crystallography using tilted samples.

The ME method is very well suited to such tomography problems. Recently, Papoular *et al.* (1995) have used ME as a regularizing function to produce some impressive results. This is demonstrated with the PNN crystal – a phenyl-substituted α -nitroso-nitroxide – which is shown in Fig. 6(*a*). The compound is paramagnetic owing to an unpaired electron, which is in a π^* molecular orbital. To find out how the spin density is distributed in the molecule, a single-crystal polarized-neutron diffraction experiment was per-

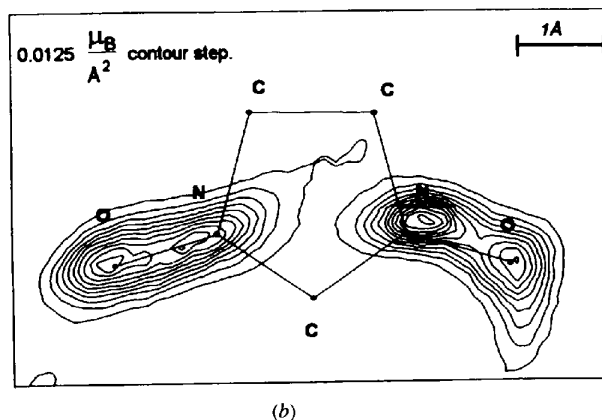
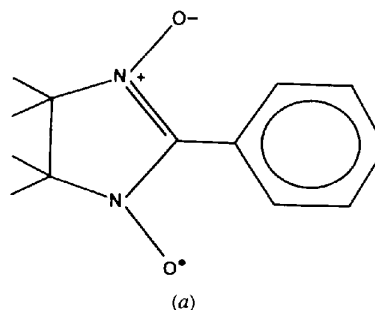


Fig. 6. (*a*) The structural formula of PNN. (*b*) Spin density in a PNN crystal projected orthogonally onto the O—N—C—N—O molecular plane reconstructed by 3D ME from 187 Fourier coefficients. $0.0125 \mu_B \text{ \AA}^{-2}$ contours. (From Papoular, Zheludev, Ressouche & Schweizer, 1995.)

formed. The three-dimensional ME reconstruction is shown in Fig. 6(b). It is highly convincing, although a conventional map is not shown for comparison. Zheludev, Papoular, Ressouche & Schweizer (1995) have extended their method significantly by the use of non-uniform priors. In an application to [TCNE]⁻[Bu₄N]⁺, the use of an appropriate prior gave impressive results. The problem of spin densities is less difficult and contentious than the prediction of accurate electron densities because there is less susceptibility towards producing spurious artefacts, and we are not requiring super-resolution in more than one direction; the ME method should therefore become the technique of choice, especially when non-uniform priors are used.

6. Particle sizes from diffraction experiments

6.1. Small-angle scattering

Small-angle scattering (SAS) techniques are subdivided into small-angle X-ray scattering (SAXS) and the neutron equivalent (SANS). They provide a direct way of obtaining particle size information from monodisperse and polydisperse solutions. For a brief survey of the method, see Glatter (1992), and for a detailed monograph see Feigin & Svergun (1987). In a SAXS experiment, a set of M scattering intensities, I_j , are measured as a function of increasing incident X-ray angle, θ . The experiment operates at very low θ values. From such an experiment, we obtain a distance distribution function $p(r)$ via the relationship

$$I(q) = \int_0^r p(r)S(q, r) dr, \quad (33)$$

where $q = 4\pi \sin \theta / \lambda$, θ is the incident angle, λ is the radiation wavelength and $S(q, r)$ depends on the assumed particle shape. For example, for a sphere of radius r ,

$$S(q, r) = N(\rho - \rho_s)[3(\sin qr - qr \cos qr)]^2 / (qr)^6, \quad (34)$$

where ρ and ρ_s are the scattering-length densities of the particles and solvent, respectively, and N is the number of particles per unit volume.

In practice, $p(r)$ may not be uniquely determined because: (i) the observations are subject to experimental error; there are random errors generated from counting statistics, and smearing arising from instrument resolution; (ii) experimentally, we only sample M points of $I(q)$:

$$I_j(q) = \int_0^r p(r)S_j(qr) dr, \quad j = 1, 2, \dots, M, \quad (35)$$

and there is a limit on how fine this sampling can be, imposed by the experimental constraints in (i). Standard techniques to extract a feasible solution for $p(r)$ use data smoothing, interpolation and model functions, all of

which have the potential of introducing errors and biases into the solution, so SAXS is an obvious candidate for ME methods. In SANS, there is negative contrast and, in consequence, the associated probability distributions measured the probability of finding density lower than the solvent, but the methodology is similar.

The literature is not extensive but it is all interesting, although the approaches are not dissimilar, differing only in detail. Potton, Daniell & Rainford (1988a) looked at particle size distributions from SANS data via ME using a uniform prior and developed an algorithm that successively generates trial structures and subjects them to entropy maximization using the Skilling algorithm (see §7.3), which is well suited to this purpose (Skilling & Bryan, 1984). The fit between observed and calculated intensities is measured by the traditional χ^2 statistic, thus maintaining a suitable lack of fit that incorporates experimental errors. This paper deals only with model data in which experimental noise is modelled via random errors having a suitable Gaussian distribution and so, although the conclusions are encouraging, they are by no means definitive. However, it is always sound to start a new approach to problems of this sort with simulated data, and, in a subsequent paper (Potton, Daniell & Rainford, 1988b), the authors study the annealing of stainless steel at different temperatures, again using SANS technology. The results correlate well with those from traditional data-processing methods, throwing up no unexpected surprises.

Hansen & Pedersen (1991) examine the ME method in SAXS comparing its efficacy against two conventional approaches based on indirect Fourier transform methods that use least-squares and various interpolation procedures, and employing both real and simulated data. The conclusions are those that one would expect: there is a lack of spurious artifacts and oscillations, and a reduced sensitivity to prior assumptions concerning particle size and shape when the ME method is employed. Scattering profiles are also sharpened. Morrison, Corcoran & Lewis (1992) give a very detailed account of the ME method, again using the Skilling MEMSYS package, on simulated and real data looking especially at bi- and trimodal distribution functions $p(r)$. With real data, which are the main topic of interest here, results are presented for poly(ethyl acrylate) of concentration 3% in D₂O, which is proposed as representative of a typical polymer latex, and for a siloxane-ethoxy copolymer also in D₂O. The results correlate well with alternative approaches, but very polydisperse intensity distributions are problematic. This is not the fault of the method but a consequence of the nature of the scattering equations for such systems. Steenstrup & Hansen (1994) have used a variant of the ME method to determine $p(r)$ from (33), in which the positivity constraint is relaxed to determine $p(r)$ using both simulated and experimental data, the

latter using sodium dodecylsulfate in solution. These results are also encouraging. Finally, Hansen & Wilkins (1994) have used simulated SAXS data for cytochrome *c* to develop ideas of error propagation in the ME approach. A non-uniform prior is used into which uncertainties are incorporated and these errors are allowed to propagate along with the data variances through to the final distribution functions. With this simulated data, the results look convincing; in addition, they suggest possible extensions to crystal structure determination by selectively weighting the prior maps to reflect a degree of belief in their accuracy.

6.2. Crystallite sizes from powder profiles

Particle size in powders is determined by studying line broadening in diffraction patterns. The traditional approach is exemplified by the now classic Warren & Averbach (1950) paper, which depicts the sample as a column-like structure of unit cells perpendicular to the diffracting plane. Each of these columns makes an independent contribution to the powder diffraction pattern. The standard nomenclature is somewhat different from that used in SAS: an interference function of the form

$$G(N, \psi) = \sin^2(N\psi) / \sin^2 \psi \quad (36)$$

arises, where $\psi = 2\pi d \sin \theta / \lambda$ with d the d spacing. If we measure the intensity $I(\psi)$ at θ , the required distribution function for particle sizes is given by

$$I(\psi) = K \sum p(N) G(N, \psi), \quad (37)$$

where K is usually considered to be a constant and $p(N)$ is the required size-distribution function, conventionally determined by Fourier methods. The parallels with (33) are obvious. A main source of error in conventional Fourier approaches arises from problems with peak truncation: a truncation of 20% leads to size errors of over 100% (Young, Gerdes & Wilson, 1967); this poses problems in the case of the overlap of tails in adjacent diffraction peaks. A conventional ME approach to determining $p(N)$ showing many similarities with the SAS calculations in §6.1 is described by Guérin, Alvarez, Rebollo Neira, Plastino & Bonetto (1986). As seems usual, both ideal and real data are used as trials, the latter that of montmorillonite with very small particle size. The conclusion is quite dramatic: truncation errors of 20% lead to only 5% errors in the size distribution. The ME formalism is thus an obvious candidate for data sets where peak overlap is a problem.

Mazumder, Bhagwat & Sequeria (1995) have extended these ideas to extract particle size distributions from multiple small-angle scattering experiments using different priors. The comparisons with more conventional methods look convincing.

6.3. Conclusions

What are the conclusions here? It is obvious that the determination of particle size distributions is, in general, a much simpler and less controversial problem than others that we have addressed in this review, and one that seems to produce good results, in general, with well established inversion techniques. Nonetheless, the ME method has its usual improvements to offer. It can propagate and use experimental errors, it avoids interpolation problems, it admits the use of non-uniform priors, and it is less sensitive than conventional methods to both initial assumptions concerning particle size and shape and to the extent of the measured data. If a computer program using the ME formalism was offered to the SAS community, it would probably be used quite routinely. There is, however, in the author's view much that can still be done especially using non-uniform prior distributions.

7. Algorithms for entropy maximization

7.1. Introduction

Throughout this review, the process of entropy maximization itself has been taken for granted. In this final section, we discuss briefly entropy maximization algorithms. The problem takes two forms:

(i) *Unconstrained*: as in scaling reflections measured under different conditions (Collins, 1984). This is a relatively straightforward problem and is discussed by Prince & Collins (1992), who also give the example of linear fitting using entropy.

(ii) *Constrained*: this is much more difficult. There are already two brief surveys in a crystallographic context by Bricogne (1984, 1993). His nomenclature will be used here since it is best suited for crystallographic purposes and is also that used in §4.3.

Let $q(\mathbf{x})$ be non-uniform prior. It is to have its entropy maximized to generate $q^{\text{ME}}(\mathbf{x})$ relative to a prior $m(\mathbf{x})$ (which, of course, can be uniform), *i.e.* we wish to maximize the entropy, S :

$$S(q) = - \int_V q(\mathbf{x}) \log[q(\mathbf{x})/m(\mathbf{x})] d^3\mathbf{x} \quad (38)$$

or, in the case of N discrete pixels,

$$S(q) = - \sum_{i=1}^N q_i \log q_i / m_i, \quad (39)$$

subject to a set of M constraints. The most useful example is that using M phased reflections (*i.e.* those with known amplitudes and phases) as constraints, but this is easily extendable to other situations. We will define a set $\{\mathbf{H}\}$ of unitary structure factors with amplitude $|U_{\mathbf{h}}|$ and associated phase $\varphi_{\mathbf{h}}$, in which case the constraint equations may be written

$$\int_V q^{\text{ME}}(\mathbf{x}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) d^3 \mathbf{x} = |U_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) = c_j$$

$$(j = 1, 2, \dots, M), \quad (40)$$

i.e. $q^{\text{ME}}(\mathbf{x})$ must reproduce the amplitudes and phases of the U 's in $\{\mathbf{H}\}$ via a Fourier transform. In the discrete case, these can be written

$$\sum_{j=1}^N q_j \exp(2\pi i \mathbf{h}_i \cdot \mathbf{r}_j) = U_{\mathbf{h}_i} \exp(i\varphi_{\mathbf{h}_i})$$

$$= \sum_{j=1}^N q_j A_{ij}, \quad i = 1, \dots, M, \quad (41)$$

where the terms A_{ij} are the constraints, q_j is the number of electrons in the pixel located at \mathbf{r}_j and \mathbf{h}_i is the vector of Miller indices for reflection i . The constraints are convex and so the maximum entropy equations derived from them should lead to a unique solution. In the case of amplitude constraints, without phase information, this property is lost and we are confronted by the so-called 'branching problem' with its associated multimodality problems.

The required constrained maximum is equivalent to the unconstrained minimum of the dual function:

$$L(q) = -S(q) + \sum_{j=1}^M \lambda_j C(q). \quad (42)$$

In principle, the optimization problem is solved by the application of the method of Lagrange multipliers. Let the multiplier associated with constraint i be λ_i and let the vector of such multipliers be λ , then the required maximum entropy distribution is

$$q^{\text{ME}}(\mathbf{x}) = [m(\mathbf{x})/Z(\lambda_1, \lambda_2, \dots, \lambda_M)]$$

$$\times \exp \left[\sum_{j=1}^M -\lambda_j \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) \right]. \quad (43)$$

The λ 's are the Lagrange multipliers, one per constraint. If the constraints are complex, then so are the corresponding multipliers. The normalization requirement gives

$$Z(\lambda_1, \lambda_2, \dots, \lambda_M)$$

$$= \int_V m(\mathbf{x}) \exp \left[\sum_{j=1}^M -\lambda_j \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) \right] d^3 \mathbf{x} \quad (44)$$

and, finally,

$$\partial(\log Z)/\partial \lambda_j = c_j. \quad (45)$$

In the discrete case, the equations are

$$p_i = m_i/Z(\lambda) \exp \left(- \sum_{r=1}^M \lambda_r A_{ri} \right), \quad (46)$$

where

$$Z(\lambda) = \sum_{i=1}^N m_i \exp \left(- \sum_{r=1}^M \lambda_r A_{ri} \right). \quad (47)$$

We now have a set of M coupled non-linear equations and need an efficient algorithm for their solution, *i.e.* we need to determine the individual λ 's to generate $q^{\text{ME}}(\mathbf{x})$. Note, and this is important, that we are not attempting to maximize the entropy with respect to the λ 's (Bricogne, 1991b, 1993).

Perhaps the most obvious method is the Gauss-Newton technique (Agmon, Alhassid & Levine, 1979; Bricogne, 1984; Navaza, 1985, 1991). Let $\lambda^{(i)}$ be a vector of trial Lagrange multipliers at iteration i . Then (Bricogne, 1984),

$$\lambda^{(i+1)} = \lambda^{(i)} + [\mathbf{H}(\log Z^{(i)})]^{-1} (\mathbf{c}^{(i+1)} - \mathbf{c}^{(i)}), \quad (48)$$

where

$$Z^{(i)} = \int_V m(\mathbf{x}) \exp \left[\sum_{j=1}^M -\lambda_j \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) \right] d^3 \mathbf{x} \quad (49)$$

and

$$q^{(i)}(\mathbf{x}) = [m(\mathbf{x})/Z^{(i)}] \exp \left[\sum_{j=1}^M -\lambda_j^{(i)} \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) \right] \quad (50)$$

with

$$c_j^{(i)} = \int_V q^{(i)}(\mathbf{x}) \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) d^3 \mathbf{x}. \quad (51)$$

In the same paper, Bricogne has shown that Bertaut's linearization techniques can be used to generate the elements of $[\mathbf{H}(\log Z^{(i)})]$ in a straightforward way.

The explicit use of the dual function for solving the maximum entropy equations was first suggested by Agmon, Alhassid & Levine (1979). It has been extensively adapted by Bricogne (1993) and used in a different form by Prince (1989, 1993). This method of solving the ME equations is still Gauss-Newton based but works explicitly with (45). In all cases, the shifts predicted by the *NR* method need to be treated with caution since it assumes a locally quadratic function, and the entropy functional is markedly non-quadratic for large shifts. Also, the method may not be globally convergent unless the shifts are controlled. Often, a line-search method along the Newton direction is employed to suitably damp them (Bricogne, 1984, 1993; Prince, 1989). In fact, in all the algorithms described here, the use of line and plane searches (Prince & Boggs, 1992; Bricogne & Gilmore, 1990) is essential in conferring stability.

However, as M increases, the building and inversion of the Hessian matrix can become very slow, as well as posing stability problems with very large matrices; this is a standard problem with the Newton method, but can be overcome by using suitable approximations. These are now discussed.

7.2. Exponential modelling

The key to understanding the exponential modelling method (Collins, 1982; Collins & Mahar, 1983; Bricogne & Gilmore, 1990; Stewart Collins, Watenpugh, Prince & Hall, 1993; Collins & Prince, 1991) is the approximation:

$$[\mathbf{H}(\log \mathbf{Z}^{(i)})]^{-1} \simeq 1/q^{(i)}. \quad (52)$$

The process of entropy maximization can then be re-written as a real-space scheme of map multiplication and division on a pixel-by-pixel basis. In its simplest form, the method is often unstable largely due to the process of taking the reciprocal of q but, with the proper control and shaping of the shifts using line searches along the Newton direction or plane searches along the Newton and constraint directions, it is a method of great stability and versatility that is used in the *MICE* computer program (Bricogne & Gilmore, 1990) with basis set sizes that range from $M = 1$ to several thousand. Since it is a Fourier method employing FFTs, the computational complexity is of order $N \log N$ for N reflections.

7.3. The Skilling algorithm and related techniques

Skilling's algorithm (Skilling & Bryan, 1984; Bricogne, 1984; Titterton, 1985) was designed as a general-purpose method for entropy maximization with incomplete and noisy data, and is not specific to crystallographic problems although it has been used in this environment (*e.g.* Livesey & Skilling, 1985; Potton, Daniell & Rainford, 1988*a,b*). This method addresses the problems of the Hessian matrix by defining a search subspace in n vectors with n typically 3, 4 or 6. The constraints and the entropy are modelled as locally quadratic functions around the current position, giving rise to scalar products which are readily evaluated, and an entropy metric is also defined, which makes it easier to control the size of the shifts that are produced. Bricogne (1984) has produced an elegant algorithm that extends the scope of this method to non-convex (in this case amplitude only) constraints with the associated potential of solving the branching problem in some aspects of the crystallographic phase problem. The algorithm is reported to be very stable and, just as for exponential modelling, has a computational complexity of $N \log N$ for N reflections.

7.4. Pixel methods

In pixel methods, we look explicitly at formulations of the ME equations that explicitly treat the individual map pixels. The first of these involves the single-pixel equations of Wilkins *et al.* (Wilkins, Varghese & Lehmann, 1983; Wilkins & Stuart, 1986; Wilkins, 1983*a,b*). In this approach, each pixel is considered to be independent of any other with its own Lagrangian

multiplier and the exponent in the ME equations is Taylor expanded to first order in q_j . The Hessian matrix is then reduced to a diagonal form by only including those terms $\partial^2 A_{ij}/\partial q_i \partial q_j$ for which $i = j$. Each pixel then becomes independent of any other (hence its name, the single-pixel approximation), and the solution of the ME equations becomes relatively simple and quick to compute by a Newton method. An accelerated convergence procedure is also proposed (Wilkins, 1983*b*) using the contrast between successive iterations. This approach is much used in electron density studies and is examined in this context by Kumazawa, Takata & Sakata (1995*a*) to see if the underlying approximations are valid or could be sufficient to prevent a correct solution. They conclude, from studying a very simple two-pixel model, that, providing the Lagrangian multipliers associated with each pixel remain small, the solutions produced are exact. This constraint on the magnitude of the multipliers is somewhat worrying, but large values would introduce divergences in the calculations, which should be easy to trap in a suitable computer program.

Johnson's *METRIC* algorithm (Johnson, 1987) is also a pixel method that linearizes the exponential in the ME equations *via* the approximation

$$\exp(b\omega) = (1 + a\omega)^{b/a} \quad (53)$$

for a suitable choice of a . To what extent these single-pixel methods face stability problems because of their underlying assumptions is not clear, but current usage of the technique seems restricted.

7.5. Annealing methods

Simulated annealing is an established optimization technique, which is invaluable in situations where the global maximum or minimum is difficult to attain because of the presence of local optimum points. For convex constraints, which have been the major topic of the previous section, this is not a problem but, where amplitude information is used, the non-convexity of the constraints because a problem. McLachlan (1993) has applied annealing techniques to the situation where only amplitude constraints, without phase information, are available. He has called the method entropy phase dynamics. His published application, however, is to a multisolution phase environment using phased subsets of reflections for the membrane protein bacteriorhodopsin (Henderson, Baldwin, Downing, Lepault & Zemlin, 1986). In this case, the procedure seems to find the correct entropy maximum.

7.6. Which method is the most useful for crystallographers?

In some sense, this is an academic question. There are as many variants of methodology of entropy

maximization as there are workers in the field, and there are very few, if any, freely available computer programs or subroutine libraries that perform these calculations in a crystallographic environment. The most notable exception is in the *XTAL* package. *XTAL3.2* (Stewart, Collins, Watenpaugh, Prince & Hall, 1993) has several routines (e.g. *MERUN*, *MEDENS*, *MEFFIT*, *MESTAR* and *MEPHAS*) for generating maximum entropy maps using phased reflections as constraints. In general, however, most workers in the field have code that is perhaps too specialized or too experimental to make it suitable for public distribution. So one must often write one's own, which is a non-trivial task. For relatively simple situations, i.e. unconstrained optimization, it can be relatively straightforward, but for situations where there are multiple constraints the problem can be difficult. Most published algorithms seem to work effectively.

8. Conclusions

Writing and researching this review has been a mixed experience. There is excitement at all the crystallographic problems that are being addressed, from data processing, the phase problem in a wide variety of situations and particle sizes through to accurate electron density determinations, combined with some frustration that it has not yet reached mainstream crystallography. The reasons for this are made clear throughout this review, and they all hinge partly on available software and partly on the relative newness of the ME approach. This will change as the subject matures and with it, I am convinced, will come an explosion of activity. Let us finish with a few words from Jaynes (1989): 'Although the terms "Bayesian" and "Maximum Entropy" appear prominently in the announcements of our meetings, our efforts are somewhat more general. Stated broadly, we are concerned with this: "What are the theoretically valid, and pragmatically useful ways of applying probability theory in science?'''. Later, in the same paper, Jaynes quotes Maxwell: 'Probability theory is itself the true logic of science.'

We hope that we have given some flavour of how maximum entropy and Bayesian statistics are better ways of applying probability theory and logic in crystallography; they are theoretically valid and the string of results demonstrates their pragmatism. The author's own view is summed up in a final quotation from that great American climber, Royal Robbins: '...The other is faith. You have to have faith in what you're doing. You have to have faith that a certain approach is the best one, that it's going to give you the best chance of getting what you're after.' I have that faith in the ME formalism.

My thanks above all to Gérard Bricogne (MRC, Cambridge, England) for his endless enthusiasm and knowledge with a willingness to explain; to Charles Carter (University of North Carolina, USA), to Doug Dorset (HWI, Buffalo) and to everyone who has worked with me in this field for the same commitment. One referee of this paper read it with great care, and made invaluable suggestions for improvement for which I am very grateful.

References

- Agmon, N., Alhassid, Y. & Levine, R. D. (1979). *The Maximum Entropy Formalism*, edited by R. D. Levine & M. Tribus, pp. 207-209. Cambridge: MIT Press.
- Anderson, D. M., Martin, D. C. & Thomas, E. L. (1989). *Acta Cryst.* **A45**, 686-698.
- Baird, T., Gall, J. H., MacNicol, D. D., Mallinson, P. R. & Mitchie, C. R. (1988). *J. Chem. Soc. Chem. Commun.* p. 1471.
- Baldwin, J. M., Henderson, R., Beckman, E. & Zemlin, F. (1988). *J. Mol. Biol.* **202**, 585-591.
- Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3-58.
- Blessing, R. H. (1989). *J. Appl. Cryst.* **22**, 396-397.
- Bricogne, G. (1982). *Computational Crystallography*, edited by D. Sayre, pp. 258-264. Oxford: Clarendon Press.
- Bricogne, G. (1984). *Acta Cryst.* **A40**, 410-445.
- Bricogne, G. (1988a). *Acta Cryst.* **A44**, 517-545.
- Bricogne, G. (1988b). *Crystallographic Computing 4: Techniques and New Technologies*, edited by N. W. Isaacs & M. R. Taylor, pp. 60-79. Oxford: Clarendon Press.
- Bricogne, G. (1991a). *Maximum Entropy in Action*, edited by B. Buck & V. A. Macaulay, pp. 187-216. Oxford University Press.
- Bricogne, G. (1991b). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 257-297. Oxford University Press.
- Bricogne, G. (1991c). *Acta Cryst.* **A47**, 803-829.
- Bricogne, G. (1992). *The Molecular Replacement Method. Proceedings of the Study Weekend held at Daresbury*, edited by W. Wolf, E. J. Dodson & S. Glover, pp. 60-68. Warrington: Daresbury Laboratory.
- Bricogne, G. (1993). *Acta Cryst.* **D49**, 37-60.
- Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* **A46**, 284-297.
- Brisson, A. & Moser, G. (1991). *J. Electron Microsc. Tech.* **18**, 387-393.
- Bryan, R. K. (1988). *Scanning Microscopy*, Supplement 2, pp. 99-105. Chicago: Scanning Microscopy International.
- Bryan, R. K. (1989). *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, pp. 214-224. Dordrecht: Kluwer.
- Bryan, R. K. & Banner, D. W. (1987). *Acta Cryst.* **A43**, 556-564.
- Bryan, R. K., Bansal, M., Folkhard, W., Nave, C. & Marvin, D. A. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4728-4731.
- Buck, B. & Macaulay, V. A. (1991). *Maximum Entropy in Action*. Oxford: Clarendon Press.

- Carter, C. W. Jr (1994). *From First Map to Final Model. Proceedings of the Study Weekend held at Daresbury*, edited by W. Wolf, E. J. Dodson & S. Glover, pp. 41–58. Warrington: Daresbury Laboratory.
- Carter, C. W. Jr & Yin, Y. (1994). *Acta Cryst.* **D50**, 579–590.
- Carvalho, C. A. M., Hashizume, H., Stevenson, A. W. & Robinson, I. K. (1996). *Physica (Utrecht) B*. In the press.
- Castleden, I. R. & Fortier, S. (1994). *Acta Cryst.* **A50**, 9–17.
- CCP4 (1993). *The CCP4 Suite – Computer Programs for Protein Crystallography*. Warrington: Daresbury Laboratory.
- Clegg, W. (1981). *Acta Cryst.* **A37**, 22–28.
- Collins, D. M. (1982). *Nature (London)*, **298**, 49–51.
- Collins, D. M. (1984). *Acta Cryst.* **A40**, 705–708.
- Collins, D. M. (1985). *Structure and Statistics in Crystallography*, edited A. J. C. Wilson, pp. 109–112. New York: Adenine Press.
- Collins, D. M. & Mahar, M. C. (1983). *Acta Cryst.* **A39**, 252–256.
- Collins, D. M. & Prince, E. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 308–316. Oxford University Press.
- Conway, J. H. & Sloane, N. J. A. (1993). *Sphere Packings, Lattices and Groups*. Berlin: Springer-Verlag.
- Daniels, M. E. (1954). *Ann. Math. Stat.* **25**, 631–650.
- David, W. I. F. (1987). *J. Appl. Cryst.* **20**, 316–319.
- David, W. I. F. (1990). *Nature (London)*, **346**, 731–734.
- Dong, W., Baird, T., Fryer, J. R., Gilmore, C., MacNicol, D. D., Bricogne, G., Smith, D. J., O'Keefe, M. A. & Hovmöller, S. (1992). *Nature (London)*, **355**, 605–609.
- Donoho, D. L., Johnstone, I. M., Hoch, J. C. & Stern, A. S. (1992). *J. R. Stat. Soc. Ser. B*, **54**, 41–82.
- Dorset, D. L. (1991a). *Proc. Natl Acad. Sci. USA*, **88**, 5499–5502.
- Dorset, D. L. (1991b). *Acta Cryst.* **A47**, 510–515.
- Dorset, D. L. (1991c). *Ultramicroscopy*, **38**, 23–40.
- Dorset, D. L. (1993). *MSA Bull.* **23**, 99–108.
- Dorset, D. L. (1995). *Structural Electron Crystallography*. New York: Plenum Press.
- Dorset, D. L. & McCourt, (1994). *Acta Cryst.* **A50**, 287–292, 344–356.
- Doublié, S., Bricogne, G., Gilmore, C. J. & Carter, C. W. C. Jr (1995). *Structure*, **3**, 17–31.
- Doublié, S., Xiang, S., Gilmore, C. J., Bricogne, G. & Carter, C. W. C. Jr (1994). *Acta Cryst.* **A50**, 164–182.
- Edwards, A. W. F. (1972). *Likelihood*. Cambridge University Press.
- Estermann, M. A., McCusker, L. B. & Baerlocher, C. (1992). *J. Appl. Cryst.* **25**, 539–543.
- Fan, H., Zhong, Z., Zheng, C. & Li, F. (1985). *Acta Cryst.* **A41**, 163–165.
- Farrow, N. A. & Ottensmeyer, P. (1988). *Scanning Microscopy*, Supplement 2, pp. 75–81. Chicago: Scanning Microscopy International.
- Feigin, A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*. New York: Plenum Press.
- French, S. (1978). *Acta Cryst.* **A34**, 728–738.
- French, S. & Wilson, K. (1978). *Acta Cryst.* **A34**, 517–525.
- Frieden, B. R. (1988). *Scanning Microscopy*, Supplement 2, pp. 107–111. Chicago: Scanning Microscopy International.
- Fryer, J. R. & Gilmore, C. J. (1992). *Trans. Am. Crystallogr. Assoc.* **28**, 57–75.
- Giacovazzo, C. (1993). *Crystallographic Computing 6: a Window on Modern Crystallography*, edited by H. D. Flack, L. Párkányi & K. Simon, pp. 73–84. Oxford University Press.
- Gilmore, C. J. (1992). *Curr. Opin. Struct. Biol.* **2**, 806–810.
- Gilmore, C. J. (1993). *Crystallographic Computing 6: a Window on Modern Crystallography*, edited by H. D. Flack, L. Párkányi & K. Simon, pp. 25–46. Oxford University Press.
- Gilmore, C. J. & Bricogne, G. (1992). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 298–307. Oxford University Press.
- Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *Acta Cryst.* **A46**, 297–308.
- Gilmore, C. J., Henderson, A. N. & Bricogne, G. (1991). *Acta Cryst.* **A47**, 842–846.
- Gilmore, C. J., Henderson, K. & Bricogne, G. (1991). *Acta Cryst.* **A47**, 830–841.
- Gilmore, C. J. & Nicholson, W. N. (1996). *Trans. Am. Crystallogr. Assoc.* In the press.
- Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Proc. R. Soc. London Ser. A*, **442**, 97–111.
- Gilmore, C. J., Shankland, K. & Fryer, J. R. (1992). *Trans. Am. Crystallogr. Assoc.* **28**, 129–139.
- Gilmore, C. J., Shankland, K. & Fryer, J. R. (1993). *Ultramicroscopy*, **49**, 132–146.
- Glatter, O. (1992). *International Tables for Crystallography*, Vol. C, *Mathematical, Physical and Chemical Tables*, pp. 89–136. Dordrecht: Kluwer.
- Glover, I., Haneef, I., Pitts, J. E., Wood, S. P., Moss, D., Tickle, I. J. & Blundell, T. (1983). *Biopolymers*, **22**, 293–304.
- Gonzales, A., Nave, C. & Marvin, D. A. (1995). *Acta Cryst.* **D51**, 792–804.
- Guérin, D. M. A., Alvarez, A. G., Rebollo Neira, L. E., Plastino, A. & Bonetto, R. D. (1986). *Acta Cryst.* **A42**, 30–35.
- Gull, S. F., Livesey, A. K. & Sivia, D. S. (1987). *Acta Cryst.* **A43**, 112–117.
- Hall, S. R. & Stewart, J. M. (1992). Editors. *Xtal3.2 Reference Manual*. Universities of Western Australia, Australia, and Maryland, USA.
- Han, F., Fan, H. & Li, F. (1986). *Acta Cryst.* **A42**, 353–356.
- Hansen, S. & Pedersen, J. S. (1991). *J. Appl. Cryst.* **24**, 541–548.
- Hansen, S. & Wilkins, S. W. (1994). *Acta Cryst.* **A50**, 547–550.
- Harrison, R. W. (1989). *Acta Cryst.* **A45**, 4–10.
- Hauptman, H. A., Weeks, C. M., Smith, G. D., Teeter, M. M. & Miller, R. (1993). Proc. of the American Crystallographic Association Meeting, Albuquerque, USA, P118.
- Havelka, W. A., Henderson, R., Heymann, J. A. W. & Oesterhelt, D. (1993) *J. Mol. Biol.* **234**, 837–846.
- Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. (1986). *Ultramicroscopy*, **19**, 147–178.

- Hill, R. (1991). *A First Course in Coding Theory*. Oxford: Clarendon Press.
- Jauch, W. (1994). *Acta Cryst.* **A50**, 650–652.
- Jauch, W. & Palmer, A. (1993). *Acta Cryst.* **A49**, 590–591.
- Jaynes, E. T. (1957). *Phys. Rev.* **106**, 620–630.
- Jaynes, E. T. (1979). *The Maximum Entropy Formalism*, edited by R. D. Levine & M. Tribus, pp. 15–118. Cambridge: MIT Press.
- Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, edited by R. D. Rosenkrantz. *Synthese Library*, Vol. 158. Dordrecht: Reidel.
- Jaynes, E. T. (1986). *Maximum Entropy and Bayesian Methods in Applied Statistics*, edited by J. H. Justice, pp. 26–58. Cambridge University Press.
- Jaynes, E. T. (1987). *Contemp. Phys.* **28**, 501–502.
- Jaynes, E. T. (1989). *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, pp. 1–28. Dordrecht: Kluwer.
- Jeffreys, H. (1939). *Theory of Probability*. Oxford University Press.
- Johnson, M. W. (1987). *Metric: an Algorithm for Maximum Entropy Reconstruction Using Iterative Multiple Constraints*. Chilton: Rutherford Appleton Laboratory.
- Koritsansky, T. (1993). *Crystallographic Computing 6: a Window on Modern Crystallography*, edited by H. D. Flack, L. Párkányi & K. Simon, pp. 145–168. Oxford University Press.
- Kumazawa, S., Kubota, Y., Takata, M. & Sakata, M. (1993). *J. Appl. Cryst.* **26**, 453–457.
- Kumazawa, S., Takata, M. & Sakata, M. (1995a). *Acta Cryst.* **A51**, 47–53.
- Kumazawa, S., Takata, M. & Sakata, M. (1995b). *Acta Cryst.* **A51**, 651–658.
- Lapthorn, A. J., Harris, D. C., Littlejohn, A., Lustbader, J. W., Canfield, R. E., Machin, K. J., Morgan, F. J. & Isaacs, N. W. (1994). *Nature (London)*, **369**, 455–461.
- Lehmann, M. S., Robinson, T. E. & Wilkins, S. W. (1986). *J. Phys. (Paris)*, **47**, C5, 55–60.
- Lemaréchal, C. & Navaza, J. (1991). *Acta Cryst.* **A47**, 631–632.
- Lenstra, A. T. H., Verbruggen, M., Bracke, B., Vanhouteghen, F., Reyniers, F. & Borremans, F. (1991). *Acta Cryst.* **B47**, 92–97.
- Lightfoot, P., Tremayne, M., Harris, K. D. M., Glidewell, C., Shankland, K., Gilmore, C. J. & Bruce, P. G. (1993). *Mater. Sci. Forum*, **133–136**, 207–212.
- Liu, Y., Fan, H. & Zheng, C. (1988). *Acta Cryst.* **A44**, 61–63.
- Livesey, A. & Skilling, J. (1985). *Acta Cryst.* **A41**, 113–122.
- McCusker, L. (1988). *J. Appl. Cryst.* **21**, 305–310.
- McLachlan, A. D. (1993). *Acta Cryst.* **D49**, 75–85.
- MacWilliams, F. J. & Sloane, N. J. A. (1977). *The Theory of Error Correcting Codes*. Amsterdam: North-Holland.
- Marvin, D. A., Bryan, R. K. & Nave, C. (1987). *J. Mol. Biol.* **193**, 315–343.
- Mazumder, S., Bhagwat, K. V. & Sequeria, A. (1995). *J. Phys. Condens. Matter*, **7**, 9737–9746.
- Millane, R. P. (1988). *Crystallographic Computing 4: Techniques and New Technologies*, edited by N. W. Isaacs & M. R. Taylor, pp. 169–186. Oxford University Press.
- Molina, E. C. (1963). *Two Papers by Bayes with Commentaries*. New York: Hafner Publishing Company.
- Morris, R. E., Harrison, W. T. A., Nicol, J. M., Wilkinson, A. P. & Cheetham, A. K. (1992). *Nature (London)*, **359**, 519–522.
- Morrison, J. D., Corcoran, J. D. & Lewis, K. E. (1992). *J. Appl. Cryst.* **21**, 504–513.
- Nakahigashi, K. & Higashimine, K. (1995). *J. Phys. Condens. Matter*, **7**, 6961–6966.
- Narayan, R., Nityanada, R. & Vani, G. V. (1983). *Contributions to Crystallography*, edited by C. N. R. Rao, pp. 51–68. Bangalore: Indian Academy of Sciences.
- Navaza, J. (1985). *Acta Cryst.* **A41**, 232–244.
- Navaza, J. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 317–323. Oxford University Press.
- Neyman, J. & Pearson, E. S. (1933a). *Proc. Cambridge Philos. Soc.* **29**, 492–510.
- Neyman, J. & Pearson, E. S. (1933b). *Philos. Trans. R. Soc. London Ser. A*, **231**, 289–337.
- Oatley, S. & French, S. (1982). *Acta Cryst.* **A38**, 537–549.
- O'Hagen, A. (1988). *Probability, Methods and Measurement*. London: Chapman and Hall.
- O'Hagen, A. (1994). *Kendall's Advanced Theory of Statistics*, Vol. 2B, *Bayesian Inference*. London: Edward Arnold.
- Papoular, R. J. & Delapalme, A. (1994). *Phys. Rev. Lett.* **72**, 1486–1489.
- Papoular, R. J. & Gillon, B. (1990). *Neutron Scattering Data Analysis 1990*, edited by M. W. Johnson. *Inst. Phys. Conf. Ser.* No. 107, pp. 101–116.
- Papoular, R. J., Zheludev, A., Ressouche, E. & Schweizer, J. (1995). *Acta Cryst.* **A51**, 295–300.
- Potter, J. A., Daniell, G. J. & Rainford, B. D. (1988a). *J. Appl. Cryst.* **21**, 663–668.
- Potter, J. A., Daniell, G. J. & Rainford, B. D. (1988b). *J. Appl. Cryst.* **21**, 891–897.
- Prince, E. (1989). *Acta Cryst.* **A45**, 200–203.
- Prince, E. (1993). *Acta Cryst.* **D49**, 61–65.
- Prince, E. & Boggs, P. T. (1992). *International Tables for Crystallography*, Vol. C, *Mathematical, Physical and Chemical Tables*, pp. 594–604. Dordrecht: Kluwer.
- Prince, E. & Collins, D. (1992). *International Tables for Crystallography*, Vol. C, *Mathematical, Physical and Chemical Tables*, pp. 605–608. Dordrecht: Kluwer.
- Prince, E., Sjölin, L. & Alenljung, R. (1988). *Acta Cryst.* **A44**, 216–222.
- Reid, N. (1988). *Stat. Sci.* **3**, 213–238.
- Restori, R., Schwarzenbach, D. & Schneider, J. R. (1987). *Acta Cryst.* **B43**, 251–257.
- Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
- Roth, M. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 229–248. Oxford University Press.
- Rowlinson, J. S. (1970). *Nature (London)*, **225**, 1196–1198.
- Saka, T. & Kato, N. (1986). *Acta Cryst.* **A42**, 469–478.
- Sakata, M. & Sato, M. (1990). *Acta Cryst.* **A46**, 263–270.
- Sakata, M., Uno, T., Takata, M. & Mori, R. (1992). *Acta Cryst.* **B48**, 591–598.
- Sato, T. (1992). *Acta Cryst.* **A48**, 842–850.
- Sayre, D. (1993). *Acta Cryst.* **D49**, 107.

- Schlutzen, F., Volkmann, N., Thygesen, J., Hansen, H. A. S., Harms, J., Bennett, W. S. & Yonath, A. (1994). Proc. of the American Crystallographic Association Meeting, Atlanta, USA, TRN15.
- Schotte, K.-D., Schotte, U., Bleif, H.-J. & Papoular, R. (1995). *Acta Cryst.* **A51**, 739–746.
- Shankland, K. (1996). *Trans. Am. Crystallogr. Assoc.* In the press.
- Shankland, K., Gilmore, C. J., Bricogne, G. & Hashizume, H. (1993). *Acta Cryst.* **A49**, 493–501.
- Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Shirley, R. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. R. Hall & T. Ashida, pp. 411–437. Oxford: Clarendon Press.
- Shore, J. E. & Johnson, R. W. (1980a). *IEEE Trans. Inf. Theory*, **IT-26**, 26–39.
- Shore, J. E. & Johnson, R. W. (1980b). *IEEE Trans. Inf. Theory*, **IT-29**, 942–943.
- Sivia, D. S. & David, W. I. F. (1995). *Acta Cryst.* **A50**, 703–714.
- Sivia, D. S., David, W. I. F., Knight, K. K. & Gull, S. F. (1993). *Physica (Utrecht)*, **D66**, 234–242.
- Sjölin, L., Prince, E., Svensson, L. A. & Gilliland, G. L. (1991). *Acta Cryst.* **A47**, 216–223.
- Sjölin, L., Svensson, L. A., Prince, E. & Sundell, S. (1990). *Acta Cryst.* **B46**, 209–215.
- Skilling, J. & Bryan, R. K. (1984). *Mon. Not. R. Astron. Soc.* **211**, 111–124.
- Steenstrup, S. & Hansen, S. (1994). *J. Appl. Cryst.* **27**, 574–580.
- Steenstrup, S. & Wilkins, S. W. (1989). *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, pp. 203–212. Dordrecht: Kluwer.
- Stewart, J. M., Collins, D. M., Watenpaugh, K., Prince, E. & Hall, S. R. (1993). *Acta Cryst.* **D49**, 100–106.
- Sudo, N., Hashizume, H. & Carvalho, C. A. M. (1995). *Powder Diffr.* **10**, 34–39.
- Takata, M., Umeda, B., Nishibori, E., Sakata, M., Saito, Y., Ohno, M. & Shinohara, H. (1995). *Nature (London)*, **37**, 46–49.
- Tikochinsky, T., Tishby, N. Z. & Levine, R. D. (1984). *Phys. Rev. Lett.* **52**, 1357–1360.
- Titterton, D. M. (1984). *Nature (London)*, **312**, 381–382.
- Titterton, D. M. (1985). *Astron. Astrophys.* **144**, 381–387.
- Tremayne, M., Lightfoot, P., Glidewell, C., Mehta, M. A., Bruce, P. G., Harris, K. D. M., Shankland, K., Gilmore, C. J. & Bricogne, G. (1992). *J. Solid State Chem.* **100**, 191–196.
- Tremayne, M., Lightfoot, P., Harris, K. D. M., Shankland, K., Gilmore, C. J., Bricogne, G. & Bruce, P. G. (1992). *J. Mater. Chem.* **2**, 1301–1302.
- Tribus, M. (1961). *J. Appl. Mech.* **28**, 1–8.
- Tribus, M. (1979). *The Maximum Entropy Formalism*, edited by R. D. Levine & M. Tribus, pp 1–14. Cambridge: MIT Press.
- Vainshtein, B. K. (1955). *Zh. Fiz. Kim.* **29**, 327–344.
- Vainshtein, B. K. (1964). *Structure Analysis by Electron Diffraction*. Oxford: Pergamon Press.
- Vainshtein, B. K., Zvyagin, B. B. & Avilov, A. S. (1992). *Electron Diffraction Techniques*, Vol. 1, edited by J. M. Cowley, pp. 216–312. Oxford: Clarendon Press.
- Voigt-Martin, I. G., Yan, D. H., Gilmore, C. J., Shankland, K. & Bricogne, G. (1994). *Ultramicroscopy*, **56**, 271–288.
- Voigt-Martin, I. G., Yan, D. H., Yakimansky, A., Schollmeyer, D., Gilmore, C. J. & Bricogne, G. (1995). *Acta Cryst.* **A51**, 849–868.
- Voronova, A. A. & Vainshtein, B. K. (1958). *Sov. Phys. Crystallogr.* **3**, 445–451.
- Vries, R. Y. de, Briels, W. J & Feil, D. (1994). *Acta Cryst.* **A50**, 383–391.
- Warren, B. E. & Averbach, B. L. (1950). *J. Appl. Phys.* **21**, 595–599.
- Wilkins, S. W. (1983a). *Acta Cryst.* **A39**, 892–896.
- Wilkins, S. W. (1983b). *Acta Cryst.* **A39**, 896–898.
- Wilkins, S. W. & Stuart, D. (1986). *Acta Cryst.* **A43**, 556–564.
- Wilkins, S. W., Varghese, J. N. & Lehmann, M. S. (1983). *Acta Cryst.* **A39**, 47–60.
- Wilkinson, C., Khamis, H. A., Stansfield, R. F. & McIntyre, G. J. (1988). *J. Appl. Cryst.* **21**, 471–478.
- Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
- Woolfson, M. M. (1987). *Acta Cryst.* **A43**, 593–612.
- Woolfson, M. M. & Fan, H. (1995). *Physical and Non-physical Methods of Solving Crystal Structures*. Cambridge University Press.
- Xiang, S., Carter, C. W. C. Jr, Bricogne, G. & Gilmore, C. J. (1993). *Acta Cryst.* **D49**, 193–212.
- Young, R. A., Gerdes, R. J. & Wilson, A. J. C. (1967). *Acta Cryst.* **22**, 155–162.
- Zandbergen, H. W. & Cava, R. J. (1995). *Nature (London)*, **374**, 682.
- Zheludev, A., Papoular, R. J., Ressouche, E. & Schweizer, J. (1995). *Acta Cryst.* **A51**, 450–455.