

Journal of

**Applied  
Crystallography**

ISSN 0021-8898

Editor: **Gernot Kosterz**

## **Extending the charge-flipping method towards structure solution from incomplete data sets**

**Lukáš Palatinus, Walter Steurer and Gervais Chapuis**

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

# Extending the charge-flipping method towards structure solution from incomplete data sets

Lukáš Palatinus,<sup>a\*‡</sup> Walter Steurer<sup>b</sup> and Gervais Chapuis<sup>a</sup>

<sup>a</sup>Ecole Polytechnique Fédérale de Lausanne, Laboratoire de Cristallographie, BSP, CH-1015 Lausanne, Switzerland, and <sup>b</sup>Laboratory of Crystallography, Department of Materials, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland. Correspondence e-mail: palat@fzu.cz

The charge-flipping method tends to fail if applied to an incomplete diffraction data set. The reason is artifacts induced in the density maps by Fourier transforming the data. It is shown that the missing data can be sufficiently well approximated on the basis of the Patterson map of the unknown structure optimized by the maximum entropy method (MEM). Structures that could not be solved by the original charge-flipping algorithm can be solved by the proposed method. The method has been tested on experimental data of one inorganic and two organic structures and on several types of missing data. In many cases, up to 50% of missing reflections, or even more, can be tolerated and the structure can still be reconstructed by charge flipping.

© 2007 International Union of Crystallography  
Printed in Singapore – all rights reserved

## 1. Introduction

Despite the widespread usage and success of direct methods in solving crystal structures, problems still exist that are difficult or impossible to tackle by the well established methods. Some of these problems can be solved by alternative structure solution methods. Charge flipping is such an alternative method that can be used to solve crystal structures for up to several hundreds of atoms in the unit cell from diffraction data with atomic resolution (Oszlányi & Sütö, 2004, 2005). Charge flipping has been shown to be applicable to standard crystal structures (Wu *et al.*, 2004; Oszlányi *et al.*, 2006), but it has also proved very useful in the solution of incommensurately modulated structures (Palatinus, 2004; Zuñiga *et al.*, 2006; Palatinus *et al.*, 2006), quasicrystals (Katrach *et al.*, 2007) and structures solved from powder diffraction data (Wu *et al.*, 2006; Baerlocher *et al.*, 2007). From the practical point of view, the method has only one serious disadvantage, namely the requirement of a complete or almost complete data set. A small proportion of missing reflections can prevent the structure solution completely.

The maximum entropy method (MEM) is a well known and widely used image-enhancement method (Buck & Macaulay, 1991; von der Linden *et al.*, 1999). It has been used in crystallography both in the initial processes of structure solution for phasing structure factors (Bricogne, 1993, 1997; Gilmore, 1996; Gilmore *et al.*, 1999) and in the final stages of refinement for extracting detailed information about the electron density in the unit cell (Steurer *et al.*, 1993; Iversen & Larsen, 1995; Haibach & Steurer, 1996; Papoular *et al.*, 1996, 2002; Roversi *et al.*, 1998; Dinnebier *et al.*, 1999; Wang *et al.*, 2001). It has been noticed that the nonlinear properties of the MEM could

be used also for extrapolation of the amplitudes and phases of unobserved structure factors (Sakata & Sato, 1990), but, to our knowledge, this idea was not exploited in detail, probably because the research was focused on the investigation of the electron density in real space and the MEM was applied to accurate electron density studies, where the data completeness is usually not an issue.

This article focuses on a combination of the MEM with charge flipping for structure solution from incomplete X-ray diffraction data. It will be shown that the MEM can be used to extrapolate the intensities of the missing reflections by optimizing the Patterson function. Combining the extrapolated reflection intensities with the experimental data greatly enhances the ability of charge flipping to solve a crystal structure from an incomplete data set.

## 2. Method

### 2.1. Charge flipping

Charge flipping is an iterative method for *ab initio* reconstruction of electron densities from diffraction data (Oszlányi & Sütö, 2004, 2005). It uses as an input only the cell parameters of the structure, the reflection indices and intensities. Neither chemical information nor the symmetry is explicitly used in the structure solution process. The electron density is sampled on a discrete rectangular grid of pixels with values  $\rho_i, i = 1, \dots, N_{\text{pix}}$ . The algorithm proceeds in iteration cycles. Before the iteration, a starting set of structure factors is created by combining the experimental structure-factor amplitudes  $|F(\mathbf{H})|$  with random phases. One iteration cycle involves four steps.

(1) A trial electron density  $\rho^{(n)}$  is obtained by inverse Fourier transform of the structure factors  $F^{(n)}(\mathbf{H})$ :

‡ Permanent address: Institute of Physics, Academy of Sciences of the Czech Republic, Na Slovance 2, 182 21 Prague, Czech Republic.

$$\rho^{(n)} = \text{FT}^{-1}[F^{(n)}(\mathbf{H})].$$

(2) A modified density  $g^{(n)}$  is obtained from  $\rho^{(n)}$  by reversing the sign (flipping) of all density pixels with density below a certain positive threshold  $\delta$ :

$$g_i^{(n)} = \begin{cases} \rho_i^{(n)} & \text{if } \rho_i^{(n)} > \delta \\ -\rho_i^{(n)} & \text{if } \rho_i^{(n)} \leq \delta. \end{cases}$$

(3) The structure factors of this modified density are obtained by Fourier transform of  $g^{(n)}$ :

$$G^{(n)}(\mathbf{H}) = \text{FT}[g^{(n)}].$$

(4) The structure factors  $F^{(n+1)}(\mathbf{H})$  are obtained from  $F^{(n)}(\mathbf{H})$  and  $G^{(n+1)}(\mathbf{H}) = |G^{(n+1)}(\mathbf{H})| \exp[2\pi i\varphi_G(\mathbf{H})]$  according to the following scheme:

$$F^{(n+1)}(\mathbf{H}) = |F(\mathbf{H})| \exp[2\pi i\varphi_G(\mathbf{H})]$$

for  $F(\mathbf{H})$  observed and strong,

$$F^{(n+1)}(\mathbf{H}) = |G^{(n)}(\mathbf{H})| \exp\{2\pi i[\varphi_G(\mathbf{H}) + 0.25]\}$$

for  $F(\mathbf{H})$  observed and weak,

$$F^{(n+1)}(\mathbf{H}) = 0$$

for  $F(\mathbf{H})$  unobserved, and

$$F^{(n+1)}(\mathbf{H}) = G^{(n)}(\mathbf{H})$$

for  $\mathbf{H} = 0$ .

The new set of structure factors enters the next cycle of iteration. The iteration cycles are repeated until the calculation converges.  $\delta$  is the main parameter of the iteration. It must be determined by trial and error, but this search can be automated. The second variable parameter of the algorithm is the proportion of the reflections considered weak in the fourth step of the iteration cycle. Experience shows that about 20–40% of all reflections can be considered weak. The usage of the phase shifting of the weak reflections significantly improves the performance of the algorithm in cases of more complex structures (Oszlányi & Sütö, 2005).

The algorithm seeks a Fourier map that is stable against repeated flipping of all density regions below  $\delta$ . Obviously, a large number of missing reflections will make the algorithm less efficient, because the missing reflections are assigned a zero amplitude, which induces large termination ripples in the Fourier map. The underlying assumption of the algorithm that the density is close to zero in large regions of the unit cell and positive in small parts of the unit cell is no longer fulfilled and the algorithm fails. The way to improve the performance of the algorithm is thus to avoid the occurrence of large termination ripples.

## 2.2. The maximum entropy method

The MEM originates from information theory and is based on the assumption that the information content of a positive additive distribution is monotonically related to the entropy of

the distribution (Shannon, 1948; Sivia, 1996). Originally the theory was formulated only for probability distributions, but later it was generalized towards any positive additive distribution, including electron densities. The Shannon (information) entropy is defined by

$$S = - \int p(x) \log \frac{p(x)}{\tau(x)} dx,$$

where  $p(x)$  is the probability distribution of a variable  $x$  and  $\tau(x)$  is the prior probability distribution. The distribution  $\tau(x)$  can include any information known about the distribution  $p(x)$  prior to including the experimental data. For many applications  $\tau(x)$  is taken to be constant, and such prior distribution is usually called a uniform or flat prior. The theory underlying the method shows that the most probable among several distributions complying with the experimental information is the one with the maximum value of entropy, while satisfying one or more constraints. In the study of the electron densities two constraints are usually used. The first fixes the total number of electrons in the unit cell to the number defined by the chemical composition, and the second defines the agreement of the reconstructed distribution with the experimental data. In the studies of the electron density, this constraint usually takes the form of a  $\chi^2$  constraint:

$$\chi^2 = \frac{1}{N_F} \sum_{\mathbf{H}} \left\{ \frac{|F_{\text{obs}}(\mathbf{H}) - F_{\text{MEM}}(\mathbf{H})|}{\sigma[F_{\text{obs}}(\mathbf{H})]} \right\}^2. \quad (1)$$

The summation runs over all diffraction vectors  $\mathbf{H}$  with intensities known from the experiment, and the subscripts obs and MEM refer to the experimental and calculated structure factors, respectively. The expectation value of  $\chi^2$  is 1. An overview of applications of the MEM in crystallography was given by Gilmore (1996).

The precise meaning and reliability of the distributions reconstructed by the MEM has been subject to long discussions and remains somewhat controversial. However, an obvious property of the reconstructed distributions apart from their positivity is their smoothness with respect to  $\tau(x)$ , because the entropy decreases with increasing deviation of the distribution  $p(x)$  from the prior distribution  $\tau(x)$ .

In this work, the purpose of the application of the MEM is to enhance the input data for a structure solution method. Thus, the distribution to be optimized by the MEM cannot be the electron density itself, since it is unknown. Instead, its autoconvolution, the Patterson function, can be used. The Patterson function  $P(\mathbf{r})$  is related to the diffracted intensities by the Fourier transform, and as such can be calculated directly from the experimental data:

$$P(r) = \text{FT}^{-1}|F_{\mathbf{H}}|^2.$$

The relation between the Patterson map and reflection intensities is the same as the relation of the electron density and the structure factors. Thus, the MEM formalism can be applied directly to the reconstructions of the Patterson function, with the structure factors  $F(\mathbf{H})$  in the expression for

**Table 1**

Characteristics of the structures used for the tests.

Name	Code	Composition	Z	Symmetry	$V_{UC}$ (Å <sup>3</sup> )	Resolution (sin $\theta/\lambda$ )
Hexachlorotetra- $\mu_3$ -chlorotetra- $\mu_3$ -tellurooctohexarhenium(III)	br1121	Re <sub>6</sub> Te <sub>4</sub> Cl <sub>10</sub>	2	$P\bar{1}$	1035.5	0.61
3,5-Dihydroxybenzoic acid-4,4'-bipyridine (2/3)	sk1293	C <sub>44</sub> H <sub>36</sub> N <sub>6</sub> O <sub>8</sub>	2	$P\bar{1}$	1809.4	0.60
Nicotinamide adenine dinucleotide tetrahydrate	gs1072	C <sub>21</sub> H <sub>35</sub> N <sub>7</sub> O <sub>18</sub> P <sub>2</sub>	1	$P1$	773.7	0.83

References: br1121: Mironov *et al.* (1996); sk1293: Wheatley *et al.* (1999); gs1072: Guillot *et al.* (2000).

the constraint [equation (1)] replaced by the intensities  $I(\mathbf{H}) = |F(\mathbf{H})|^2$ .

If the Patterson function is calculated from a data set with many missing reflections, it contains sharp ripples caused by the missing data and can even exhibit negative regions. On the other hand, the Patterson function optimized by the MEM will be everywhere positive and smooth owing to the properties of the MEM. The smoothing and positivity are achieved by adjusting the intensities of the observed reflections within the limits given by their standard deviations (noise filtering) and by assigning non-zero amplitudes to Fourier coefficients that are not present in the input data and thus not restricted (intensity extrapolation). It is this second property of the MEM that is exploited in this work.

The MEM calculations were performed using the program *BayMEM* (van Smaalen *et al.*, 2003). We used the Sakata–Sato algorithm (Sakata & Sato, 1990) with flat prior and the  $\chi^2$  constraint [equation (1)]. To obtain an extrapolated set of reflections, the experimental data set is analyzed and the list of missing reflections in a sphere of desired resolution is produced. The experimentally known reflection intensities are used as an input to *BayMEM*. The program is designed for reconstructions of the electron densities, and it takes as an input the reflection indices together with the real and imaginary part of the structure factor. No modification to the program was necessary to make it reconstruct the Patterson function. It is just necessary to supply the reflection intensity at the place of a real part of the structure factor and to set all imaginary parts of the structure factors to zero. Moreover, the number of electrons in the unit cell [which corresponds to the value of  $F(\mathbf{0})$ , *i.e.* the amplitude of the forward scattering] must be replaced by its square. If  $F(\mathbf{0})$  is not known *a priori*, its approximate value must be determined using a Wilson plot.

*BayMEM* has an option to output the values of Fourier coefficients of the reconstructed distribution from a list supplied by the user. Using this option, the extrapolated intensities of the missing reflections are obtained. These intensities are merged with the observed data to form a complete reflection set that can be used as an input for the charge-flipping algorithm. The structure solution by charge flipping proceeds further in a completely standard manner as described in §2.1, using the computer program *Superflip* (Palatinus & Chapuis, 2006).

### 3. Testing procedure

We decided to test the new method on experimental data sets rather than data sets generated from a known structure. The

main advantage of using generated data sets is that possible influence of experimental errors on the performance of the method is avoided, and the answer to the problem is known *a priori*. However, the aim of the current method is to solve crystal structures, *i.e.* to find the positions of the atoms in the unit cell, and the quality of the X-ray diffraction data is nowadays often so good that the influence of possible experimental error on the structure solution process is negligible and the refined structure model can be considered unequivocal. Thus, the advantages of using theoretical data are minor. On the other hand, using the experimental data in testing the method has the advantage that the second testing is avoided; this would otherwise be needed to prove that the method works equally well on experimental data. Moreover, the MEM formalism requires that the standard deviations of the observed intensities are known, and using experimental data is the most easy and realistic way to obtain them.

We selected three structures for which a good experimental data set was available, that is, a data set with good completeness and good resolution, which yielded a structure solution with acceptable *R* value. The structures represent one inorganic and two organic structures; among the latter two, one is centrosymmetric and one non-centrosymmetric. Their characteristics are summarized in Table 1. The structures were selected so that they were reasonably complex, but at the same time could be solved easily from the complete data set, and the solution from the complete data set would yield a density with all atomic positions easily recognizable.

All three structures are triclinic. Higher than triclinic symmetry would mean that special care would have to be taken when constructing the incomplete data sets in cases of special geometries of missing regions (see below). Moreover, structures with higher symmetry have special directions and planes. A coincidence of such special direction with a specific direction of the region of missing reflections could make the tests less general.

The nature of missing regions in the data set can vary greatly. We selected five prototypic cases for our tests. In each case the method of excluding reflections from the complete data set has one or more free parameters, as follows.

(i) Upper resolution limit. Several data sets were generated with progressively lower resolution. A resolution lower than optimal can occur if the crystal diffracts poorly at high angles or if the maximal diffraction angle is limited by instrumentation, for example by the opening of a furnace or a diamond anvil cell.

(ii) Lower resolution limit. The variable parameter in this method is the lower limit on the accessible diffraction angle.

All reflections with diffraction angle below the limit were excluded from the data set.

(iii) Randomly omitted reflections. The variable parameter is the ratio of the omitted reflections to all reflections. Each reflection from the complete data set was rejected with probability equal to that ratio. In practice a data set with missing reflections irregularly distributed in the reciprocal space can occur if some reflections have to be omitted as a result of overlaps with other reflections from the measured crystal, with impurities or with powder rings from the gasket in the diamond anvil cell.

(iv) Missing cone. A double cone of reflections around a given axis with a given opening angle is omitted. The region of omitted reflection extends along the axis in both directions from the origin. The variable parameter is the vector of the cone axis and the opening half-angle of the cone.

(v) Present cone. Similarly to the previous case, a double cone is defined by its axis and opening angle, but in this case the reflections in the double cone are preserved and the rest are omitted. In practice a situation with missing or present cones can occur if the crystal is rotated only around one rotation axis during the data collection, or in measurements under non-ambient conditions, where part of the reciprocal space is not accessible because of bulky equipment.

The testing data sets were generated from the complete data sets of each structure by excluding reflections according to one of the above schemes. A large range of each variable parameter was exploited to obtain reliable statistics. Two inputs for the charge-flipping calculation were then created from each generated data set. The first input contained only the reflections present in the data set. This input represented the classical application of charge flipping to an incomplete data set and was used as a reference. Secondly, the intensities of reflections excluded from the data set were extrapolated by the method described in §2.2 and merged with the original data set. The completed data set was then submitted to the charge-flipping calculation. The electron densities of the centrosymmetric structures (br1121 and sk1293) were averaged according to the expected symmetry of the density. Because the charge-flipping algorithm is initiated by assigning random phases, different runs of the algorithm can lead to different results. To obtain good statistics of the reliability of the reconstruction, 25 charge-flipping runs were performed on each input data set. The value of  $F(\mathbf{0})$  for the MEM calculation was calculated from the known structure model. This is the only difference of the present testing from the real-world application. It was used to limit the number of variables that could influence the results and should not have a severe impact on the practical applications.

The tests were performed on hundreds of different data sets, and therefore an automated and unified method had to be developed for comparison of the results. There are essentially two possible approaches to the evaluation of the quality of the structure solution: in reciprocal space and in direct space.

Since in principle the structure solution is a search for the phases of the structure factors, a natural way to express the quality of the reconstruction is to use some measure of the

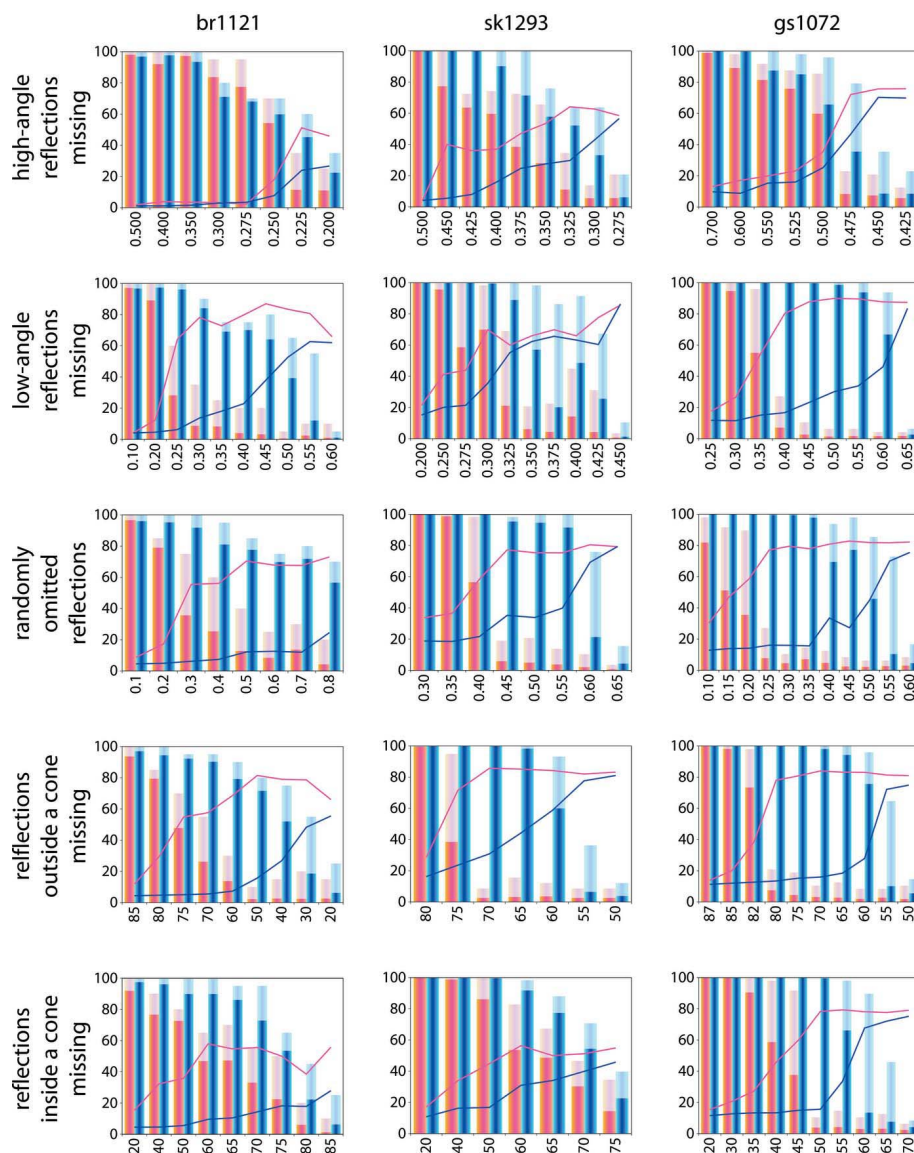
phase difference between the phases obtained by charge flipping and the phases calculated from the refined structure model.

Throughout this work we use a weighted mean square phase difference defined as

$$w\Delta^2(\varphi) = 100 \times \frac{1}{\Delta_{\text{rnd}}^2} \frac{\sum_{\mathbf{H}_{\text{obs}}} |F(\mathbf{H})|^2 \Delta(\varphi)^2}{\sum_{\mathbf{H}_{\text{obs}}} |F(\mathbf{H})|^2}.$$

$\Delta(\varphi)$  is the difference between the structure-factor phase calculated from the structure model and the phase obtained from charge flipping, mapped to the interval  $(-\pi, \pi)$  by subtracting or adding integer multiples of  $2\pi$ . The normalization constant  $\Delta_{\text{rnd}}^2$  is selected so that the expected value of  $w\Delta^2(\varphi)$  for a completely random density is 100.  $w\Delta^2(\varphi)$  gives a good overall measure of the quality of the reconstruction, especially for different reconstructions within the same type of omitted reflections. It is especially useful for the reconstructions from difficult data sets, where it is hard to evaluate the density automatically in real space because of the abundant noise. On the other hand,  $w\Delta^2(\varphi)$  alone cannot be used as the ultimate measure of the quality of the reconstruction. This is because two electron densities that give the same  $w\Delta^2(\varphi)$  can still be very different in terms of the number and clarity of the atomic maxima they contain, depending on which reflections contribute most to  $w\Delta^2(\varphi)$ . Finally it should be noted that the origin of the density reconstructed by charge flipping is in general shifted with respect to the structure model. This origin shift induces a phase shift in the structure factors that, if not taken into account, would lead to completely wrong values of  $w\Delta^2(\varphi)$ . To avoid this problem, the origin of each reconstructed electron density was shifted to align it to the structure model, and only then  $w\Delta^2(\varphi)$  was evaluated. The alignment was performed using the density-matching option of the program *Superflip* and is based on the evaluation of the correlation between the actual density and the reference density.

The other obvious possibility to evaluate the quality of the reconstructed electron density is to count the number of atoms that can be identified in the density. The positions of the atoms in the structure are the values of interest and thus this criterion better reflects the practical usefulness of the reconstruction than  $w\Delta^2(\varphi)$ . The following procedure was adopted to count the number of atoms in the electron density. First the electron density is analyzed and the list of maxima is created; this contains the first  $N_{\text{str}}$  maxima ( $N_{\text{str}}$  being the number of independent atoms in the structure model) plus all maxima with peak electron density higher than 80% of the density of the  $N_{\text{str}}$ th maximum. The maxima that are closer than 0.55 Å to a higher maximum in the list are then eliminated. This reduced list of maxima is compared with the list of atomic positions from the structure model and each maximum closer than 0.55 Å to an atom from the structure model is considered to represent that atom. If more than one maximum lies in the vicinity of an atom, only the nearest maximum is taken into account. The list of maxima located in this way is the first



**Figure 1**

The number of located atoms and  $w\Delta^2(\varphi)$ . Each plot summarizes the results for the type of missing reflections indicated on the left and for the structure indicated at the top. The horizontal axis of each plot represents the variable parameter of the type of missing reflections ( $\sin \theta/\lambda$  in rows 1 and 2, ratio of missing reflections in row 3, and opening half-angle of the cone in rows 4 and 5; see also §3). The orange and blue columns represent the percentage of the atoms located in the charge-flipping results from the incomplete and MEM-enhanced data sets, respectively. The solid part of the columns corresponds to the average percentage of located atoms from the given data set; the transparent part shows the maximum number of located atoms among the 25 calculations on each data set. The orange and blue lines represent the average  $w\Delta^2(\varphi)$  for the calculation on the incomplete and MEM-enhanced data sets, respectively.

tentative list of identified atoms. It is then filtered to decrease the influence of the assignment of noise to the atoms,<sup>1</sup> and the number of assigned maxima after the filtering is considered to represent the number of located atoms. This method has several parameters and steps that are to some extent arbitrary. The method itself and the parameters were tuned so that the

<sup>1</sup> The filtering is based on the observation that among the maxima belonging to the noise there are also a large number of non-assigned maxima. Thus, at the point where the non-assigned maxima start to prevail above the assigned ones, no more assigned atoms are accepted.

results of the analysis roughly reflect the visual impression about the quality of the analyzed densities and the ability of a human eye to read out the information from the density. For the present purposes, the exact number of located atoms is not as relevant, as the comparison between the results and the presented method suits the purpose of comparison very well.

The main advantage of the evaluation in direct space is that it is directly related to the information content of the electron density. Moreover, unlike  $w\Delta^2(\varphi)$ , which is calculated only from the experimental reflections, the electron density from the calculation including the extrapolated reflections contains also information from these reflections. On the other hand, the distinction between the converged and unconverged calculation is not so obvious in the number of located atoms, since the method can indicate a certain number of ‘located’ atoms even in a completely random density.

#### 4. Results and discussion

The method was tested on many data sets and many different values of the variable parameter for each type of omitted reflections and each structure. Here only a representative subset of the results is presented. In particular, the cone-shaped regions depend on the opening angle of the cone and the direction of the cone axis. We have performed calculations with cone axes along the directions [100], [010] and [001] for each structure. The calculations showed that the results were qualitatively similar for various directions of the cone axis, and therefore only the calculation with cone axis along [001] is presented here. The results are summarized in Fig. 1. The plot shows an

average number of located atoms for the 25 calculations performed on each data set, and the average  $w\Delta^2(\varphi)$ , both quantities for a calculation without and with the completion of the data set by the MEM. In addition to the average number of located atoms the maximum number of located atoms for each set is indicated. This is because for difficult data sets not every calculation converges successfully, and in practice it is possible to perform several structure solution attempts and pick the best one. Thus the maximum number of located atoms among several runs is also an interesting quantity.

The first and main impression from Fig. 1 is that the MEM-enhanced data sets lead in all cases to an improvement of the performance of charge flipping, and this improvement is often substantial. A more detailed comparison among the results for each structure shows that the improvement is most remarkable for missing low-angle reflections, randomly omitted reflections and missing reflections outside a cone. The improvement is smaller for missing high-angle reflections (*i.e.* limited resolution of data) and for missing reflections inside a cone. Only in the case of missing high-angle reflections in br1121 and gs1072 can the improvement be considered insignificant.

The behavior just described can be readily understood if we recall the basic properties of the MEM and the underlying assumptions of charge flipping. Charge flipping works best on a density that is nearly zero in large regions of the unit cell and non-zero only at a few places. In short, the density has to be sparse. Two effects related to the data completeness can influence this property. First, the large termination ripples in a Fourier map calculated from the data with missing reflections inside the resolution sphere break the sparseness of the density by generating false significant maxima and minima. Second, a data set with missing high-angle reflections causes the corresponding Fourier map to be less peaked and more smeared. As a result the proportion of density that is almost zero decreases and the sparseness condition is less well fulfilled. The severity of the impact of the termination ripples and density smearing on the efficiency of charge flipping depends on the sharpness and height of the atomic maxima. Therefore, charge flipping is more efficient on inorganic than organic structures.

The effect of the MEM exploited in this work is primarily the smoothing of the Patterson map in order to remove the termination ripples. It is thus clear that the improvement should be best in cases when the termination ripples are the largest, but where significant structural information is still preserved. This is the case for the randomly omitted reflections and also for the missing reflections at low angles. In contrast, the density map calculated from a complete data set with low resolution is already quite smooth and thus further smoothing by the MEM does not lead to any significant improvement. The observed marginal improvement is caused by estimating the missing high-angle reflections from the requirement on the positivity of the Patterson map.

The cone-shaped omitted regions represent a combination of the two effects mentioned above: they induce both the termination ripples and smearing of the density along certain directions. The missing reflections outside a cone lead to a deformation of the atomic maxima into discs with short axis along the cone axis. In an extreme case of reflections present only along the cone axis, the density corresponds to the projection of the structure onto that axis. The case of reflections missing inside the cone leads analogically to a smearing of the density along the cone axes and ultimately to a density corresponding to the projection of the structure onto the plane perpendicular to the cone axis. The latter case is thus a two-dimensional projection, while the former case is one-dimensional.

The two-dimensional projection preserves more low-density regions than the more degenerate one-dimensional projection. This difference is responsible for decreased performance of the basic charge-flipping algorithm for missing reflections outside a cone compared with missing reflections inside a cone.

The exact limits on the applicability of charge flipping on incomplete data depend strongly on the nature of the structure, as can be seen from comparison of the plots of different structures. Despite this several general statements can be deduced from the plots.

(i) The improvement of the data with low resolution is marginal, but it can make or break the solution in cases of data resolution just below the limit accessible for the original charge-flipping algorithm. The resolution limit between  $\sin \theta / \lambda_{\max} = 0.35$  and 0.50 is frequently encountered in structures that crystallize poorly or do not diffract at high angles.

(ii) A problem of missing low-angle reflections can be overcome quite efficiently with the new method, as long as the percentage of the missing reflections remains below *ca* 30%.

(iii) Up to about 50% of randomly missing reflections can be tolerated.

(iv) The opening half-angle of a cone of missing reflections can be up to 55–65°, which amounts to roughly 50% coverage of reciprocal space. For missing reflections outside a cone the limit is also about 60° or 50% coverage.

(v) All the limits just given apply to the two organic structures tested in the present study. For the inorganic structures these limits can be usually relaxed by 10 or even 20% in terms of the proportion of missing reflections. On the other hand, for more complex structures the limits are likely to be more severe.

It is interesting to compare the reflection intensities extrapolated by the MEM with the experimental values. The match between the extrapolated amplitudes  $|F_{\text{ext}}|$  and the experimental amplitudes  $|F_{\text{obs}}|$  of the structure factors can be evaluated by the conventional  $R_F$  value:

$$R_F = \frac{\sum ||F_{\text{ext}}| - |F_{\text{obs}}||}{\sum |F_{\text{obs}}|}.$$

The  $R$  values strongly differ among the structures and they depend also on the amount and type of missing reflections. For the structures br1121 and gs1072, a typical  $R$  value is between 30 and 40%, and exceeds 60% only in the case of a large number of missing reflections.  $R$  values around 35% illustrate that in most cases the MEM extrapolation performs quite well. The case of sk1293 is quite different. The lowest  $R$  value among all data sets is 57.6% and most  $R$  values range between 60 and 70%. This effect can be attributed neither to the difference in data resolution ( $\sin \theta / \lambda_{\max} = 0.6$  for sk1293 and 0.83 for pg1072) nor to the slightly higher overall Debye–Waller parameter for sk1293 ( $B_{\text{iso}} \simeq 2.5 \text{ \AA}^2$  versus  $1.1 \text{ \AA}^2$  for gs1072). This has been shown by a few tests on simulated data of sk1293, which yielded similarly high  $R$  values independently of  $B_{\text{iso}}$  or resolution. The most likely explanation is thus that the higher  $R$  values for sk1293 are caused by the much larger

unit-cell volume (see Table 1) and consequently larger number of reflections. It is, however, interesting that despite the much worse overall accuracy of the extrapolation, the improvement of the performance of charge flipping is still remarkable and does not differ significantly from the other two structures. This confirms the assumption that it is the elimination of the termination ripples rather than the accurate prediction of amplitudes that causes the better performance of charge flipping.

The enhancement of the incomplete data set by the MEM is, naturally, not limited to the combination with charge flipping. Other structure solution methods could profit from it as well. However, the problem of incomplete data is not so severe in direct methods, because they are not primarily Fourier methods, and thus do not directly suffer from the truncation effects. Moreover, direct methods more than charge flipping rely on accurate reflection intensities, and these are not reconstructed very well by the MEM extrapolation. The possible contribution of the data enhancement by the MEM to the structure solution methods other than density modification methods remains open for discussion.

### 5. Conclusions

The combination of the maximum entropy method with charge flipping overcomes the main shortcoming of the charge-flipping algorithm, namely the failure when applied to an incomplete data set. We have shown that if the data set is completed with reflection intensities calculated from the MEM-optimized Patterson map, charge flipping can successfully reconstruct structures from diffraction data with as little as 50% coverage of the reciprocal space. The improvement compared with charge flipping on the incomplete data set is most significant if the missing reflections are randomly distributed between the present reflections, and least significant if the data set has limited angular resolution. This is related to the number of the Fourier artifacts generated by the different types of data incompleteness. The more Fourier artifacts are present in the Patterson map of the incomplete data, the better is the relative enhancement of such a map by the MEM.

There is no principal difference between the structure solution of periodic structures and modulated structures or quasicrystals by charge flipping. Therefore it is almost certain that the conclusions made in this work will qualitatively apply also to aperiodic structures, where charge flipping turns out to be particularly useful.

The contribution of the Swiss National Science Foundation, grant No. 20-105325, is gratefully acknowledged.

### References

- Baerlocher, C., McCusker, L. & Palatinus, L. (2007). *Z. Kristallogr.* **222**, 47–53.
- Bricogne, G. (1993). *Acta Cryst.* **D49**, 37–60.
- Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.
- Buck, B. & Macaulay, V. A. (1991). Editors. *Maximum Entropy in Action*. Oxford: Clarendon Press.
- Dinnebier, R. E., Schneider, M., Van Smaalen, S., Olbrich, F. & Behrens, U. (1999). *Acta Cryst.* **B55**, 35–44.
- Gilmore, C. J. (1996). *Acta Cryst.* **A52**, 561–589.
- Gilmore, C. J., Dong, W. & Bricogne, G. (1999). *Acta Cryst.* **A55**, 70–83.
- Guillot, B., Jelsch, C. & Lecomte, C. (2000). *Acta Cryst.* **C56**, 726–728.
- Haibach, T. & Steurer, W. (1996). *Acta Cryst.* **A52**, 277–286.
- Iversen, B. B. & Larsen, F. K. (1995). *Acta Cryst.* **B51**, 580–591.
- Katrych, S., Weber, T., Kobas, M., Massüger, L., Palatinus, L., Chapuis, G. & Steurer, W. (2007). *J. Alloys Compd.* **428**, 164–172.
- Linden, W. von der, Dose, V., Fisher, R. & Preuss, R. (1999). Editors. *Maximum Entropy and Bayesian Methods*. Dordrecht: Kluwer.
- Mironov, Y. V., Cody, J. A. & Ibers, J. A. (1996). *Acta Cryst.* **C52**, 281–283.
- Oszlányi, G. & Sütö, A. (2004). *Acta Cryst.* **A60**, 134–141.
- Oszlányi, G. & Sütö, A. (2005). *Acta Cryst.* **A61**, 147–152.
- Oszlányi, G., Sütö, A., Czugler, M. & Párkányi, L. (2006). *J. Am. Chem. Soc.* **128**, 8392–8393.
- Palatinus, L. (2004). *Acta Cryst.* **A60**, 604–610.
- Palatinus, L. & Chapuis, G. (2006). *Superflip*. EPFL Lausanne, Switzerland. (<http://superspace.epfl.ch/superflip>.)
- Palatinus, L., Dušek, M., Glaum, R. & El Bali, B. (2006). *Acta Cryst.* **B62**, 556–566.
- Papoular, R. J., Collin, G., Colson, D. & Viallet, V. (2002). *Proceedings of the 21st Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by B. Fry. Melville, NY: American Institute of Physics.
- Papoular, R. J., Vekhter, Y. & Coppens, P. (1996). *Acta Cryst.* **A52**, 397–407.
- Roversi, P., Irwin, J. J. & Bricogne, G. (1998). *Acta Cryst.* **A54**, 971–996.
- Sakata, M. & Sato, M. (1990). *Acta Cryst.* **A46**, 263–270.
- Shannon, C. E. (1948). *Bell Sys. Tech. J.* **27**, 379–423, 623–656.
- Sivia, D. S. (1996). *Data Analysis – A Bayesian Tutorial*. Oxford: Clarendon Press.
- Smaalen, S. van, Palatinus, L. & Schneider, M. (2003). *Acta Cryst.* **A59**, 459–469.
- Steurer, W., Haibach, T., Zhang, B., Kek, S. & Lück, R. (1993). *Acta Cryst.* **B49**, 661–675.
- Wang, C.-R., Tsutomu, K., Tomiyama, T., Yoshida, T., Kobayashi, Y., Nishibori, E., Takata, M., Sakata, M. & Shinohara, H. (2001). *Angew. Chem. Int. Ed.* **40**(2), 397–399.
- Wheatley, P. S., Lough, A. J., Ferguson, G. & Glidewell, C. (1999). *Acta Cryst.* **C55**, 1489–1492.
- Wu, J. S., Leinenweber, K., Spence, J. C. H. & O’Keeffe, M. (2006). *Nat. Mater.* **5**, 647–652.
- Wu, J. S., Spence, J. C. H., O’Keeffe, M. & Groy, T. L. (2004). *Acta Cryst.* **A60**, 326–330.
- Zuñiga, F. J., Palatinus, L., Cabildo, P., Claramunt, R. M. & Elguero, J. (2006). *Z. Kristallogr.* **221**, 281–287.