Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 25 June 2007
Accepted 13 November 2007

Solving the crystal structures of zeolites using electron diffraction data. I. The use of potential-density histograms

Christopher J. Gilmore,[a]* Wei Dong[a]‡ and Douglas L. Dorset[b]

[a]WestCHEM, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, and [b]Advanced Characterization, ExxonMobil Research and Engineering Co., 1545 Route 22 East, Annandale, NJ 08801, USA. Correspondence e-mail: chris@chem.gla.ac.uk

The maximum-entropy and likelihood method for solving zeolite crystal structures from electron diffraction data is modified to use potential-map-density histograms as an additional figure of merit. The experimental histogram is compared to an idealized one (based on known zeolite structures) using Pearson and Spearman correlation coefficients. These supplement the use of log-likelihood estimates as figures of merit to select the optimal solution from a collection of phase sets. The method has been applied with success to seven zeolite and one inorganic crystal structures that have varying associated data quality. The technique works easily even with two-dimensional data sets of less than 50 unique diffraction data and a resolution of less than 2 Å. The method is very fast, and the computer time needed on a modest PC was never more than a few minutes.

© 2008 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

The difficulties of solving crystal structures from electron diffraction (ED) data are well known (Dorset, 1995) and equally well documented: the intensity data are subject to systematic errors arising from *n*-beam dynamical scattering, secondary scattering, sample bending and radiation damage. Nonetheless, the technique can yield structural information when all other diffraction methods fail. Solving some structures can be routine especially those from organic materials with well sampled data (Dorset, 2007) but, in general, the process is difficult as is the validation of the proposed model. Direct methods, maximum entropy, Patterson techniques, model building and the use of image data can all be employed (Dorset, 1995), and structures of high complexity can be solved and refined in favourable cases, including those of some inorganic materials (*e.g.* Mo *et al.*, 1992; Zandbergen *et al.*, 1997; Gjønnes *et al.*, 1998; Weirich, 2001; Gemmi *et al.*, 2003; Sinkler *et al.*, 2004). Even low-resolution protein data can be analysed in this way (Dorset, 1995), and it is an important tool in surface crystallography (*e.g.* Marks *et al.*, 1997). For a full discussion of the problems and the range of structures that have been solved, see Dorset (1995, 2005) and a review by Gilmore (2003).

Because of their predominant preference to form microcrystals, zeolites and similar microporous materials are commonly characterized by powder X-ray diffraction (PXRD) techniques, including various *ab initio* structure solution techniques especially *FOCUS* (Grosse-Kunstleve *et al.*, 1999), which uses the constraints of zeolite geometry in an active way in the phasing process. Often the nature of the material is such that peak overlap can make structure solution with powder diffraction difficult, even with synchrotron radiation, and it is then natural to turn to electron diffraction to provide useful single-crystal data from individual microcrystals. Best known (Cheetham, 1995) is the ability to map out a three-dimensional reciprocal lattice to facilitate unit-cell and space-group determination; less well known is the quantitative use of ED intensity data for *ab initio* structure determination.

The first application of electron diffraction to the determination of a zeolite structure *via* direct methods was made for MCM-22 by Nicolopoulos *et al.* (1995). This was soon followed by the three-dimensional solution of SSZ-48 (Wagner *et al.*, 1999) again by direct methods. Analyses of various zonal projections for MCM-22, MCM-49 (Dorset, 2003*c*), ZSM-5, ZSM-11 (Dorset, 2003*a*) and mordenite (Dorset, 2003*b*) were also carried out. These studies revealed that secondary scattering (Cowley *et al.*, 1951) can degrade the observed selected-area diffraction intensities to the extent that only partial structure determinations are possible. It is certainly true that there is no routine method for solving zeolites using any diffraction method except in very favourable cases and that it is exceptionally difficult with ED data.

Sometimes the related technique of high-resolution electron microscopy has been used as a supplement to PXRD, as in the solution of TNU-9 by Gramm *et al.* (2006) and IM-5 by Baerlocher, Gramm *et al.* (2007). In this case, crystallographic

‡ Present address: Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England.

phases from the image Fourier transform were used as constraints in the *FOCUS* program, but suitable image data can be very difficult to obtain. Many other examples of electron-microscope (EM) images used for zeolite structure analysis have been published (*e.g.* Leonowicz *et al.*, 1994; Liu *et al.*, 2001).

The maximum-entropy (ME) method has been used to study intermetallic compounds (Gjønnes *et al.*, 1998), organic materials (Dong *et al.*, 1992; Voigt-Martin *et al.*, 1994, 1995, 1997, 1999, 2000), membrane proteins (Gilmore *et al.*, 1996) and has also recently been applied to zeolites (Dorset *et al.*, 2005, 2007). In this paper, we first explain how the ME method is used with zeolite data (§2), then in §§3 and 4 we present an extension of the technique based on histogram matching, followed by a set of eight examples in §5. There is a brief discussion including the problems of validation in the final part of the paper.

2. The maximum-entropy method as applied to electron crystallography data

Recently, we (Dorset *et al.*, 2005) have developed a methodology based on earlier work by Bricogne (1984) and Gilmore (1996) for solving zeolites using electron diffraction data and the *MICE* computer program (Bricogne & Gilmore, 1990; Gilmore *et al.*, 1990; Gilmore & Bricogne, 1997). It works as follows.

1. The intensities of the measured reflections along with their estimated standard uncertainties are normalized using Wilson's method to give unitary structure factors $|U_h|^{\rm obs} = |E_h|^{\rm obs}/\sqrt{N}$ and the associated standard deviation $\sigma(|U_{\bf h}|^{\rm obs})$. N is the number of Si atoms in the unit cell. Where possible, no overall temperature factor is imposed on the normalization.

2. An origin is defined in the usual way (Rogers, 1980) by fixing the phases of several strong U values subject to certain constraints and a maximum-entropy map, $q^{\rm ME}({\bf x})$, is generated by maximizing the map entropy subject to the constraints of the phases and intensities of the origin-defining reflections. The phased reflections define a basis set $\{H\}$ whereas the remaining unphased measured ones form a set $\{K\}$. There is a third set of unobserved reflections {O} which arises from the missing cone or the use of zonal data only. The Fourier transform of $q^{\rm ME}({\bf x})$ generates estimates of amplitudes ($|U_{\bf h}^{\rm ME}|$) and phases ($\varphi_{\bf h}^{\rm ME}$) for reflections in $\{K\}$ and {O}. This process is called extrapolation.

3. A set of reflections is now chosen which optimally enlarges the second neighbourhood of the origin set. There are two types of reflections: acentric, which have no phase restrictions and have two degrees of freedom which are used to fix the quadrant (one for the real part and one for the imaginary part of the structure factor), and centric which are constrained to two possible values (*e.g.* 0, π) and have one degree of freedom that makes the binary phase choice. In this paper, all the space groups are centrosymmetric, so we are only dealing with the latter possibility; we choose a set of reflections with 16 degrees of freedom.

4. The phases of set $\{K\}$ are unknown and are therefore given permuted values; each combination of phases is used as a constraint in entropy maximization and defines a node. A full factorial design would require 2^{16} = 65536 phase combinations which, although feasible, poses problems with computer time and abstracting the correct phase set. For these reasons, we use error-correcting codes (Gilmore *et al.*, 1999), and in this particular case the Nordström–Robinson code which reduces the number of phase combinations to 256 with the caveat that the best solution will have a maximum of three phases incorrectly assigned.

5. Each node is subjected to constrained entropy maximization. As a figure of merit, log-likelihood gain is used (Bricogne, 1984; Gilmore *et al.*, 1990). For each centric extrapolated non-basis-set reflection ${\bf k}$, the likelihood measure, in its diagonal approximation, can be written as

$$\Lambda_{\bf k} = \frac{2|U_{\bf k}|^{\rm obs}}{\pi(2\varepsilon_{\bf k}\Sigma + \sigma_{\bf k}^2)} \exp\left\{-\tfrac{1}{2}\frac{(|U_{\bf k}|^{\rm obs})^2 + |U_{\bf k}^{\rm ME}|^2}{2\varepsilon_{\bf k}\Sigma + \sigma_{\bf k}^2}\right\} \times \cosh\left(\frac{|U_{\bf k}|^{\rm obs}|U_{\bf k}^{\rm ME}|}{2\varepsilon_{\bf k}\Sigma + \sigma_{\bf k}^2}\right), \quad (1)$$

where $\varepsilon_{\bf k}$ is the statistical weight of reflection ${\bf k}$, $\sigma_{\bf k}^2$ the variance of $|U_{\bf k}|^{\rm obs}$ and Σ a refinable measure of unit-cell contents, $\Sigma \approx 1/(2N)$. Note that this expression is a measure of agreement between $|U_{\bf k}|^{\rm obs}$ and $|U_{\bf k}^{\rm ME}|$, indeed it has a maximum where $|U_{\bf k}|^{\rm obs} = |U_{\bf k}^{\rm ME}|$. As in traditional likelihood analysis, a corresponding null hypothesis, $\Lambda_{\bf k}^0$, is defined for the situation of null extrapolation, $|U_{\bf k}^{\rm ME}| = 0$, which gives

$$\Lambda_{\bf k}^0 = \frac{2|U_{\bf k}|^{\rm obs}}{\pi(2\varepsilon_{\bf k}\Sigma + \sigma_{\bf k}^2)} \exp\left\{-\tfrac{1}{2}\frac{(|U_{\bf k}|^{\rm obs})^2}{2\varepsilon_{\bf k}\Sigma + \sigma_{\bf k}^2}\right\}. \quad (2)$$

Define

$$L_{\bf k} = \log\frac{\Lambda_{\bf k}}{\Lambda_{\bf k}^0}. \quad (3)$$

Then the global log-likelihood gain (LLG) is given by summing all the extrapolates:

$$\mathrm{LLG} = \sum_{\bf k} L_{\bf k}. \quad (4)$$

6. Rather than just choosing those phase sets with high associated LLG, which is a somewhat subjective process, the Student t-test is used in which the LLGs are analysed for phase indications (Shankland *et al.*, 1993). The simplest example involves the detection of the sign (corresponding, for example, to the 0, π) 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 + (*i.e.* 0 phase). The calculation is then repeated for those sets in which the same sign is $-$ (π) to give the corresponding μ^- and variance V^-. The t-statistic is then

$$t = \frac{|\mu^+ - \mu^-|}{\sqrt{V^+ + V^-}}. \quad (5)$$

The use of t-tests enables a sign choice to be derived with an associated significance level. This calculation is repeated for all the single-phase indications and is then extended to

combinations of two phases. Higher phase relationships cannot be used with this code because of aliasing errors or confounding. In general, when error-correcting codes are used for phase permutation only relationships with associated significance levels of <20% are used. Each of the m phase relationships so generated is given an associated weight w_i,

$$w_i = \left[1 - \frac{I_1(s_i)}{I_0(s_i)}\right], \qquad (6)$$

where I_1 and I_0 are the appropriate Bessel functions and s_i is the significance level of the ith relationship from the t-test. This weighting function reflects the need for a scheme in which the absolute values of the significance levels are not given undue emphasis since they are themselves subject to errors arising from the diagonal nature of the likelihood function used and the lack of error estimates for the LLGs themselves. Each node n is now given a score, s_n,

$$s_n = \mathrm{LLG}_n \sum_{j=1}^{m} w_j, \qquad (7)$$

where the summation spans *only* those phase relationships where there is agreement between the basis-set phases and the t-test derived phase relationships. The scores are sorted and only the top-ranked nodes are kept, the rest are discarded.

7. Potential centroid maps are used to examine the solutions. These maps use both the basis-set reflections and the extrapolates with the following Fourier coefficients:

$$|U_{\mathbf{k}}|^{\mathrm{obs}} \tanh(X_{\mathbf{k}}) \exp(i\varphi_{\mathbf{k}}^{\mathrm{ME}}), \qquad (8)$$

where

$$X_{\mathbf{k}} = (N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{\mathrm{obs}}|U_{\mathbf{k}}^{\mathrm{ME}}|. \qquad (9)$$

Ideally only the top eight maps are examined, but this procedure when applied to electron diffraction intensity data from zeolites (Dorset *et al.*, 2005, 2007) often requires the user to examine more than 30 maps. Because the structures are small (one is often only looking for less than 12 atoms), it is not necessary to build extensive phasing trees with new secondary sets of nodes being generated from highly ranked first-level phase sets.

The method has several advantages when dealing with zeolite ED data: it is robust with respect to experimental errors, which are modelled in the process and it is not reliant on data resolution. For example, it has also worked effectively with membrane protein electron diffraction data from *Halobacterium halobium* at 15 Å resolution (Gilmore, Shankland & Bricogne, 1993; Gilmore, Shankland & Fryer, 1993).

3. Histograms and ED data

Although successful (see, for example, Dorset *et al.*, 2005, 2007), the procedure raises some issues.

1. It is necessary to scan a large number of maps. In these circumstances, identifying a correct or near-correct solution can be difficult.

2. Is a usable solution present in the phase set that is not being considered?

The problem arises from difficulties with LLG estimates as figures of merit. The data are sparse with systematic errors especially involving kinematically weak reflections whose magnitudes are consistently overestimated. Although the LLG estimates are robust, the nature of ED data perturbed by multiple scattering weakens their power to discriminate phase sets.

Density histograms and histogram matching (Zhang & Main, 1990; Main, 1990) are now a routine tool in macromolecular crystallography for improving phases. Recent work by Baerlocher, McCusker & Palatinus (2007) has shown that histogram matching can also be a powerful technique for solving zeolite structures from powder diffraction data, and it is an obvious extension to these arguments to supplement the use of ME and LLG estimates with potential-map histograms from electron diffraction data.

To test the viability of this method for ED data, centroid potential maps were generated for several zeolite structures at different resolutions using experimental intensity data. The phases were taken from solved structures, the entropy maximized and the corresponding centroid map calculated. The use of real rather than idealized data is important, given the problems with secondary and dynamical scattering and the limited data resolution.

Small-area electron diffraction intensity data used in these analyses were mostly collected at 300 kV (but sometimes 200 kV) on FEI CM-30 or CM-20 electron microscopes (see Dorset *et al.*, 2005, 2007, for details). The instruments often employed precession geometry (Vincent & Midgley, 1994; Gemmi *et al.*, 2002) to minimize multiple scattering effects. The precession instrument used was a NanoMEGAS SpinningStar P010.

Patterns were recorded either on Kodak SO-163 electron-microscope films or on Fuji imaging plates. Intensity data were extracted from digitized records using the program *ELD* (Zou *et al.*, 1994). Another data set from a non-zeolite inorganic was from an electron diffraction camera as described by Voronova & Vainshtein (1958).

Typical results are shown in Fig. 1 for the zeolites ITQ-29 (LTA) (Corma *et al.*, 2004) and MWW (Camblor *et al.*, 1998). Both are two-dimensional data sets with 71 and 155 unique reflections, respectively. A frequently found trial solution for LTA involves a single large peak often at the cell corner or on a unit-cell edge. A typical map is shown in Fig. 1(*a*) with the corresponding histogram in Fig. 1(*b*). It can be seen that the histogram is very sharp. This compares with the correct map, shown in two dimensions, in Fig. 1(*c*) and its corresponding histogram in Fig. 1(*d*). This is broader with a much greater range of map pixel values.

For MWW, the histograms are less well defined. Fig. 1(*e*) shows the true density of the Si atoms with a model superimposed, and the corresponding histogram in Fig. 1(*f*). The latter is much broader than that of LTA.

Maps from data sets listed in Table 1 were used to generate a mean-density histogram which was smoothed using weighted box-car averaging and quadratic interpolation to give the histogram in Fig. 2, and this was used throughout the structure

Table 1
The data sets used in this study.

All but the last entry are zeolites. The international zeolite code is given for reference purposes where applicable. The data resolution refers to that reflection having the maximum resolution. (The effective resolution may be less than this.) All the reflections, including those with low intensity, were included. In column 4, the number of *T*-sites in the asymmetric unit is displayed.

Structure (IZA code in parentheses)	Space group	Unit-cell dimensions (Å)	No of *T*-sites in a.u.	Two- or three-dimensional data?	No. of unique reflections	Data resolution (Å)	Reference
ZSM-5 (MFI)	*Pnma*	a = 20.07, b = 19.92, c = 13.42	12	2	50	1.92	Kokotailo *et al.* (1978)
ZSM-10 (MOZ)	*P*6/*mmm*	a = 31.58, c = 7.525	6	2	29	2.73	Foster *et al.* (2005); Dorset (2006)
MCM-68 (MSE)	$P4_2/mnm$	a = 18.268, c = 20.208	8	2	42	1.83	Dorset *et al.* (2006)
MWW (MWW)	*P*6/*mmm*	a = 14.21, c = 24.94	8	3	155	1.31	Camblor *et al.* (1998)
Mordenite (MOR)	*Cmcm*	a = 18.11, b = 20.17, c = 7.53	4	2	27	1.97	Meier (1961)
ITQ-7 (ISV)	$P4_2/mmc$	a = 12.85, c = 25.1	5	2	32	1.56	Villaescusa *et al.* (1999)
ITQ-29 (LTA)	$Pm\bar{3}m$	a = 11.87	1	3	71	1.02	Corma *et al.* (2004)
$CuCl_2.3Cu(OH)_2$	*P*2/*m*	a = 5.73, b = 6.12, c = 5.64, β = 93.75°		3	120	0.73	Voronova & Vainshtein (1958)

analyses which are subsequently described as a reference. It is not symmetric about the mean but exhibits positive kurtosis with a tail at higher pixel densities. The illustrated histogram contains 100 bins.

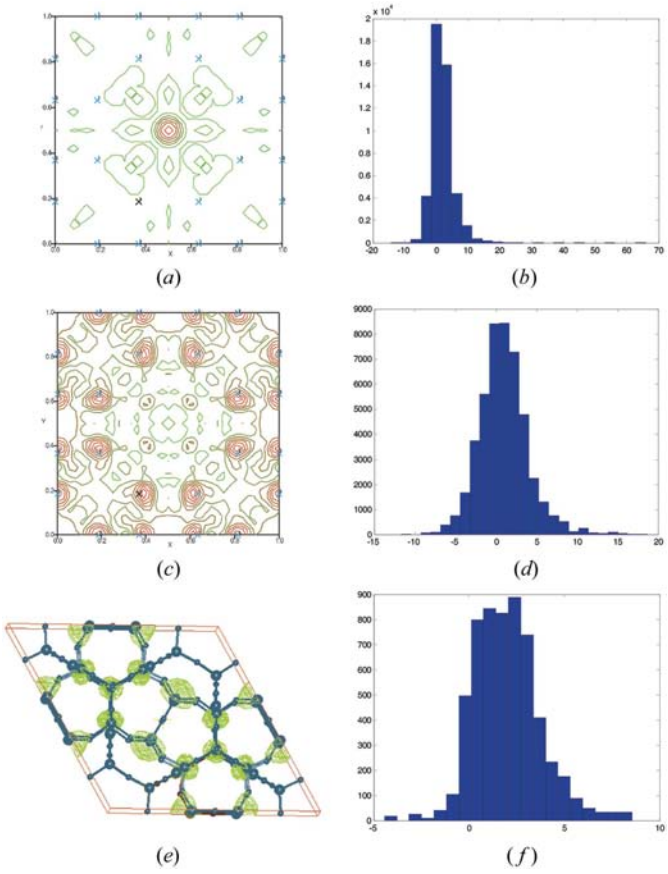

Figure 1
Representative histograms for two-dimensional diffraction data (ITQ-29). (*a*) False solution with density concentrated around (0.5, 0.5). (*b*) The corresponding potential-density histogram. (*c*) The true map based only on Si atoms. (*d*) The corresponding histogram. Representative histogram for three-dimensional data: (*e*) the true potential map for a sample three-dimensional data set for MWW, and (*f*) the corresponding histogram.

In principle, histograms are resolution dependent and will show variations between two- and three-dimensional maps, and one should modify them accordingly. However, we have found that the method we are using is robust to such changes and they are not necessary. Accordingly, we have used only the histogram in Fig. 2.

4. A revised method of structure solution

We now modify the ME formalism to use the histogram as follows.

1. Steps 1–7 of §2 are followed. For data normalization, only the Si atoms are included; the O atoms contribute less than 20% to the scattering and were never well indicated on the potential maps. For phase permutation, a Nordström–Robinson (16, 256, 6) code is used. This permutes the phases of 16 centric reflections to give 256 phase combinations, or nodes, as described in §2. Following the LLG analysis, the top 50 nodes are used to generate 50 centroid potential maps.

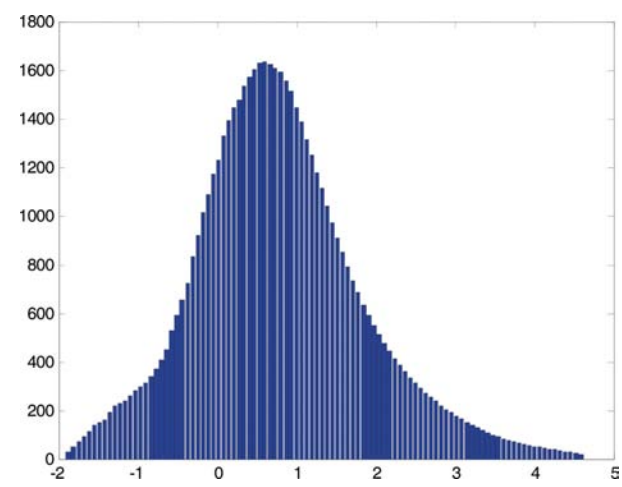

Figure 2
The histogram used for density matching for zeolites, and derived from averaging and smoothing several real ED data sets. It can be modelled by a tenth-order polynomial with coefficients 0.0002, −0.0025, 0.0127, −0.0103, −0.1014, 0.2174, 0.2820, −0.8292, −0.5706, 1.2076, 1.2296.

2. Each map generates the corresponding density histogram. Let x_i be the counts in bin i for the idealized histogram and y_i the counts in the corresponding experimental density from the ME-generated centroid maps. To compare the two, two correlation coefficients are used: the Pearson (r_P),

$$r_P = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2\right]^{1/2}}, \qquad (10)$$

and the Spearman, r_S,

$$r_S = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\left(\sum_{i=1}^{n} R(x_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{1/2}\left(\sum_{i=1}^{n} R(y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)^{1/2}}, \qquad (11)$$

where n is the number of bins (always set to 100 in this work). $R(x_i)$ and $R(y_i)$ are the ranks of the sorted histogram bins rather than their values.

The use of two correlation coefficients stems from our work in pattern matching using powder diffraction data (Barr *et al.*, 2004). The Spearman coefficient is less susceptible to outliers and the Pearson coefficient less susceptible to scaling issues, and we have found the use of these coefficients both separately and together to be optimal. Furthermore, the Spearman test does not assume that the data are normally distributed and this is clearly the case for the ideal histogram. This also helps to make the histogram independent of data resolution.

3. The following maps are examined:

(i) those with the five top scores from LLG analysis;

(ii) those with the five maximum values of r_P;

(iii) those with the five maximum values of r_S;

(iv) those with the five maximum values of $(r_P + r_S)/2$.

These four sets are not usually disjoint, and one often examines only approximately ten maps.

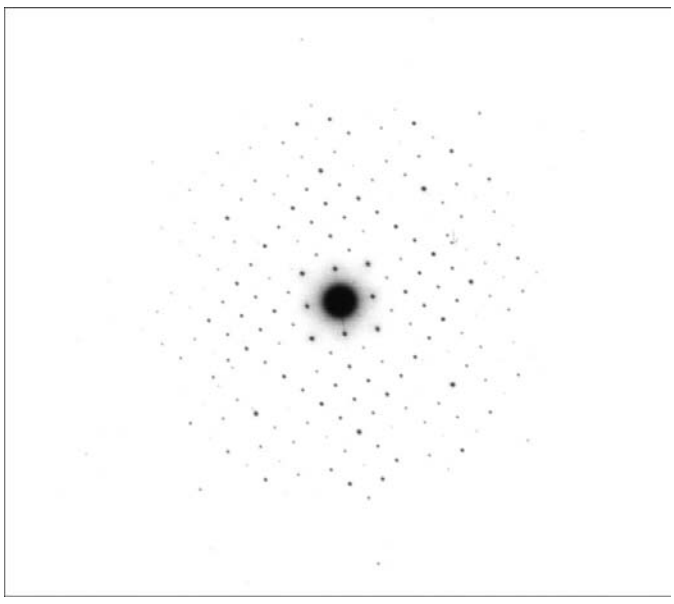

Figure 3
The $hk0$ zonal precession diffraction pattern for MCM-68.

These rules are wholly heuristic and derived from the experience of density matching but, as shown in the next sections, they have proved to be wholly effective in solving zeolite structures from ED data.

5. The test data

Eight data sets were used to verify this methodology and they are summarized in Table 1. The final entry in the table is not a zeolite but basic copper chloride, $CuCl_2 \cdot 3Cu(OH)_2$, using a data set collected by Voronova & Vainshtein (1958). This is included because it has a density histogram similar to that of zeolites and shows how the method we are using can be extended to inorganic materials. For the zeolites, all but two of the data sets are two dimensional with between 27 and 155 unique reflections, and a data resolution between 1.02 and 2.73 Å. All these structures are somewhat resistant to solution by traditional direct methods. For the examples considered, the zonal data give an optimal view of channel openings and hence are useful for structure determination. Data from tilted crystals are subject to the missing cone problem and crystal habit also constrains the use of tilt series, and the resulting maps show streaking which makes them hard to interpret. A typical diffraction pattern for MCM-68 is shown in Fig. 3. Uncertainty estimates are required for the intensity data and these were not available. To model the uncertainties, we used a simple measure that $\sigma(|F_{\mathbf{h}}|^{\rm obs}) = 0.1|F_{\mathbf{h}}|^{\rm obs}$.

6. Results

6.1. ZSM-5

This is a two-dimensional ($h0l$) data set obtained with precession geometry at 300 kV and recorded on imaging plates. At 1.92 Å resolution, there are only 50 unique reflections, somewhat restricted in terms of other ZSM-5 data sets. On the other hand, precession diffraction data were found earlier (Dorset *et al.*, 2007) to be somewhat anomalous in that it was actually easier to correctly solve a structure from 100 kV selected-area diffraction intensities (Dorset, 2003*a*) than from the 'corrected' higher-voltage data. This fact therefore presented a major challenge to this new analysis.

Data normalization gave an overall isotropic temperature factor, B, of 1.2 Å^2. The origin was defined by fixing the phases of the 403 and 303 reflections. These were selected automatically by the *MICE* computer program using an optimal neighbourhood enlargement algorithm (Gilmore *et al.*, 1990). The choice of 0 or π for these two phases is entirely arbitrary and each choice results in an origin shift. So that we can easily compare our maps to the published structures, the phases from the published models were used for origin definition. This in no way biases phasing, it is merely a device for easy map-model comparison. It should also be mentioned here that we do not quote mean phase errors or map correlation coefficients as an indicator of phase correctness in this paper. This is because they are poor indicators of map quality. Quite often a map with a good mean phase error is uninterpretable because

one or more key phases is wrong. Map correlation coefficients are confounded by one or more obviously spurious peaks. By far the best indicator of map quality is inspection.

Following entropy maximization and density matching, there were ten maps to examine. The solution that was top ranked from LLG analysis, and ranked third, second and third by Pearson, Spearman and mean-density histogram correlation coefficients, respectively, is shown in Fig. 4. The largest peaks all correspond to *T*-sites (*i.e.* Si atoms in the centres of O tetrahedra); Si(6), Si(2), Si(1) and Si(12) are not located although all the Si atoms overlap in pairs in the *b*-axis projection, so that effectively only two sites out of six are not found. There are two spurious peaks, but these are small and easily identified from geometry considerations. Other maps in the ten gave some of the framework, others had one large incorrect peak.

6.2. ZSM-10

This is a more difficult problem than ZSM-5 with only 29 unique reflections having a resolution of 2.73 Å. Two-dimensional *hk*0 precession data were collected at 300 kV and recorded on film. Data normalization gave a negative overall temperature factor, and this was reset to 5.0 Å^2. Given the space group in this projection, there are no origin reflections to be specified. Following entropy maximization on 256 nodes, the best resulting potential map (ranked eighth using LLG analysis, and second by the Pearson, Spearman and mean-density-histogram correlation coefficients) is shown in Fig. 5. There are no spurious peaks, but the site of Si(5) is not indicated and there is a tendency for a single peak to represent two sites; this is a consequence of the sparsity and resolution of the data. Thus, Si(6) is well defined, Si(5) is not indicated and the pairs Si(1)–Si(2), Si(3)–Si(4) each span a single peak. There were 12 maps to examine.

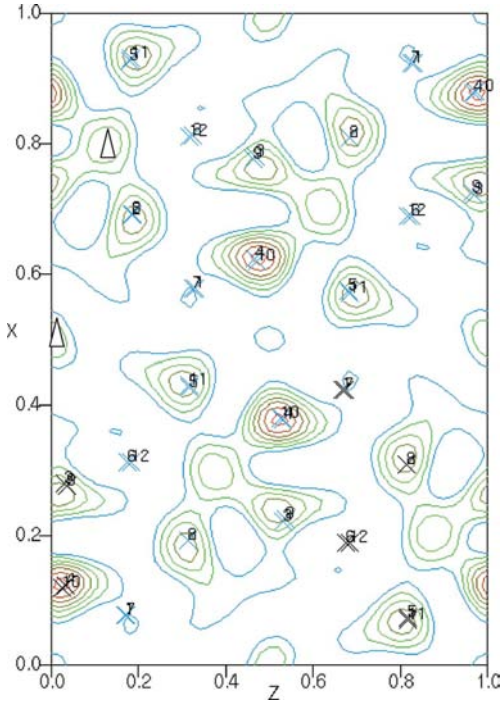

Figure 4
Structure solution for ZSM-5 using two-dimensional precession data. This potential map is the top-ranked one from LLG analysis, and ranked third, second and third by Pearson, Spearman and mean-density-histogram correlation coefficients, respectively. The crosses represent the true positions of the Si sites. The largest peaks all correspond to *T*-sites, although the positions of Si(6), Si(2), Si(1) and Si(12) are not indicated. There is sufficient information in the best map to begin the process of model building and structure completion despite the missing Si atoms. There are two unique spurious peaks marked with a triangle. In all the diagrams, a single cross is used to represent a *T*-site from the published structure. Black crosses define the asymmetric unit and blue crosses the sites generated by space-group symmetry. Since these are projection data, there is often site overlap, and for completeness this is shown by slightly offsetting the crosses when they overlap.

6.3. MCM-68

There are 42 unique precession electron diffraction data in two dimensions, collected at 300 kV with a resolution of 1.83 Å. They are two-dimensional, projected down the *c* axis. Normalization gave B = 0.75 Å^2. The origin for this structure required one phase to be fixed. Following the standard procedure, the best potential map from the solution ranked fourth using LLG analysis, and ranked fifth, fourth and fourth by the Pearson, Spearman and mean-density-histogram correlation coefficients, respectively, is shown in Fig. 6. There are two spurious peaks, but all the *T*-sites and the enclosed pore are indicated. There were ten maps to examine; all of them, except that shown in Fig. 6, were uninterpretable.

6.4. MWW

Three-dimensional data were obtained at 300 kV by tilting microcrystals around the a^* axis to find appropriate diffraction nets (Dorset *et al.*, 2005). There are 155 unique reflections. The origin was defined *via* a single reflection $52\bar{3}$ with U magnitude 0.16. Fig. 7 gives a view down both the *c* and *a* axes using the node ranked fourth in LLG analysis, and ranked 24th by the

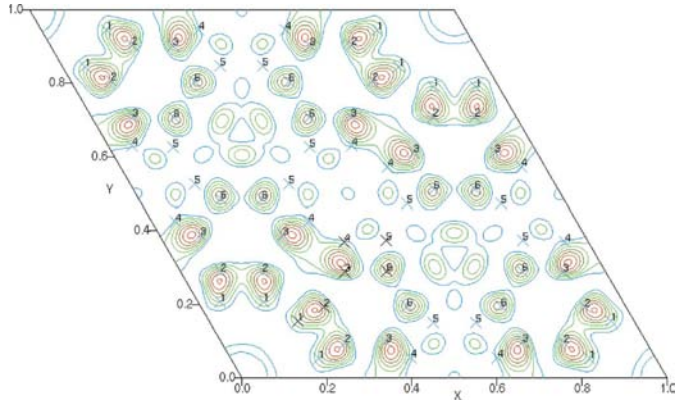

Figure 5
Structure solution for ZSM-10 using two-dimensional data. This potential map was ranked eighth using LLG analysis, and second by the Pearson, Spearman and mean-density-histogram correlation coefficients. There are no spurious peaks but the sites for Si(5) are not indicated and there is a tendency for a single peak to represent two sites; this is a consequence of the sparsity and resolution of the data.

Pearson, 35th by the Spearman and 29th using the mean-density-histogram correlation coefficients, respectively. The *c* projection is clear with all the *T*-sites indicated but with some spurious density near the origin and (0.2, 0.2, 0.0). The *a*-axis projection is much more poorly phased and very difficult to interpret owing to density elongation along the crystallographic *c* axis. This is a consequence of the undersampling provided by the limited tilt sequence, restricted to $\pm 60^\circ$. The histogram analysis performs poorly because of the data sampling and the consequential elongated shape of the potential density out of the *ab* plane. It is perhaps best to solve this structure initially in *c* projection (which is easy using this procedure) then extend the phasing into the third dimension, although we have not done this. We are currently exploring methods of improving maps of this kind.

6.5. Mordenite

The precession data, obtained at 300 kV, are two-dimensional with an effective resolution of 1.97 Å. Data normalization gave B = 11.1 Å^2. Although this is high, it is positive and the quality of the Wilson plot was acceptable, so this value was retained and used. The origin was defined using the phase of the 610 reflection having $|U| = 0.22$. Fig. 8(*a*) shows the top-ranked node using LLG analysis. The framework outline is clear although the atomic positions are not well resolved; the spurious peaks are small. This solution is ranked 28th using the Pearson correlation coefficient, third *via* the Spearman coefficient and 18th using the mean of the two. Fig. 8(*b*) shows the framework more clearly with better resolution of the atoms. There is a spurious peak at the centre of the pore but three out of four *T*-sites are clearly shown apart from Si(4). This map is poorly rated by LLG analysis (25th) and is ranked 11th using the Pearson coefficient, but fifth using Spearman coefficients and eighth *via* the mean correlation coefficient, so it is examined using the proposed procedure. *T*-sites are much better indicated than in the previous map, although there is a spurious peak in the centre of the pore.

6.6. ITQ-7

The data are two dimensional with an effective resolution of 1.56 Å. The precession data were collected at 200 kV on image plates (see Dorset *et al.*, 2007). Data normalization gave a negative B value and this was reset to 4.0 Å^2. The origin was defined using the phase of the 630 reflection having $|U| = 0.15$. Fig. 9(*a*) shows the top node as ranked by the mean correlation coefficient (41st ranked by LLG analysis, third using the Pearson coefficient and fifth *via* the Spearman coefficient). The map resolution is quite poor and the site of Si(3) is not well indicated although the pore size and position is quite

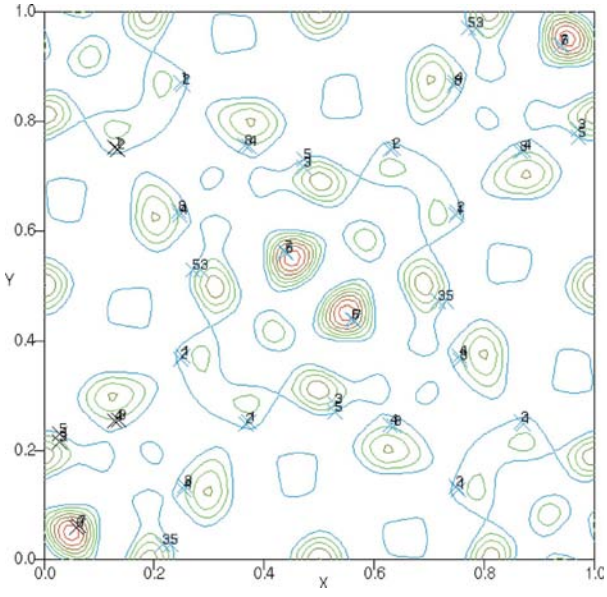

Figure 6
Structure solution for MCM-68 using two-dimensional data. This potential map was ranked fourth using LLG analysis, and fifth, fourth and fourth by the Pearson, Spearman and mean-density-histogram correlation coefficients, respectively. There are two spurious peaks, but all the *T*-sites are indicated.

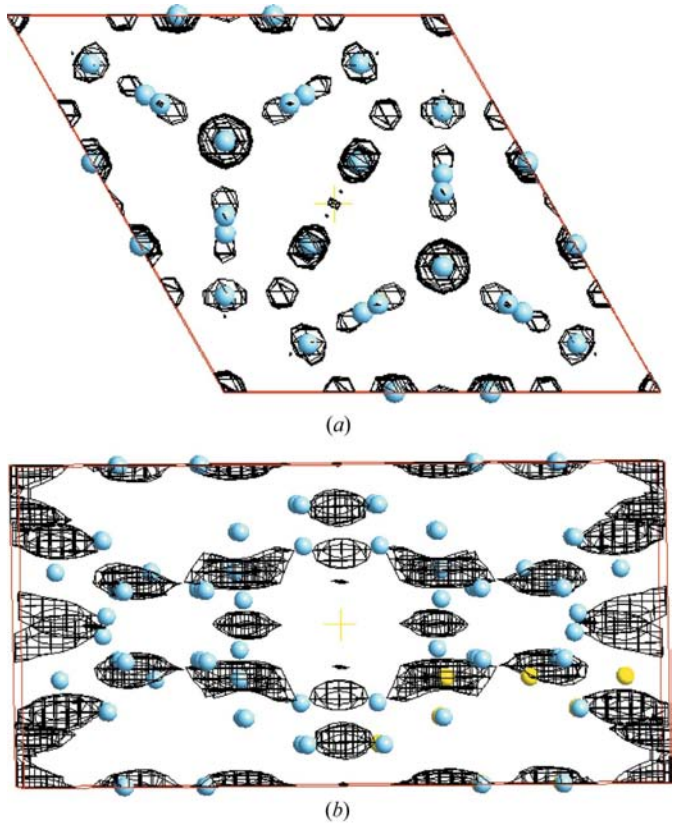

Figure 7
Potential map for the best solution for MWW. The data are three-dimensional data derived from specimen tilt along the a^* axis. (*a*) View down the *c* axis. (*b*) Looking down the *a* axis. This node is ranked fourth using LLG analysis, 24th by the Pearson, 35th by the Spearman and 29th using the mean-density-histogram correlation coefficients, respectively. The *c* projection is clear with density on all the *T*-sites indicated but with some spurious density near the origin. The *a* projection is much more poorly phased. This is a consequence of the undersampling resulting from the limited goniometric tilt sequence and explains the poor performance of density histograms as a figure of merit for this data set. The yellow atoms represent the asymmetric unit and the blue ones are symmetry equivalents.

clear. Fig. 9(*b*) shows the potential map corresponding to the top-ranked node using the Spearman correlation coefficient. This was ranked seventh *via* the Pearson, fifth ranked using the mean correlation coefficient and 42nd using LLG analysis. This shows the pore more clearly.

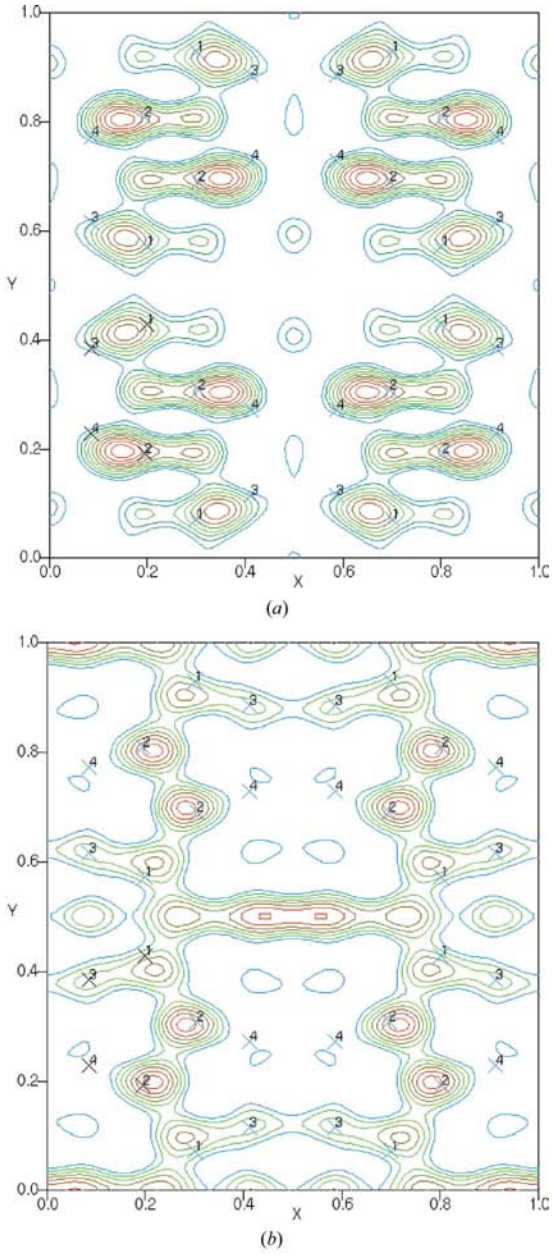

Figure 8
Maps for mordenite. (*a*) The top-ranked node using LLG. The framework outline is clear although the atomic positions are not resolved; the spurious peaks are small. This solution is ranked 28th using the Pearson correlation coefficient, third *via* the Spearman coefficient and 18th using the mean of the two. (*b*) An alternative solution which shows the framework more clearly with better resolution of the atoms. There is a spurious peak at the centre of the pore, but the *T*-sites are clearly shown apart from Si(4). This map is poorly rated by LLG analysis (25th) and is ranked 11th using the Pearson coefficient, but fifth using Spearman coefficients and eighth *via* the mean correlation coefficient.

6.7. ITQ-29 (LTA)

Precession (*hkk*) data were observed for ITQ-29 at 200 kV and recorded on imaging plates (see Dorset *et al.*, 2007, for details) with resolution of 1.02 Å. Because of the oblique projection in a cubic space group, a three-dimensional solution was found. Data normalization gave a negative *B* value and this was reset to 4.0 $Å^2$. The origin was defined by the 933 reflection with $|U| = 0.38$. The structure solution did not quite follow the usual pattern. This was clear from the low values of histogram correlation coefficients (all less than 0.5) and there were no interpretable maps. A simple measure of restricting

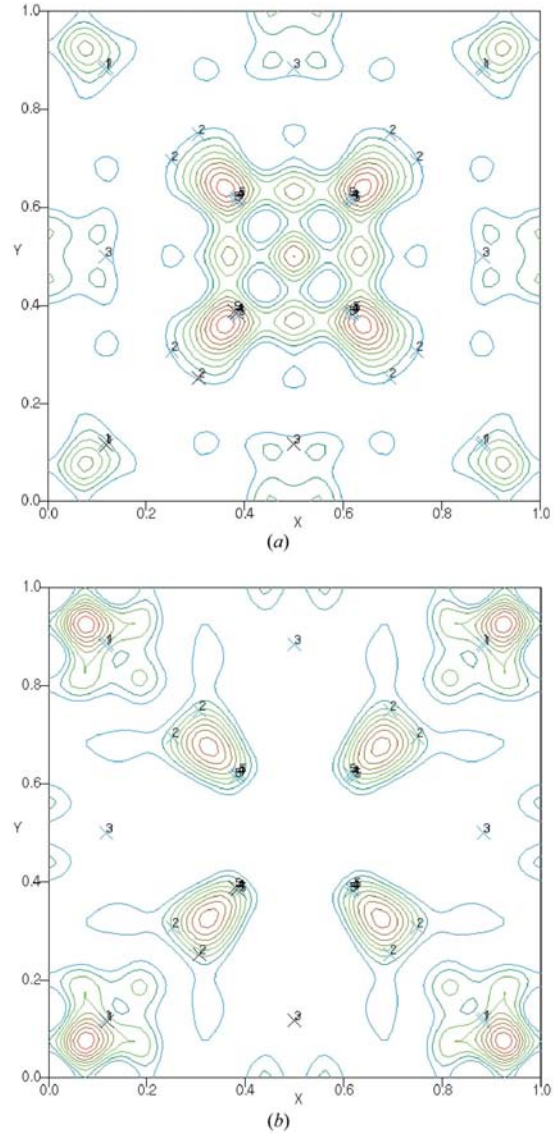

Figure 9
Potential maps for ITQ-7. (*a*) The best node as ranked by the mean correlation coefficient (41st ranked by LLG analysis, third using the Pearson coefficient and fifth *via* the Spearman coefficient). The resolution is quite poor, and Si(3) is not well indicated. (*b*) The top-ranked node using the Spearman correlation coefficient, seventh *via* the Pearson and the fifth ranked using the mean correlation coefficient. It was ranked 42 using LLG analysis. The framework is better defined, although the position of Si(3) is still not indicated.

the basis set to the top 20 U values (rather than leaving it unrestricted as in the other calculations) fixed this problem. Histogram correlation coefficients were now greater than 0.7 and the best solution as measured by LLG analysis produced a clear map with minimal noise that was easily interpreted as shown in Fig. 10. The solution was ranked seventh using the Pearson coefficient, sixth using the Spearman and sixth using the mean correlation coefficient.

6.8. $CuCl_2 \cdot 3Cu(OH)_2$

Although not a zeolite, this is an inorganic structure with an associated data set of high quality that should display an idealized histogram close to that of zeolites. The three-dimensional oblique texture data, collected on an electron diffraction camera (Voronova & Vainshtein, 1958), probably at 55–60 kV, are extensive: 120 three-dimensional reflections with a resolution of 0.73 Å. The large illuminated specimen area leading to this texture pattern produced data averaging similar to that found for precession devices in modern electron microscopes (Vincent & Midgley, 1994). The data had been processed using a Lorentz correction (Voronova & Vainshtein, 1958). The data quality had been indicated by a previous successful phase determination *via* symbolic addition (Dorset, 1994).

Data normalization gave a negative B value and this was reset to 4.0 $Å^2$. The origin was defined by three reflections, 221, 321 and 111, with U magnitudes of 0.51, 0.59 and 0.33, respectively. Several maps were interpretable; the best is shown in Fig. 11. This node was ranked fifth using the Spearman coefficient, but 24th by LLG analysis, 11th using the Pearson coefficient and seventh *via* the mean coefficient. There is very little noise and all the atoms except the Cl^- ions are easily located including the O atoms. Maps with missing heavy atoms are a common experience with ED data. [See, for example, Sinkler *et al.* (1998) and Sinkler & Marks (1999).]

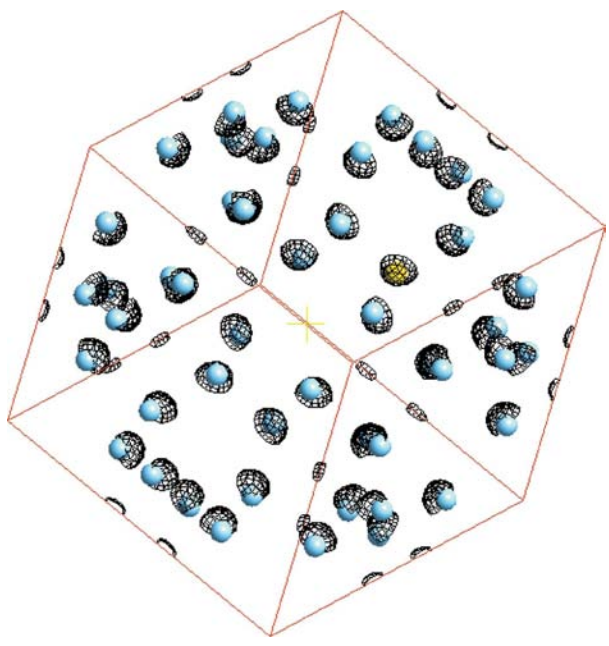

Figure 10
Potential map for ITQ-29. This map shows the framework clearly and was ranked seventh using the Pearson coefficient, sixth using the Spearman, and sixth using the mean coefficient. However, it was the top-ranked node from LLG analysis.

6.9. Density modification

It is, of course, possible to use histogram matching not just as a figure of merit but as a technique for density modification and, indeed, this is the main use of the method in macromolecular crystallography. To do this, pixels are modified as follows. The histograms are in n bins. Let ρ be the experimental histogram with bins ρ_i and let ρ' be the ideal density with bins ρ'_i. Every pixel, i, in the observed potential density map is replaced with

$$\text{pixel}_{\text{new}} = a \times \text{pixel}_{\text{old}} + b, \tag{12}$$

$$a = \frac{\rho'_{i+1} - \rho'_i}{\rho_{i+1} - \rho_i}, \tag{13}$$

$$b = \frac{\rho'_i \rho_{i+1} - \rho'_{i+1} \rho_i}{\rho_{i+1} - \rho_i}. \tag{14}$$

The method does not, in general, work well with ED data. It has a tendency to introduce small satellite peaks in clusters around large peaks and does not significantly improve the interpretability of the maps. A good example, where this problem does not arise, however, is shown for ZSM-5 in Fig. 12. The differences between the unmodified and modified maps are small but the latter is slightly cleaner and better defined. The site of Si(7) is better indicated. However, because of the problems described above, we have not used this technique routinely.

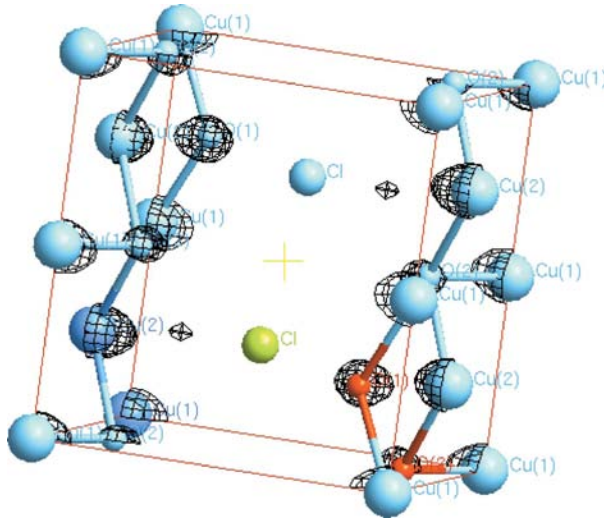

Figure 11
Potential map for $CuCl_2 \cdot 3Cu(OH)_2$. This was ranked fifth using the Spearman coefficient, but 24th by LLG analysis, 11th using the Pearson coefficient and seventh *via* the mean coefficient. There is very little noise and all the atoms except the Cl^- ions are easily located including the O atoms.

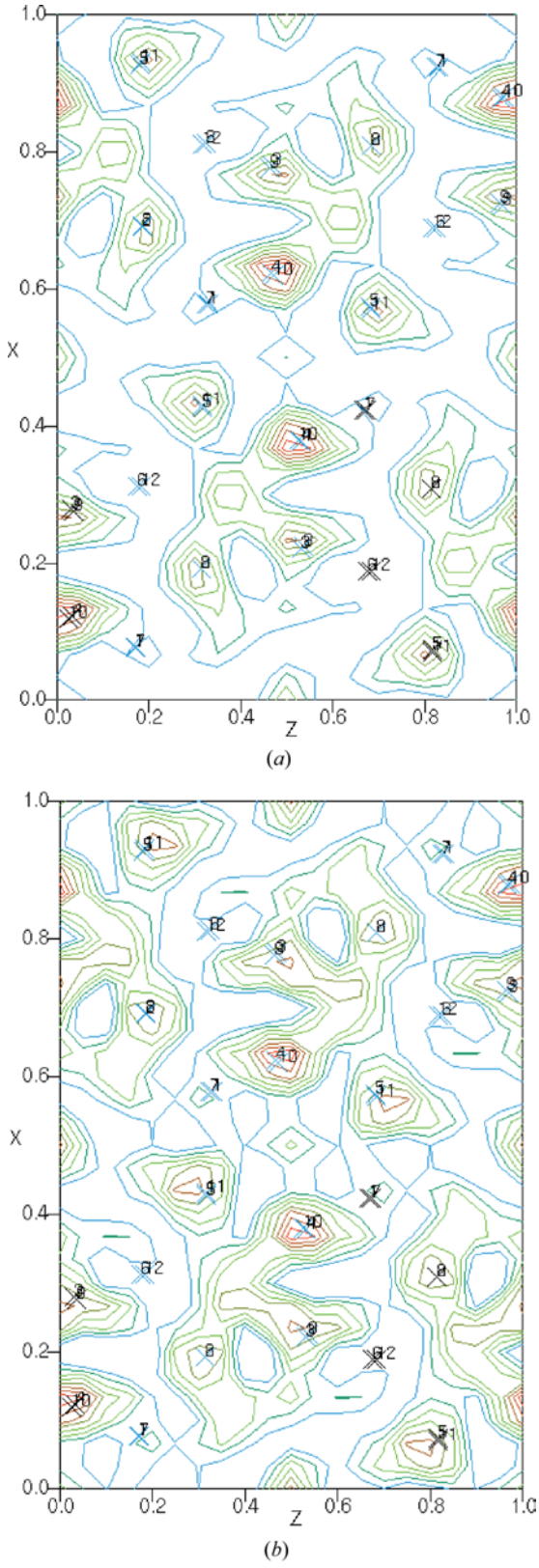

Figure 12
Density-matching calculations for ZSM-5. (*a*) The preferred solution before histogram matching and (*b*) after matching. The differences are small but the modified map is slightly cleaner and better defined. The resolution of these maps is less than that of the other examples because it has been chosen to prevent artefacts appearing during the matching process.

7. Discussion and conclusions

We have described a simple modification to the ME method for solving zeolite structures from ED data which works in a routine way and has been successful with eight structures of varying data quality and resolution. The technique works easily even with data sets of less than 50 unique diffraction data and a resolution less than 2 Å. It is also very fast: the computer time needed on a modest PC running at 2.8 GHz with 1 Gbyte of RAM and running Windows XP was never more than a few minutes, and for two-dimensional data was always less than 2 min.

There are, however, still problems to be addressed. In particular, the method does not give a unique solution, and it may still be necessary to scan up to 20 potential maps, some of which will be incorrect but still look plausible. A validation method is needed for these circumstances and a future paper in this series will explore the use of ME in a validation/structure completion environment.

From the analyses of electron diffraction data above, a further question arises: is LLG needed at all? Can the maps be assessed and sorted solely on the criteria of matching histograms? The short answer is 'no'; LLG is needed as an initial filter before histogram matching takes place. Histogram matching merely repositions the trial solutions so that the most meaningful phase set is often located among the first few in the ranking. On the other hand, there is a case (MWW) where LLG found the best solution among the top-ranked sets when the density histograms did not. This anomaly may be the consequence of the incomplete three-dimensional sampling of the diffraction intensity data (*i.e.* the 'missing cone') and the effect of the resultant density elongation along the *c* axis on the calculated histograms.

The general method could also be extended to organic molecules and even membrane proteins, although new idealized histograms will be needed. A pilot study on the halorhodopsin membrane structure at 6 Å resolution was promising (Havelka *et al.*, 1993; Gilmore *et al.*, 1996). Another form of density histogram matching in protein crystallography has also been successful for structure determination at low resolution (Dorset, 2000*a*,*b*,*c*).

Density-building functions will be discussed in the next paper; they provide an alternative way to build up zeolite frameworks in a ME environment and can also be used in conjunction with density-matching algorithms.

CJG acknowledges support from ExxonMobil Research and Engineering Co., and the University of Glasgow.

References

Baerlocher, Ch., Gramm, F., Massüger, L., McCusker, L. B., He, Z., Hovmöller, S. & Zou, X. (2007). *Science*, **315**, 1113–1116.
Baerlocher, Ch., McCuster, L. B. & Palatinus, L. (2007). *Z. Kristallogr.* **222**, 47–53.
Barr, G., Dong, W. & Gilmore, C. J. (2004). *J. Appl. Cryst.* **37**, 243–252.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* A**46**, 284–297.

Camblor, M. A., Corma, A., Diaz-Cabanas, M.-J. & Baerlocher, C. (1998). *J. Phys. Chem. B*, **102**, 44–51.

Cheetham, A. K. (1995). *The Rietveld Method*, edited by R. A. Young, pp. 276–292. Oxford University Press.

Corma, A., Rey, F., Rius, J., Sabater, M. J. & Valencia, S. (2004). *Nature (London)*, **431**, 287–290.

Cowley, J. M., Rees, A. L. G. & Spink, J. A. (1951). *Proc. Phys. Soc. London Sect. A*, **64**, 609–619.

Dong, W., Baird, T., Fryer, J. R., Gilmore, C. J., MacNicol, D. D., Bricogne, G., Smith, D. J., O'Keefe, M. A. & Hovmöller, S. (1992). *Nature (London)*, **355**, 605–609.

Dorset, D. L. (1994). *J. Chem. Crystallogr.* **24**, 219–224.

Dorset, D. L. (1995). *Structural Electron Crystallography.* New York: Plenum Press.

Dorset, D. L. (2000*a*). *Proc. Natl Acad. Sci. USA*, **97**, 3982–3986.

Dorset, D. L. (2000*b*). *Z. Kristallogr.* **215**, 265–271.

Dorset, D. L. (2000*c*). *Acta Cryst.* A**56**, 529–535.

Dorset, D. L. (2003*a*). *Z. Kristallogr.* **218**, 458–465.

Dorset, D. L. (2003*b*). *Z. Kristallogr.* **218**, 525–530.

Dorset, D. L. (2003*c*). *Z. Kristallogr.* **218**, 612–616.

Dorset, D. L. (2005). *Crystallography of the Polymethylene Chain.* Oxford University Press/IUCr.

Dorset, D. L. (2006). *Z. Kristallogr.* **221**, 260–265.

Dorset, D. L. (2007). *Ultramicroscopy*, **107**, 453–461.

Dorset, D. L., Gilmore, C. J., Jorda, J. L. & Nicolopoulos, S. (2007). *Ultramicroscopy*, **107**, 462–473.

Dorset, D. L., Roth, W. J. & Gilmore, C. J. (2005). *Acta Cryst.* A**61**, 516–527.

Dorset, D. L., Weston, S. C. & Dhingra, S. S. (2006). *J. Phys. Chem. B*, **110**, 2045–2050.

Foster, M. D., Treacy, M. M. J., Higgins, J. B., Rivin, I., Balkovsky, E. & Randall, K. H. (2005). *J. Appl. Cryst.* **38**, 1028–1030.

Gemmi, M., Calestrani, G. & Migliori, A. (2002). *Adv. Imag. Electron Phys.* **123**, 311–325.

Gemmi, M., Zou, X. D., Hovmöller, S., Migliori, A., Vennström, M. & Andersson, Y. (2003). *Acta Cryst.* A**59**, 117–126.

Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.

Gilmore, C. J. (2003). *Crystallogr. Rev.* **9**, 17–32.

Gilmore, C. J. & Bricogne, G. (1997). *Methods in Enzymology*, edited by C. W. Carter Jr & R. M. Sweet, pp. 65–78. New York: Academic Press.

Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *Acta Cryst.* A**46,** 297–308.

Gilmore, C., Dong, W. & Bricogne, G. (1999). *Acta Cryst.* A**55**, 70–83.

Gilmore, C. J., Nicholson, W. V. & Dorset, D. L. (1996). *Acta Cryst.* A**52**, 937–946.

Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Proc. R. Soc. London*, **442**, 97–111.

Gilmore, C. J., Shankland, K. & Fryer, J. (1993). *Ultramicroscopy*, **49**, 132–146.

Gjønnes, J., Hansen, V., Berg, B. S., Runde, P., Cheng, Y. F., Gjønnes, K., Dorset, D. L. & Gilmore, C. J. (1998). *Acta Cryst.* A**54**, 306–319.

Gramm, F., Baerlocher, Ch., McCusker, L. B., Warrender, S. J., Wright, P. A., Han, B., Hong, S. B., Zheng, L., Ohsuna, T. & Terasaki, O. (2006). *Nature (London)*, **444**, 79–81.

Grosse-Kunstleve, R. W., McCusker, L. B. & Baerlocher, C. (1999). *J. Appl. Cryst.* **32**, 536–542.

Havelka, W. A., Henderson, R., Heynmann, J. A. W. & Osterhelt, D. (1993). *J. Mol. Biol.* **234**, 837–846.

Kokotailo, G. T., Lawton, S. L., Olson, D. H. & Meier, W. M. (1978). *Nature (London)*, **272**, 437–438.

Leonowicz, M. E., Lawton, J. A., Lawton, S. L. & Rubin, M. K. (1994). *Science*, **264**, 1910–1913.

Liu, Z., Ohsuna, T., Terasaki, O., Camblor, M. A., Diaz-Cabanas, M. J. & Hiraga, K. (2001). *J. Am. Chem. Soc.* **123**, 5370–5371.

Main, P. (1990). *Acta Cryst.* A**46**, 507–509.

Marks, L. D., Plass, R. & Dorset, D. L. (1997). *Surf. Rev. Lett.* **4**, 1–8.

Meier, W. M. (1961). *Z. Kristallogr.* **115**, 439–450.

Mo, Y. D., Cheng, T. Z., Fan, H. F., Li, J. Q., Sha, B. D., Zheng, C. D., Li, F. H. & Zhao, Z. X. (1992). *Supercond. Sci. Technol.* **5**, 69–72.

Nicolopoulos, S., Gonzalez-Calbet, J. M., Vallet-Regi, M., Corma, A., Corell, C., Guil, J. M. & Perez-Pariente, J. (1995). *J. Am. Chem. Soc.* **117**, 8947–8956.

Rogers, D. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 23–92. New York: Plenum Press.

Shankland, K., Gilmore, C. J., Bricogne, G. & Hashizume, H. (1993). *Acta Cryst.* A**49,** 493–501.

Sinkler, W. & Marks, L. D. (1999). *Ultramicroscopy*, **75**, 251–268.

Sinkler, W., Marks, L. D., Edwards, D. D., Mason, T. O., Poeppelmeier, K. R., Hu, Z. & Jorgensen, J. D. (1998). *J. Solid State Chem.* **136**, 145–149.

Sinkler, W., Own, C. S. & Marks, L. D. (2004). *Microsc. Microanal.* **10** (Suppl. 2), 40–41.

Villaescusa, L. A., Barrett, P. A. & Camblor, M. A. (1999). *Angew. Chem. Int. Ed.* **38**, 1997–2000.

Vincent, R. & Midgley, P. A. (1994). *Ultramicroscopy*, **53**, 271–282.

Voigt-Martin, I. G., Gao Li, I. G., Kolb, U., Kothe, H., Yaminsky, A. V., Tenkovtsev, A. V. & Gilmore, C. J. (1999). *Phys. Rev. B*, **59**, 6722–6735.

Voigt-Martin, I. G., Kothe, H., Yaminsky, A. V., Tenkovtsev, A. V., Zandbergen, H., Jansen, J. & Gilmore, C. J. (2000). *Ultramicroscopy*, **83**, 33–59.

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.* A**51**, 849–868.

Voigt-Martin, I. G., Zhang, Z. H., Kolb, U. & Gilmore, C. J. (1997). *Ultramicroscopy*, **68**, 43–59.

Voronova, A. A. & Vainshtein, B. K. (1958). *Sov. Phys. Crystallogr.* **3**, 445–451.

Wagner, P., Terasaki, O., Ritsch, S., Nery, J. G., Zones, S. I., Davis, M. E. & Hiraga, K. (1999). *J. Phys. Chem. B*, **103**, 8245–8250.

Weirich, T. E. (2001). *Acta Cryst.* A**57**, 183–191.

Zandbergen, H. W., Andersen, C. J. & Jansen, J. (1997). *Science*, **277**, 1221–1225.

Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.

Zou, X. D., Sukharev, Yu. & Hovmöller, S. (1994). *Ultramicroscopy*, **53**, 271–282.