

## A Minimum-Entropy Algorithm for Surface Phasing Problems

L. D. MARKS\* AND E. LANDREE

Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA

(Received 24 March 1997; accepted 14 November 1997)

### Abstract

Methods of determining the phases for noisy and incomplete surface diffraction data (X-ray or transmission electron) are investigated. It is found that, while conventional methods do not always work, a minimum-entropy method that uses the relative entropy is more effective in finding the correct solutions.

### 1. Introduction

Despite the large number of different surface probes that now exist, very few surface structures are known. This is because low-energy electron diffraction (LEED), transmission electron diffraction (TED) and surface X-ray diffraction all work well for refining the structure but determining a viable starting point is often trial and error. While scanning tunneling microscopy (STM), atomic force microscopy (AFM) and high-resolution transmission electron microscopy (HREM) do offer information on possible starting structures, each suffers limitations. In particular, STM and AFM are only sensitive to the top layer of atoms. In addition, STM probes the surface electron density, which is not necessarily representative of the underlying surface structure. While HREM is sensitive to surface relaxations that extend into the bulk and reveals actual atom positions (*e.g.* Bengu *et al.*, 1996), the signal-to-noise ratio is often small and difficult to analyze. In the bulk, direct methods are frequently used to generate an initial structure or parts of the structure. While direct methods are well developed for bulk materials, their application to surface structure determination has not been explored in any detail; until very recently, it was accepted that direct methods could not be used for surfaces, and TED or X-ray diffraction experiments relied upon Patterson-map analyses. For completeness, we are only dealing herein with surface diffraction data, not bulk TED data where dynamical effects need to be considered.

Surface diffraction data differs substantially from bulk data not only in the much higher noise levels but because the data sets are incomplete, missing critical reflections. It is useful to define the character of the experimental data and what one hopes to achieve. A set of experimental intensities in two dimensions from a diffraction pattern (X-ray or electron) have been

determined. These span 2–3 orders of magnitude, with errors of 5–10% in the strongest beams ranging to 100% or more in the weaker ones. For TED data, some of the smaller reflections may be missing (obscured by the transmitted beam and associated energy-loss electrons) and in both cases reflections coincident with the primitive two-dimensional surface mesh must be excluded from the data set since it is not possible in any simple fashion to separate the bulk and surface contributions. [One can extract information by making assumptions about the bulk scattering in some cases as discussed by Ferrer (1993) but this then introduces other complications such as the surface coverage.] Note that we cannot assume that the surface content of these unmeasurable reflections are weak; they can be and often are among the strongest surface reflections. At present, the measurement errors for TED data are a little smaller, offset by weak dynamical perturbations provided that strong bulk diffraction conditions are avoided. (This is true in an *R*-factor sense albeit not necessarily in a  $\chi^2$  sense.) In both cases, subsurface strains (Jayaram *et al.*, 1993) perturb the data at higher angles; the true phases will give positive/negative dipoles rather than atomic features. As a final point, the composition of the surface is very poorly determined. For a surface composed of material *A* on a substrate *B*, it may be possible to know moderately accurate ( $\pm 20\%$ ) values for the number of *A* atoms but not the number of *B* atoms.

For a set of *N* diffraction intensities, we seek to assign phases. In some cases, a large number of these may be correctly determined, yielding a reconstruction (map) of the two-dimensional surface potential (for TED) or charge density (for X-rays). In other cases where the structure is either more complicated or there is a large percentage of unmeasurable beams, this may not be possible. Note that while the 'correct' map (true amplitudes and phases) would be positive everywhere if all reflections were measured, owing to the unmeasurable beams it will contain negative regions. One wants enough of the structure to be defined in real-space maps that Fourier methods can be used to complete it, more so than is normally the case. A 'good' map should indicate the positions to within a fraction of an ångström of 90% of the atoms and contain few erroneous peaks – the minimum that can probably be

tolerated is 50% of the atoms. Chemical arguments such as interatomic spacings do not work in projection and little prior knowledge about the cell contents can be assumed. Furthermore, since the data are two dimensional, both the initial phase determination and structure completion can be somewhat harder than in three dimensions. For instance, in three dimensions there are many more triplet phase relationships ( $\Sigma_2$ ) available than in two dimensions, weakening probabilistic methods. Indeed, for certain surfaces there are no triplets at all, for instance a  $2 \times 1$  surface (real-space  $A$  axis twice that of the primitive surface mesh) where only reflections  $h = 2n + 1$  are measured. A viable estimate of the phases needs to be among the 'best' solutions based upon some figure of merit (FOM). To put the above into context, we need to restore the phases for a complete data set with all reflections present to some high degree of accuracy, perhaps 5–10° r.m.s. error, otherwise we have no hope of finding the correct solution when strong reflections are absent as well as with measurement errors.

Before proceeding further, it is important to clarify some issues about the crystallography of surfaces and our notation herein. It is common in the surface science literature to describe surfaces in terms of the primitive two-dimensional surface mesh or ' $1 \times 1$ ' lattice including the surface normal (in parentheses) along with the bulk material. A reconstructed surface is then described as a multiplication (or matrix in the general case) with respect to this lattice; for instance, a  $3 \times 1$  cell means that the  $A$  axis of the surface is three times that of the  $1 \times 1$  lattice. Unfortunately, this leads to indexing of the surface reciprocal lattice using fractional order reflections and loses the symmetry information encoded in the two-dimensional space group. We will retain the multiplicative notation but use the more standard two-dimensional space groups and integer values for reciprocal-lattice points. We will also only deal herein with the two-dimensional problem normal to the surface; the full three-dimensional problem where the aperiodic information along the rods (surface truncation rods) normal to the surface has been partially measured is harder.

For simple structures where there is a large complete set of measured beams, the solution may be found by application of the Sayre equation (Landree *et al.*, 1997). Such a method has already been used in reliability tests for Au on Si(111)  $\sqrt{3} \times \sqrt{3}$  and  $5 \times 2$  surfaces (Marks *et al.*, 1997) and solved In on an Si(111)  $4 \times 1$  surface (Collazo-Davila *et al.*, 1997). Alternatively, one can use a branching structure, taking options for  $\Sigma_1$  relationships, assuming phases for some of the unmeasured reflections then using symbolic methods. With experimental electron diffraction data for the Si(111)  $7 \times 7$  surface, both maximum entropy and this technique work (Gilmore *et al.*, 1997). In this paper, we address the issue of optimizing an approach using unitary

structure factors in a Sayre-type sense, with no set phases beyond those for origin definition. We will show that a minimum relative entropy approach can be very successful for noisy and incomplete data.

## 2. Unitary Sayre equations

We will start with a somewhat unconventional description of certain direct methods. No new concepts will be introduced at first but our interpretation of the equations will be rather different.

Consider a complete set of structure factors in reciprocal space  $F(\mathbf{k})$ , where  $\mathbf{k}$  corresponds to a reciprocal lattice vector. The standard Sayre equation states that for non-overlapping identical atoms

$$F(\mathbf{k}) = c(\mathbf{k}) \sum_{\mathbf{h}} F(\mathbf{k} - \mathbf{h})F(\mathbf{h}), \quad (1)$$

where  $c(\mathbf{k})$  is a known function. This can be rewritten in terms of the unitary structure factors  $U(\mathbf{k})$  for  $N$  atoms as

$$U(\mathbf{k}) = N \sum_{\mathbf{h}} U(\mathbf{k} - \mathbf{h})U(\mathbf{h}) \quad (2)$$

or in a probabilistic fashion using the normalized structure factors  $E(\mathbf{k})$  as

$$E(\mathbf{k}) = \sum_{\mathbf{h}} E(\mathbf{k} - \mathbf{h})E(\mathbf{h}). \quad (3)$$

Given some initial phases, any of equations (1)–(3) can be used to generate new ones *via* the tangent formula (Karle & Hauptman, 1956) or modified to include statistical weights as in *MULTAN* (Germain & Woolfson, 1968). These new phases are then fed back in an iterative fashion and the validity of a set of phases determined by some combination of self-consistency FOMs. In effect, this is similar to a Picard iteration or successive approximation method, and equation (2) can be somewhat better written as

$$U_{n+1}(\mathbf{k}) = N \sum_{\mathbf{h}} U_n(\mathbf{k} - \mathbf{h})U_n(\mathbf{h}), \quad (4)$$

where  $U_n(\mathbf{k})$  contains the  $n$ th iteration phase estimate and at the end of each cycle the modulus is corrected to the known value. Equation (4) [or with  $E_n(\mathbf{k})$  values] is not at all unique; one can also write

$$U_{n+1}(\mathbf{k}) = N^2 \sum_{\mathbf{h}} \sum_{\mathbf{l}} U_n(\mathbf{k} - \mathbf{h} - \mathbf{l})U_n(\mathbf{h})U_n(\mathbf{l}), \quad (5)$$

in effect including positive quartet values (Schenk, 1973; Woolfson & Fan, 1995). To extend even further, instead of using convolutions in reciprocal space it is easier to use real space. Using the standard image processing and electron diffraction notation of lower case for real-space functions that correspond (Fourier transformation) to an upper-case reciprocal-space function, equations (4)–(5) can be generalized as

$$u_{n+1}(\mathbf{r}) = \hat{\mathbf{O}}(u_n(\mathbf{r})), \quad (6)$$

where  $\hat{\mathbf{O}}$  is some sharpening operator operating on a given real-space (potential) density map,  $u_n(\mathbf{r})$ . This method is not unlike other techniques such as the holographic method (Somoza *et al.*, 1995) or the *Shake-and-Bake* algorithm (Miller *et al.*, 1994), which utilize real-space density modifications and constraints to find the correct structure.

We want some 'optimum' sharpening operator that will enhance strong features in  $u_n(\mathbf{r})$  and suppress weak ones; *a priori*, there is no reason why a classic 'Sayre' form taking the square should be at all optimal. For instance, one could use  $u_n(\mathbf{r})^3$  or  $u_n(\mathbf{r})|u_n(\mathbf{r})|$ , both of which preserve negative regions. Borrowing from information theory (Shannon & Weaver, 1949), a completely flat  $u(\mathbf{r})$  has no information so we can consider

$$S_r(\mathbf{r}) = \begin{cases} u(\mathbf{r}) \ln[u(\mathbf{r})/e\langle u(\mathbf{r}) \rangle] & u(\mathbf{r}) > 0 \\ 0 & u(\mathbf{r}) < 0, \end{cases} \quad (7)$$

where  $S_r(\mathbf{r})$  is the 'relative entropy' as a function of the real-space vector  $\mathbf{r}$  (Kullback & Leibler, 1951; Cover & Thomas, 1991), also known as the Kullback-Leibler distance with  $\langle u_n(\mathbf{r}) \rangle$  the mean value. Note that we are considering negative  $u(\mathbf{r})$  as containing no useful information. This suggests a 'minimum-entropy' iterative algorithm similar to that described above of

$$u_{n+1}(\mathbf{r}) = \begin{cases} u_n(\mathbf{r}) \ln[u_n(\mathbf{r})/\langle u_n(\mathbf{r}) \rangle] & u_n(\mathbf{r}) > 0 \\ 0 & u_n(\mathbf{r}) < 0. \end{cases} \quad (8)$$

While equations (7) and (8) have similarities to well known maximum-entropy methods, they are different. First, conventional maximum-entropy methods divide  $u(\mathbf{r})$  by the sum over all  $\mathbf{r}$  to give probabilities, rather than referencing as we do here to the completely flat case. They then use the entropy as a regularizing functional based upon probability arguments; loosely speaking, they find the most disordered  $u(\mathbf{r})$  consistent with the experimental data (see Bricogne, 1984; Bricogne & Gilmore, 1990; Gilmore *et al.*, 1990, for further details). Here we are using equation (8) to find a self-consistent solution in a Sayre-like sense, *i.e.* using the relative entropy as a self-consistent sharpening operator. At a formal mathematical level, it may have no more fundamental validity than the Sayre equation and we will not claim anything except that both are exact relationships for equal non-overlapping atoms. That said, the solution of equation (7) minimizes the relative entropy, which is equivalent to maximizing the (more conventional) entropy (Cover & Thomas, 1991). We really have here a hybrid method that has components of both conventional direct methods (Picard iteration and tangent formula) and maximum-entropy methods. One feature that it does share with maximum entropy is high nonlinearities, which are important in

the sense that they will compensate for unmeasured reflections. For later reference, one can extend equation (8) to a form

$$u_{n+1}(\mathbf{r}) = \begin{cases} u_n(\mathbf{r})^x \ln[u_n(\mathbf{r})/\langle u_n(\mathbf{r}) \rangle C] & u_n(\mathbf{r}) > 0 \\ 0 & u_n(\mathbf{r}) < 0. \end{cases} \quad (9)$$

The constant  $C$  can either be treated as part of the prior information (in a maximum-entropy sense) or an adjustable baseline parameter (Hoare, 1985). Herein we will only use  $x = 1$  (but see *Discussion*) and  $C = 1$  unless otherwise stated.

So far we have assumed in equations (1)–(9) that a complete set of reflections is available. More realistically, only reflections out to a certain resolution are available and there may be significant holes in the data owing to the unmeasurable beams. This is akin to the standard problem in image processing where a simple power spectrum of an image gives long 'ringing' effects (streaking) in reciprocal space. We can handle this by introducing a 'window function'  $W(\mathbf{k})$  (*e.g.* Press *et al.*, 1992) defined only for the experimentally measured spots and  $W(\mathbf{0})$ , chosen to satisfy for any sharpening operator:

$$w(\mathbf{r}) = \alpha \hat{\mathbf{O}}(w(\mathbf{r})), \quad (10)$$

where  $\alpha$  is some constant, yielding windowed unitaries  $U'(\mathbf{u})$  given by

$$U'(\mathbf{k}) = W(\mathbf{k})U(\mathbf{k}). \quad (11)$$

It is simple to show that these windowed unitaries satisfy equation (6) subject to a multiplicative constant for non-overlapping identical atoms. We will hereafter use these windowed unitaries, dropping the prime superscript. [For reference,  $W(\mathbf{k})$  can normally be found by a simple updating scheme similar to equation (6) starting from a Gaussian form.]

Finally, some FOM is required. Since the errors will tend to be Gaussian, this suggests a figure of merit FOM:

$$\text{FOM}^m = \frac{\sum' |U_n(\mathbf{k}) - \beta U_{n+1}(\mathbf{k})|^m}{\sum' |U_n(\mathbf{k})|^m}, \quad (12)$$

where the sum,  $\sum'$ , taken over all reflections except  $\mathbf{k} = \mathbf{0}$  with  $\beta$  chosen to minimize the FOM, and the unitaries  $U_{n+1}(\mathbf{k})$  corrected at the end of each cycle to have the experimentally determined moduli (see Fig. 1). Since we are employing a Picard-like iteration rather than a true minimization, termination of the iteration was when the FOM either increased or became stationary. This FOM differs from the traditional definition of a residual by inclusion of both the amplitudes and the phases. While a form with  $m = 2$  is conventional, a robust variant with  $m = 1$  will accommodate more experimental and numerical errors, avoiding the FOM being controlled by reflections that are poorly

determined. In practice, it performed much better. Note that the above FOM uses the windowed unitary terms; removing the window for the FOM calculation was decidedly inferior. In many cases, we used the structure factors in equation (12), *i.e.* replacing  $U(\mathbf{k})$  by the corresponding  $F(\mathbf{k})$ . In practice, there was very little difference between the two since the window cancels most of the increase upon conversion from true structure factors to unitary ones, at least for inorganic compounds with reasonably small Debye–Waller terms. [A subtle issue here that in practice could be important is errors in the conversion from  $F(\mathbf{k})$  to  $U(\mathbf{k})$ .] For reference, we have not found any method of including measurement errors that works better than ignoring them.

A final issue that needs to be addressed is normalization. If the number of atoms is known, the standard normalization procedure can be used but this will rarely be the case. Empirically, from a number of calculated data sets, the approximation

$$\sum_{\mathbf{k}} |U(\mathbf{k})|^2 = A \quad (13)$$

(excluding the zero beam from the sum), with  $A = 1$  for TED data and  $A = \pi$  for X-rays works well, slightly overestimating the number of atoms. As an alternative,

one can run calculations for different assumptions about the number of atoms. In terms of the final method, it is highly desirable that it be insensitive to normalization. For reference, except for the last numerical model, we used the crude  $A = 1$  normalization for all the calculations; better normalizations improve the results but this requires assumptions that may not be valid.

### 3. Numerical methods

The primary algorithm that we used to govern the initial phases was a genetic (evolution) method that has been previously discussed (Landree *et al.*, 1997). In essence, this uses random initial phases similar to a *RANTAN* method (Yao, 1981), guiding the choice of phases based upon an evolution algorithm and the FOM. The main input were simulated TED data for surfaces and the atomic positions and other details are shown in Table 1. Model 1 was a complete data set to a resolution of 0.1 nm with two types of atoms, In and Si. Model 2 used the same coordinates, but with only Si atoms and excluded reflections that would be unmeasurable for a  $2 \times 4$  unit cell, *i.e.* the  $A$  and  $B$  axes 2 and 4 times, respectively, the primitive surface  $1 \times 1$  lattice dimensions, thereby excluding reflections  $2n, 4m$ , which include a number of the strongest reflections (see Table 2). Model 3 was the same as model 1 but also had missing reflections for a  $2 \times 4$  unit cell. In all of the above, Gaussian noise was added to the simulated data, representative of experimental conditions. How well the methods worked was gauged by a consistency term defined as

$$\text{CFOM} = \frac{\sum_{\mathbf{k}} |U(\mathbf{k})| (1 - \cos \theta)}{\sum_{\mathbf{k}} |U(\mathbf{k})|} \quad (14)$$

at the end of a given iteration cycle, with  $\theta$  the angle between the true and algorithm-determined phases. For reference, this is similar to  $1 - X$ , where  $X$  is the normalized cross-correlation coefficient, albeit a more robust form that will pay more attention to phase errors in the weaker reflections. A given method can be considered to work well if there is a strong correlation between small FOM and small CFOM values: if the CFOM is zero, a full restoration rather than an approximate map has been achieved; if the FOM is small but the CFOM large, the method has found an incorrect solution.

The final set (model 4) was experimental X-ray data for the Si(111)  $7 \times 7$  surface (Robinson *et al.*, 1986), a centrosymmetric  $p6mm$  cell. While this surface contains only one type of atom, in projection there are overlaps. Furthermore, the strongest beams for the surface lattice (3–6 times stronger than any others) coincide with bulk 220 reflections and are unmeasurable. We permuted all combinations of the strongest ten reflections and will

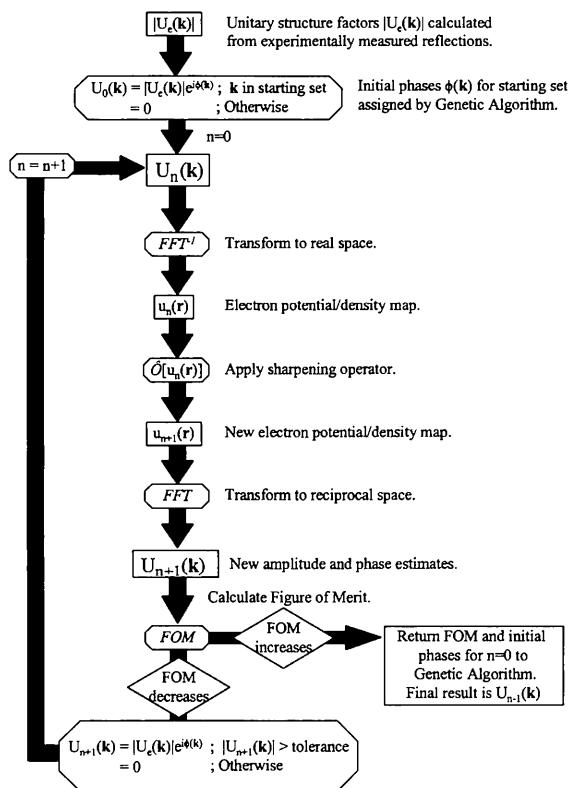


Fig. 1. Flow diagram for the methods discussed herein. The sharpening operator is completely general.

Table 1. Atom positions within the unit cell

$a = 7.68$ ,  $b = 15.36$  Å,  $\gamma = 90^\circ$ .

x	y
0.5†	0.28†
0.5†	0.72†
0.19	0.0
0.81	0.0
0.29	0.5
0.71	0.5
0.5	0.15
0.5	0.85
0.0	0.117
0.0	0.307
0.0	0.693
0.0	0.883

† In model 1 and model 3, these atoms are indium and the rest are silicon.

Table 2. The 30 strongest structure factors for models 1–3; model 1 is the same as model 3 with the missing reflections included

h	k	Models 1 and 2	Model 3
2	2	0.316569	0.369099
2	0	0.261807†	0.315425†
4	0	0.243135†	0.294317†
0	5	0.230172	0.272069
0	7	0.19207	0.231195
2	7	0.175555	0.209605
2	5	0.188602	0.209008
1	9	0.114184	0.189854
1	2	0.072682	0.182747
6	0	0.133009†	0.170699†
0	4	0.117841†	0.164648†
3	7	0.122018	0.159623
0	6	0.190422	0.152341
4	5	0.1464	0.147433
6	7	0.06872	0.130556
5	9	0.092332	0.124187
0	8	0.123191†	0.120740†
2	4	0.203030†	0.118838†
3	11	0.051	0.118047
1	4	0.063744	0.117468
4	7	0.074547	0.113024
0	10	0.135422	0.107284
4	2	0.076938	0.107225
2	12	0.084765†	0.1038†
0	1	0.104691	0.101061
3	0	0.015168	0.09692
0	9	0.026972	0.094193
1	7	0.027356	0.093642
0	2	0.012494	0.091613

† Missing reflection for model 2 and model 3.

show later how the true solution ranked among these in terms of the FOM.

These models are deceptively simple; even though they contain relatively few atoms there are far less triplet relationships than with a three-dimensional data set so classical probabilistic methods will have problems. A conventional direct-methods package (*MTHRIL*) had great difficulties even with model 1, a

point that we will return to in the discussion. There are also significant holes in the data for the last three models. To illustrate the latter, Table 2 lists the strongest reflections for models 2 and 3 with the unmeasurable ones marked.

## 4. Results

### 4.1. Model 1

For this model, the simple ‘unitary Sayre’ method [equation (2)] was not very efficient, while the minimum-entropy method worked very well. (We will return in the *Discussion* to one way that the Sayre method worked.) Shown in Fig. 2 are plots of the FOM versus the CFOM for the two. With the minimum-entropy method, the restoration is of sufficient quality that one is reconstructing the surface potential to a high degree of accuracy.

### 4.2. Model 2

For this model, the Sayre method was reasonable; two basic solutions were found, one of which is the correct one. While the solutions are not as good as for the minimum-entropy method and model 1, it would be

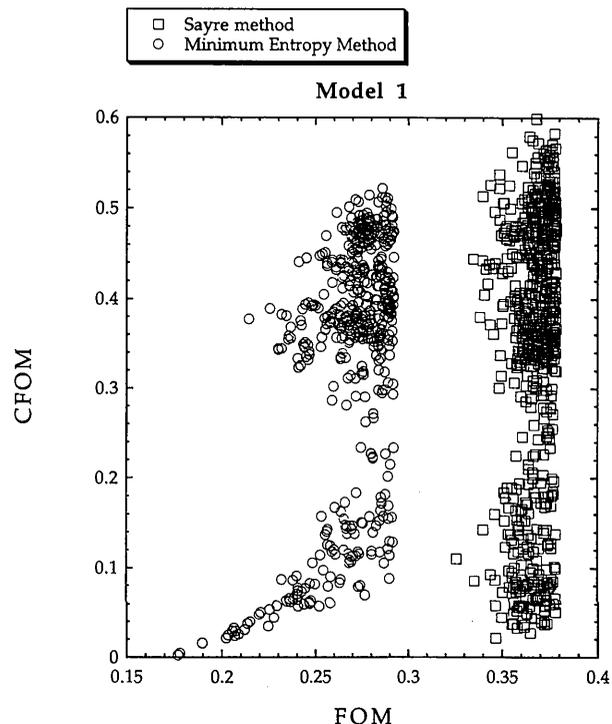


Fig. 2. Plot of the figure of merit (FOM) calculated using unitary structure factors and robust fitting [ $m = 1$ ; equation (6)] versus a ‘correctness’ factor (CFOM), which compares the calculated solution to the true solution [equation (8)] for model 1. The tail of values with low CFOM and FOM values indicates that the minimum-entropy method is achieving a full restoration of the phases.

Table 3. Position of the correct solution within the final list of generated solutions based on the estimated number of atoms within the unit cell

Cube:  $\hat{\mathbf{O}}(u(\mathbf{r})) = u(\mathbf{r})^3$ . Sayre:  $\hat{\mathbf{O}}(u(\mathbf{r})) = u(\mathbf{r})^2$ . Min Ent 1: [equation (5)]  $C = 1$ . Min Ent 0.5: [equation (5)]  $C = \exp(0.5)$ . S.F.: FOM calculated using robust form ( $m = 1$ ) with structure factors. Unitary: FOM calculated using robust form ( $m = 1$ ) with unitary structure factors [equation (6)].

S.F.	Cube	Sayre	Sayre	Min Ent 1	Min Ent 1	Min Ent 0.5	Min Ent 0.5
Unitary	•	•	•	•	•	•	•
No. of atoms							
50	25	100	132	5	5	7	5
60	19	103	111	4	4	6	5
70	33	60	87	4	4	3	3
80	34	69	86	3	3	1	1
90	18	67	83	7	4	2	2
100	7	70	82	6	3	4	2
110	7	22	41	7	2	2	2
120	6	21	35	1	1	2	1
135	2	69	98	1	1	2	2
150	2	86	102	1	2	3	2

simple to complete the solution. Similarly, the minimum-entropy method was adequate; both cases are shown in Fig. 3.

#### 4.3. Model 3

In this case, we were completely unable to find anything close to the correct solution using the Sayre method, no matter what variants were attempted in terms of the FOM. In contrast, the behavior of the

minimum-entropy method was adequate. Both sets of results are shown in Fig. 4, and we should note that the plot is a little too pessimistic in appearance; many different initial phases gave the same final phases so the probability of obtaining the correct solution was much higher than might appear to be the case. There were

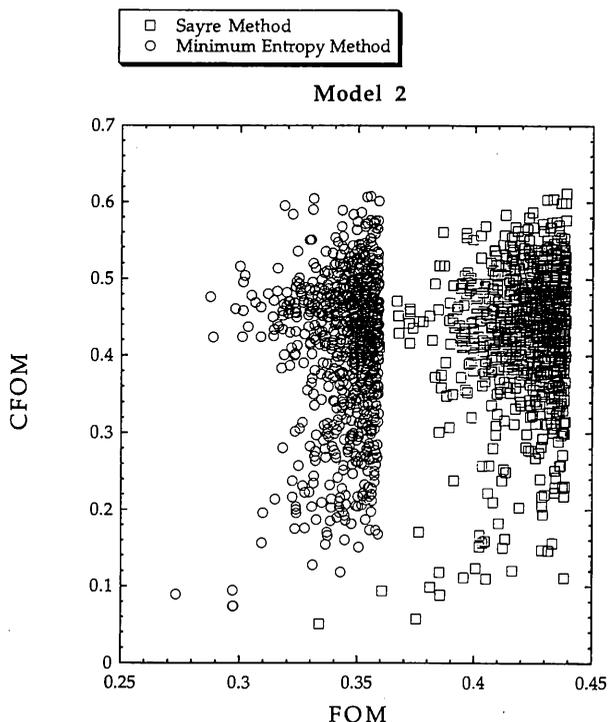


Fig. 3. Plot of the FOM versus CFOM for model 2. Both methods are finding reasonably good CFOM values ( $<0.1$ ) indicating correct assignment of the majority of the phases.

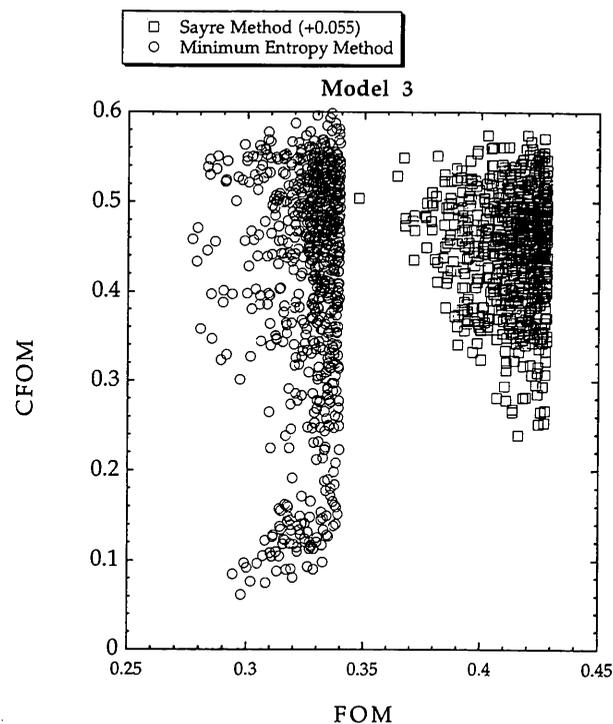


Fig. 4. Plot of the FOM versus CFOM for model 3. An absolute value of 0.55 has been added to the value of the FOM for the Sayre method so as not to overlap. The tail of low CFOM values for relatively low FOM values with the minimum-entropy method corresponds to the correct solution that could be found by analyzing in more detail the top four solutions as discussed in the text.

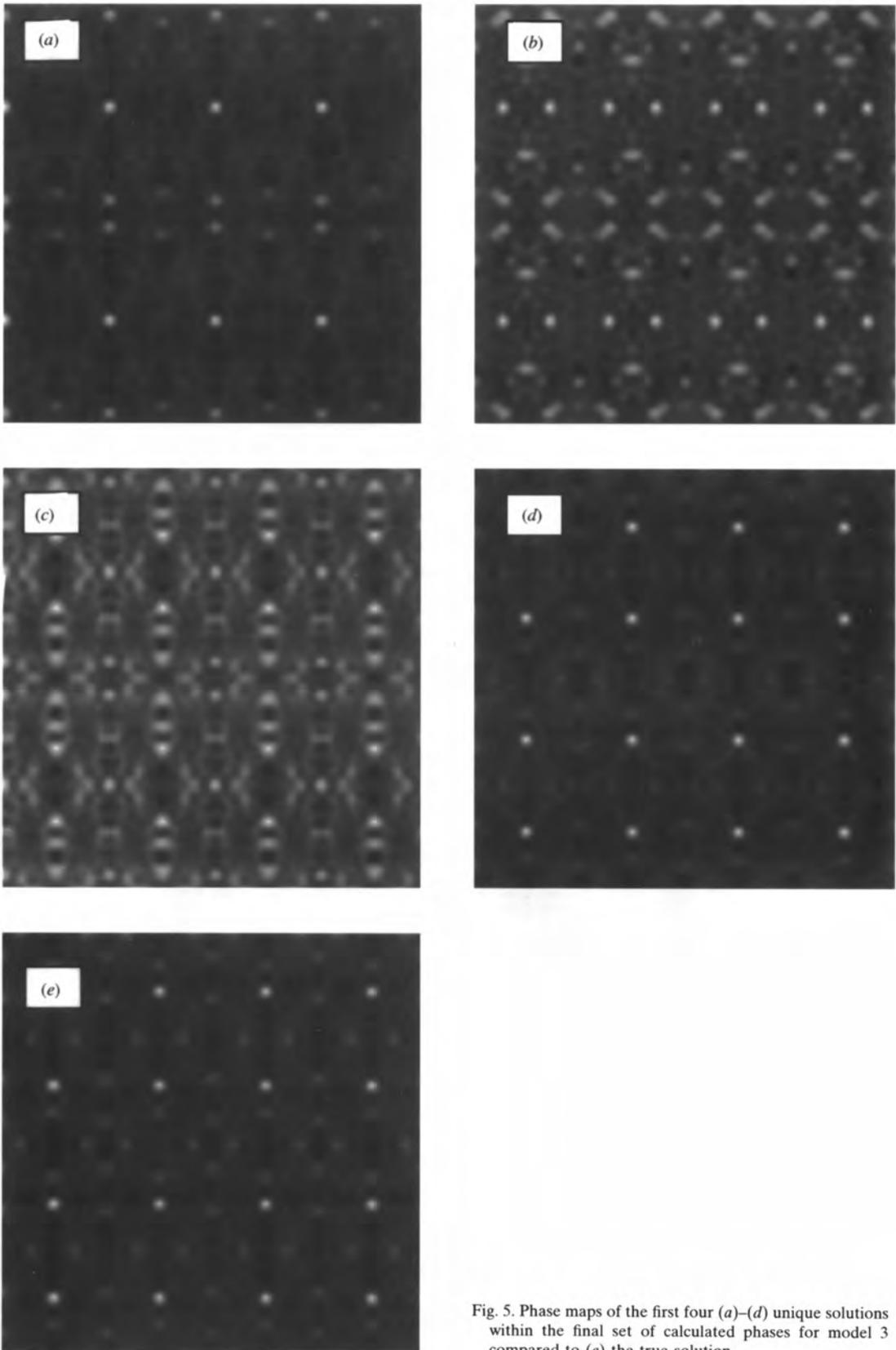


Fig. 5. Phase maps of the first four (a)–(d) unique solutions within the final set of calculated phases for model 3 compared to (e) the true solution.

only four solutions (see Fig. 5), two of which could be eliminated based on the low, noisy contrast character of the solutions (Figs. 5*b* and *c*). This leaves two unique solutions, which can be then be analyzed quantitatively to determine the correct one. A situation with more than one type of atom and strong reflections obscured by bulk diffraction spots will be the hardest case to solve. Note that we have assumed here nothing about the unit-cell contents, using the normalization  $A = 1$  from equation (13); if the correct normalization is used, this shifts the low CFOM values to lower FOM values. In other words, with correct normalization the FOM versus CFOM curve would resemble those in Figs. 2 and 3 – but this involves assumptions that one may not be entitled to make.

#### 4.4. Model 4

Similar to model 3, the Sayre method did not come close but the minimum-entropy method worked for  $C = 1$  or  $\exp(0.5)$ . [The value  $C = \exp(1.0)$  failed completely.] To explore this in more detail, we examined the results in terms of the number of atoms (an input parameter) in the unit cell – see Table 3. For the ‘true’ number of atoms, 102, both values of  $C$  are good but the larger value is much more tolerant of errors in the number of atoms and therefore much more robust. (However, it was worse for model 3.)

### 5. Discussion

Our results indicate that the minimum-entropy method is a robust method for solving surface problems, certainly superior to a simple Sayre method. The minimum-entropy method shares some similarities with maximum entropy with substantial non-linearities and some fundamental differences from traditional direct methods. It would be too strong to state that conventional direct methods will never work for surfaces; all the packages that we have been able to find are tuned for three-dimensional data and cannot be tested. Based upon testing with *MITHRIL* code, we can confidently state that they will not work in a routine fashion – perhaps after modification they will. How viable sophisticated Patterson search algorithms will be is also unknown at present, a good topic for future work.

After completing this work we used the minimum-entropy method for Au on an Si(111)  $6 \times 6$  surface (Marks *et al.*, 1998),  $p3$  with twinning, and were able to determine approximately half the atoms from the initial phasing when other methodologies were not effective. More recently, it has been applied to many more surface data sets (about 12), both ones that had previously been solved by Patterson methods and unsolved ones. One noteworthy example is for Bi on an InSb(110)  $2 \times 1$  surface (Van Gemmeren *et al.*, 1998). This is the extreme pathological case where 50% of the

reflections are unmeasurable and there are no  $\Sigma_2$  relationships. While the top six solutions were noise solutions, the seventh is very close to the true solution and of rather good quality – see Fig. 6. This directly corroborates our theoretical conclusions with the extreme cases where there are strong unmeasurable reflections (model 3). How well it will work in general, particularly for cases similar to model 3, remains to be seen, although one can also superimpose branching strategies setting certain beams (*e.g.* using  $\Sigma_1$  conditions). Preliminary tests indicate that it is also quite viable with three-dimensional data, for lower-resolution

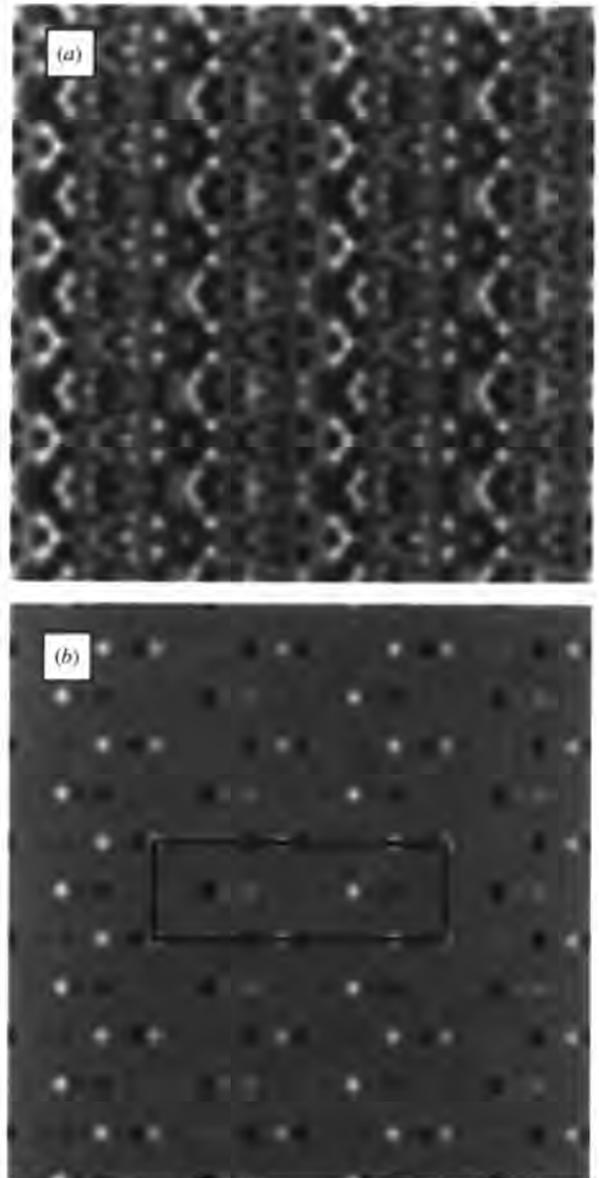


Fig. 6. Phase maps showing the top solution in (a) and the seventh solution in (b) using experimental data for Bi on an InSb(110)  $2 \times 1$  surface.

molecular envelope data or even three-dimensional surface diffraction data.

A few additional comments are appropriate about some of the technical numerical aspects of implementing the algorithm. Our experience is that obtaining an accurate form of the window function satisfying equation (10) is very important. A default Gaussian form adjusted to optimize for a complete set of reflections out to some cut-off works well and can normally be iterated to a more accurate form. However, sometimes this iteration will not work (a topic for further work). With the window, the algorithm behaves similar to a simple convolution, so array sizes of  $3N \times 3M$  are all that are needed to avoid aliasing effects with  $N$  and  $M$  the largest 'h' and 'k' values in the set of  $h, k$  reflections used in two dimensions. (The same appears to hold in three dimensions.) For reasons that are not completely obvious to us, square pixels in real space are rather important; we suspect that this eliminates astigmatism or other aberrations that will arise with uneven sampling. Except for cases with strong unmeasured reflections, the best solution generally has very small r.m.s. phase errors, with a cross-correlation coefficient of 0.9 or higher (CFOM < 0.1) relative to the true structure. This is enough to enable conclusions about the structure to be made just from the maps alone, critical in practice.

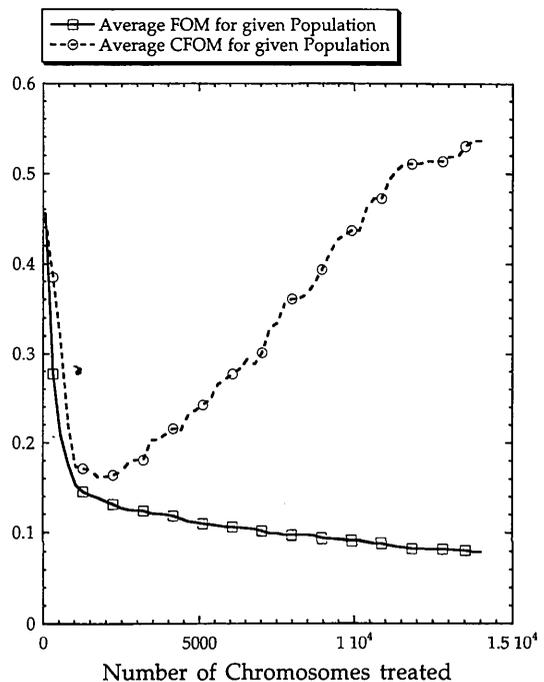


Fig. 7. Plot of the average FOM and CFOM in arbitrary units for a given population versus number of chromosomes treated. A chromosome refers to a set of assigned phases. A population is a previously determined number of chromosomes that are treated as a group and allowed to 'evolve' to a final collection of chromosomes consisting of the best solutions based upon the FOM.

The method itself is not necessarily unique and other variants are possible. For instance, one can use a higher power before the logarithm and we have experimented a little with this. Values of 1.25 or 1.5 seem to be useful, although it is not clear that they are better. An interesting avenue for future work is to pose the sharpening operator  $\hat{O}$  in terms of a calculus of variation problem and look for the 'optimum' value.

A comment should be made concerning another possible strategy. Consistently, we found that the Sayre method (and sometimes the minimum-entropy algorithm) came close to the correct solution when only a limited number of permutations were considered, but the overall FOMs for a more comprehensive analysis were sometimes not so good, see Fig. 7. This indicates that, in terms of an  $N$ -dimensional phase hypersurface, the minima associated with the true phases are large in area although not that deep. The probability of finding a particular solution in  $M$  trials depends not only upon the FOM but implicitly upon the selva regions from which other phases will convert (*via* the tangent formula) into it, *i.e.* an area-weighting term in the hyperspace. It should be possible to construct an 'area-weighted' FOM which might be very powerful, a topic for future research.

The authors would like to acknowledge the support of the National Science Foundation on grant No. DMR-9214505 in funding this work. We would also like to thank I. K. Robinson for supplying the X-ray Si(111)  $7 \times 7$  diffraction data, R. Fiedenhans<sup>1</sup> for the Bi data and C. Gilmore for useful discussions and assistance with the analysis using the *MITHRIL* code.

## References

- Bengu, E., Plass, R., Marks, L. D., Ishimiya, T., Ajayan, P. M. & Iijima, S. (1996). *Phys. Rev. Lett.* **77**, 20–24.
- Bricogne, G. (1984). *Acta Cryst.* **A40**, 410–445.
- Bricogne, G. & Gilmore, C. (1990). *Acta Cryst.* **A46**, 284–297.
- Collazo-Davila, C., Marks, L. D., Nishii, K. & Tanishiro, Y. (1997). *Surf. Rev. Lett.* **4**, 65–70.
- Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*, pp. 18–35. New York: John Wiley.
- Ferrer, S. (1993). *Surf. Sci. Lett.* **286**, L564–L570.
- Germain, G. & Woolfson, M. M. (1968). *Acta Cryst.* **B24**, 91–96.
- Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *Acta Cryst.* **A46**, 297–308.
- Gilmore, C. J., Marks, L. D., Grozea, D., Collazo, C., Landree, E. & Twesten, R. D. (1997). *Surf. Sci.* **381**, 77–91.
- Hoare, P. J. (1985). *J. Magn. Reson.* **62**, 561–567.
- Jayaram, G., Xu, P. & Marks, L. D. (1993). *Phys. Rev. Lett.* **71**, 3489–3492.
- Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635–651.
- Kullback, S. & Leibler, R. A. (1951). *Ann. Math. Stat.* **22**, 79–86.
- Landree, E., Collazo, C. & Marks, L. D. (1997). *Acta Cryst.* **B53**, 916–922.

- Marks, L. D., Grozea, D., Feidenhans'l, R., Nielsen, M. & Johnson, R. J. (1998). *Surf. Rev. Lett.* In the press.
- Marks, L. D., Plass, R. & Dorset, D. L. (1997). *Surf. Rev. Lett.* **4**, 1–8.
- Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in FORTRAN*, pp. 545–551. Cambridge University Press.
- Robinson, I. K., Waskiewicz, W. K., Fuoss, P. H., Stark, J. B. & Bennett, P. A. (1986). *Phys. Rev. B*, **33**, 7013–7016.
- Schenk, H. (1973). *Acta Cryst.* **A29**, 77–82.
- Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.
- Somoza, J. R., Szöke, H., Goodman, D. M., Béran, P., Truckses, D., Kim, S.-H. & Szöke, A. (1995). *Acta Cryst.* **A51**, 691–708.
- Van Gemmeren, T., Lottermoser, T. L., Falkenberg, G., Seehofer, L., Johnson, R. L., Gavioli, L., Mariani, C., Feidenhans'l, R., Landemark, E., Smilgies, D. & Nielsen, M. (1998). *Phys. Rev. B*. In the press.
- Woolfson, M. & Fan, H. F. (1995). *Physical and Non-Physical Methods of Solving Crystal Structures*. Cambridge University Press.
- Yao, J. X. (1981). *Acta Cryst.* **A37**, 642–644.