

Crystallographic direct methods for surfaces

L D Marks, N Erdman and A Subramanian

Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60201, USA

Received 11 June 2001, in final form 17 July 2001

Published 9 November 2001

Online at stacks.iop.org/JPhysCM/13/10677

Abstract

A brief review is given of the status of crystallographic direct methods for solving surface structure problems. The basic concepts of using the *a priori* information that the scattering comes from atoms to generate self-consistent equations and statistical constraints upon the phases are described. The use of a genetic algorithm to perform a search over possible phases is then discussed. Finally, solved cases in projection as well as the status of three-dimensional solutions for surfaces with a representative example are detailed.

1. Introduction

One of the most fundamental issues in surface science is determining the atomic scale structure. Over the years a standard methodology has been developed: experimental data are compared quantitatively (or qualitatively) with various different models and the best fit is deemed to be the correct structure. Unfortunately this approach has a fundamental problem—it requires that the true structure is very similar to one of the models chosen. Unless this is true no scientifically valid conclusions can be reached.

If the surface is very simple so that there are only two or three different possibilities, for instance oxygen adsorbed at a threefold or a bridging site, it is easy to ensure that something close to the correct model is used. However, if the structure is more complicated, for instance with more than five unique atomic sites (i.e. sites which are not related by any symmetry operations) the number of possibilities becomes unrealistically large. One has to have a method that will reduce the possibilities to a manageable number.

Over the last few years we have been developing crystallographic direct methods to overcome this problem [1–16]. (Independent of our work a slightly different direct methods approach has also proven to be viable for surfaces [17–20], although we will not discuss this here.) The basic idea is that one exploits the fact that the scattering comes from atoms to develop relationships, which connect the phases of the diffracted beams in a kinematical (i.e. single-scattering) measurement of the amplitudes using either surface x-ray diffraction or transmission electron diffraction. The phases do not need to be highly accurate; in general a root-mean-square error of 10–20 degrees is more than adequate. These approximate phases, coupled with the measured amplitudes, can then be used to produce an estimate (called a map)

of the charge density (with x-rays) or electrostatic potential (with electrons). Peaks in these maps correlate with the true atomic positions with relatively small errors, typically less than 0.1 Å. This is a standard approach in bulk crystallography, where it has been used with great success for many years. It is fair to state that it is currently almost trivial to solve a bulk structure if it does not have too many atoms, i.e. 100–200 unique sites. Few if any surface structures are even close to this complicated, which raises the question of why solving them should not be trivial.

In bulk crystallography one can routinely record more than 90% of all the possible diffraction spots to a resolution of 1 Å with good accuracy. If some are missed in one measurement run, one can always go back and measure them a second time; bulk crystals rarely change their structure when left to sit for days or weeks. One also has independent methods of determining the number of atoms in the structure. We do not have these luxuries with surfaces. Many of the surface diffraction spots overlap bulk spots which are orders of magnitude larger, and this presents special problems in handling them. While it is not uncommon for 90% of the reflections in two dimensions to be available, collecting three-dimensional data is very time consuming and often only 10–20% are measured at present. In most cases the measurement accuracies are much worse than what is considered acceptable for bulk materials. In addition, the samples are very sensitive to monolayer contamination, often rather poorly ordered, containing high concentrations of surface defects and multiple phases.

The sparsity of the number of measurements is one of the critical issues with surface diffraction data. Classical direct methods assume that all the strong reflections have been measured and only consider relationships between the measured reflections. (In effect, unmeasured reflections are taken to have zero intensity.) Therefore they can fail if there are large, unmeasured reflections.

A second major difference from bulk crystallography is the absence of good information about how many atoms there are in the reconstruction; in bulk crystallography this is almost always known ahead of time. If there is an adsorbate, one may know the coverage reasonably well, but often there are ambiguities. One rarely if ever knows how many atoms with the bulk composition are participating in the reconstruction. In any refinement against measured data one can almost always add more atoms (more variables) and improve the fit.

Another issue is the question of statistical bias due to the experimental procedure. If all the intensities were measured, we would have a representative sample. Often the experimental procedure is to perform a rather complete analysis in two dimensions, then analyse selected rods in the third. These will often be the stronger ones, which weights the measurements towards stronger reflections at the expense of the weaker ones.

Despite all these differences and additional complications for surfaces, with some care crystallographic direct methods can be applied to them. The following sections will describe some of the basics, some of the special aspects of solving a surface structure in three dimensions followed by an example and a brief discussion.

2. Basics

The principle behind crystallographic direct methods is to reduce the possible phase values by using *a priori* information as constraints; for a more detailed description of the mathematics see [15] and [21–23] for some general references on direct methods. The constraints fall into two classes: those based upon the fact that the scattering comes from atoms, and others based upon probabilities for random distributions of atoms. In particular:

- (a) We know the charge density for individual atoms, and to a very high degree of accuracy

we can ignore bonding effects and state that the charge density is some linear combination of the isolated atom charge densities, each atom being at some different location in the sample.

- (b) We know that the charge density has to be positive.
- (c) If the data have a reasonable resolution (corresponding to a reconstruction with about a 0.1 nm resolution) we know that most of the charge density is zero—the spaces between the atoms.
- (d) If we are trying to solve a structure in three dimensions, we know that the atoms are only significantly displaced from their bulk positions in a narrow region near the surface.

All these are additional types of information that we can exploit in some fashion. The oldest, and simplest is (a) above, which is known as atomicity. Consider the case of a structure with only one type of atom. In reciprocal space the structure factor can be written as

$$F(\mathbf{k}) = \sum_l f(\mathbf{k}) \exp(2\pi i \mathbf{k} \cdot \mathbf{r}_l) \exp(-2Bk^2) \quad (1)$$

for atoms at the positions \mathbf{r}_l with $f(\mathbf{k})$ the known atomic scattering factor. Here B is the isotropic temperature factor for each atom. Whereas in bulk crystallography this can often be fitted in the initial stages of the analysis using a Wilson plot [24], for surfaces this rarely seems to work and a reasonable estimate is good enough. Dividing by $Nf(\mathbf{k}) \exp(-2Bk^2)$ where N is the number of atoms gives what are called the unitary structure factors:

$$U(\mathbf{k}) = 1/N \sum_l \exp(2\pi i \mathbf{k} \cdot \mathbf{r}_l). \quad (2)$$

In real space we therefore have

$$u(\mathbf{r}) = 1/N \sum_l \delta(\mathbf{r} - \mathbf{r}_l) = Nu(\mathbf{r})^2 \quad (3)$$

or, alternatively, in reciprocal space:

$$U(\mathbf{k}) = N \sum_h U(\mathbf{k} - \mathbf{h})U(\mathbf{h}) \quad (4)$$

If we know the phases of $U(\mathbf{k} - \mathbf{h})$ and $U(\mathbf{h})$, we can therefore generate the phase for $U(\mathbf{k})$. In addition, if we guess the phases for the different reflections, we can use the difference in the phase (and amplitude) that we estimate from equation (4) and the phases that we started with to gauge how good that particular set of phases is. The ‘true’ set of phases has to obey equations similar to equation (4)—i.e. they cannot be totally random.

The above equation, which was first derived in a slightly different form by Sayre [25], is true for a structure composed of infinitely well resolved equal atoms. Unfortunately, this is never the case in practice. With some changes we can generalize it for real data with a limited resolution. Consider instead the form

$$F'(\mathbf{k}) = \sum_k A(\mathbf{k}) f(\mathbf{k}) \exp(2\pi i \mathbf{k} \cdot \mathbf{r}_l) \exp(-Bk^2) \quad (5)$$

where $A(\mathbf{k})$ is a ‘window function’ that will be defined below. We consider that $A(\mathbf{k})$ is a real function with positive values for a region of interest where the data are available (e.g. a disc in two dimensions or a sphere in three), and zero outside this. After dividing by the (known) structure factor and temperature factor term, we have a slightly different form in real space:

$$u'(\mathbf{r}) = 1/N \sum_l a(\mathbf{r} - \mathbf{r}_l) \quad (6)$$

where $a(\mathbf{r})$ is the Fourier transform of $A(\mathbf{k})$, the effective atom. Suppose further that there exists some operator O such that

$$O[a(\mathbf{r})] = a(\mathbf{r}). \quad (7)$$

Provided that the effective atoms at different locations do not overlap, it follows immediately that

$$O[u(\mathbf{r})] = u(\mathbf{r}). \quad (8)$$

a generalization of equation (3). For any particular operator (which could be the simple square as in equation (3), but does not have to be) equation (7) defines the function $A(\mathbf{k})$. Provided that it exists (in some cases it does not), we have a generalization that is independent of any resolution limits.

An alternative approach is to look at the same question using statistics and probabilistic relationships which are relatively insensitive to limited resolution. To illustrate these, consider the product

$$U(\mathbf{k} - \mathbf{h})U(\mathbf{h}) = 1/N^2 \sum_l \exp(2\pi i \mathbf{k} \cdot \mathbf{r}_l) \sum_m \exp(2\pi i \mathbf{h} \cdot [\mathbf{r}_m - \mathbf{r}_l]). \quad (9)$$

If the atoms are randomly distributed, the most probable value of the second sum is unity, so

$$U(\mathbf{k}) \sim N \langle U(\mathbf{k} - \mathbf{h})U(\mathbf{h}) \rangle. \quad (10)$$

Let us next consider

$$\begin{aligned} |U(\mathbf{k}) - NU(\mathbf{k} - \mathbf{h})U(\mathbf{h})|^2 &= |U(\mathbf{k})|^2 + N^2 |U(\mathbf{k} - \mathbf{h})U(\mathbf{h})|^2 \\ &\quad - 2N |U(\mathbf{k})U(\mathbf{k} - \mathbf{h})U(\mathbf{h})| \cos[\phi(\mathbf{k}) - \phi(\mathbf{k} - \mathbf{h}) - \phi(\mathbf{h})] \end{aligned} \quad (11)$$

where $\phi(\mathbf{k})$ is the phase of $U(\mathbf{k})$. In a statistical sense we will have a distribution of values. A standard theorem in statistics is the central limit theorem: all distributions tend towards being Gaussian. Hence the probability takes the limiting form

$$P(U(\mathbf{k}) - NU(\mathbf{k} - \mathbf{h})U(\mathbf{h})) \sim C \exp(-|U(\mathbf{k}) - NU(\mathbf{k} - \mathbf{h})U(\mathbf{h})|^2) \quad (12)$$

where C is a normalization constant. Taking all the amplitudes as known we can reduce this to what is called the Cochran distribution [26]

$$\begin{aligned} P(U(\mathbf{k}) - NU(\mathbf{k} - \mathbf{h})U(\mathbf{h})) \\ \sim C \exp(2N |U(\mathbf{k})U(\mathbf{k} - \mathbf{h})U(\mathbf{h})| \cos[\phi(\mathbf{k}) - \phi(\mathbf{k} - \mathbf{h}) - \phi(\mathbf{h})]). \end{aligned} \quad (13)$$

If the unitary structure factors are large, the sum of the phases

$$\phi(\mathbf{k}) - \phi(\mathbf{k} - \mathbf{h}) - \phi(\mathbf{h}) \sim 2n\pi \quad n = 0, 1, 2, \dots \quad (14)$$

This is known as the triplet phase relationship, abbreviated as \sum_2 . It is essentially the same as the unitary Sayre equation described earlier, derived in a statistical sense. By considering different powers or combinations of structure factors others can be derived. The triplet is generally considered the most useful for bulk crystallography, although the next term (called a positive quartet) has been used to solve some structures [17–20].

In practice [13, 15, 16] we use a variant of the generalized Sayre equation given in equation (8). The main information about the scattering of an atom is in the region near to its centre, so we want to emphasize the strong central peak, not any weak features far away. The operator therefore has to be sharpening: reinforce large regions and reduce weak ones. We also want it to be as unbiased as possible. The null hypothesis, i.e. the case without any information, would be a completely flat real-space form. The least biased case would then minimize the information for a Kullback–Liebler metric (e.g. [27]) relative to this, i.e.

$$d(u(\mathbf{r})) \equiv \langle u(\mathbf{r}) \rangle = \sum_r u(\mathbf{r}) \ln\{u(\mathbf{r})/\langle u(\mathbf{r}) \rangle\} - \sum_r u(\mathbf{r}) + \sum_r \langle u(\mathbf{r}) \rangle. \quad (15)$$

Consider the operator

$$O[u(\mathbf{r})] \equiv u(\mathbf{r}) \ln\{u(\mathbf{r})/\langle u(\mathbf{r}) \rangle\} \quad u(\mathbf{r}) > 0 \quad \equiv 0 \quad u(\mathbf{r}) < 0 \quad (16)$$

where $\langle a(\mathbf{r}) \rangle$ is the expectation value (mean) for a single atom-like feature. With some scaling term α , if

$$O[u(\mathbf{r})] = \alpha u(\mathbf{r}) \quad (17)$$

we have an atomistic-type constraint which also combines the idea of minimum information (minimum bias) and sharpening since it minimizes the metric in equation (15). We convert this to an iteration of the form

$$\begin{aligned} u^1(\mathbf{r}) = O[u_{n-1}(\mathbf{r})] &\equiv u_{n-1}(\mathbf{r}) \ln\{u_{n-1}(\mathbf{r})/\langle a(\mathbf{r}) \rangle\} & u_{n-1}(\mathbf{r}) > 0 \\ &\equiv 0 & u_{n-1}(\mathbf{r}) < 0 \end{aligned} \quad (18)$$

where $u^1(\mathbf{r})$ is the estimate for an improved charge density which is projected onto the observed structure factor moduli as well as any other constraints.

The particular form of $a(\mathbf{r})$ is determined to satisfy equation (17) exactly. This is done in practice by a Picard iteration of the form

$$a_n(\mathbf{r}) = \beta a_{n-1}(\mathbf{r}) \ln\{a_{n-1}(\mathbf{r})/\langle a(\mathbf{r}) \rangle\} \quad (19)$$

for the n th iteration, with β chosen to minimize $\sum |A_n(\mathbf{k}) - \beta A_{n-1}(\mathbf{k})|^2$. The actual form is very close to (in reciprocal space)

$$A(\mathbf{k}) = 1 - k/k_{max}. \quad (20)$$

A subtle, and important question is how to determine the scaling term α . We have some experimental structure factors, and for these we will always correct the modulus to the known values. We also have unmeasured reflections, and we want the scaling term to provide reasonable estimates of the moduli of these. For this we can find the value of the scaling by minimizing

$$\sum_{\mathbf{k}} |\alpha U^1(\mathbf{k}) - U_{n-1}(\mathbf{k})|^2 \quad (21)$$

for all the measured structure factors. (Using both the measured and unmeasured structure factors to determine the scaling is not very stable.)

There is no reasonable expansion of the Fourier transform of the logarithm, so the operation is done in real space. With relatively little extra cost in computer speed, one can encode a second operation into the imaginary part of the function when doing the inverse Fourier Transform. What we currently use for this second operation is

$$U^2(\mathbf{k}) = W(\mathbf{k}) \mathcal{F}^{-1}\{-u_{n-1}(\mathbf{r}) \ln[Nu_{n-1}(\mathbf{r})]\} \quad (22)$$

with \mathcal{F} an abbreviation for a Fourier transform, where $W(\mathbf{k})$ is defined as

$$W(\mathbf{k}) = A(\mathbf{k})/\mathcal{F}^{-1}\{-a(\mathbf{r}) \ln a(\mathbf{r})\} \quad (23)$$

This is similar to forcing a combination of atom-like features to maximize the entropy. Similar to the minimum information, a scaling term is constructed from a least squares fit. This pseudo-maximum entropy approach for a second operator works better than a positivity constraint that was used previously.

The figure of merit (FOM) used is then the least-squares error for the minimum information which is generated at the same time as the scaling. The structure of the overall algorithm is as follows:

- (1) Include phases for some of the stronger reflections.
- (2) Use the minimum entropy iteration to obtain an improved estimate for the wave, generating the FOM and rescaling the values.
- (3) Use the pseudo-maximum entropy iteration to obtain a second estimate, also appropriately rescaled.

- (4) Combine the two to form a new, intermediate estimate.
- (5) Over-relax the phases and correct the known structure factor moduli to give a new estimate.
- (6) Go to (2) and continue the iteration provided that the FOM is reducing. It should be noted that the operators involved are *not* unconditionally contractive, so the iteration will not converge beyond a certain point after which the FOM often increases. Since we only seek an approximate solution this is not a problem.

An important feature of the algorithm is the interpolation power of the entropy-type formulation, which couples many different reciprocal space values. The logarithm is a highly nonlinear function which is equivalent to considering not just square terms (as in the Sayre equation earlier) but also cubic contributions and many higher-order powers. This intrinsically builds in interpolation of the values for the reflections which have not been measured, a requirement for surfaces.

3. Genetic algorithms

The previous section gave a brief background to the basic ideas and the algorithm that we use in practice. If, independent of the starting phases, we could guarantee to end up with the same final solution, the overall problem would be relatively simple. Unfortunately this is not the case. Rather than there being just one minimum, i.e. the best FOM into which all iterations find, in most cases there are multiple minima. It is not true that the smallest FOM is always the correct solution, but one can say with confidence that the correct solution will be among the top five to ten unique solutions

One therefore has to test a relatively large number (1000–10 000) of different starting points for the analysis and collect the best solutions. These are then the models against which the data is quantitatively refined. To handle this, we use a genetic algorithm [14]. These are good at finding approximate solutions, and work exceptionally well in finding multiple local minima.

Genetic algorithms (e.g. [28]) use an approach very similar to natural evolution. Writing the set of phases as a binary string or chromosome, the concept is to look at a population of chromosomes. For each of these one evaluates the FOM. Taking these chromosomes as ‘parents’, the concept of natural selection (better FOMs preferred) is used to select which parents can reproduce. New children are next created by crosslinking, i.e. using part of one parent, part of the other. At this stage mutation of some of the bits can also be performed, although this is generally a weak phenomenon which can be neglected.

The basic structure of any genetic algorithm is, starting with some population, to calculate the FOM for each, produce new children and continue cycling until a reasonable convergence is achieved. Unfortunately, without some additional steps they do not work for finding multiple local minima. If the algorithm finds one reasonably good set of phases early on, it may produce many copies of the good schemata (or fragment of the chromosome) and drive others out. There may be other phase combinations which give a good FOM; they can be lost by this early dominance of one type. This is known as premature convergence.

Fortunately methods of preventing this from occurring are very well established. The methods involve reducing the FOM weighting when selecting children by a number that depends upon how similar the particular chromosome is to others in the population. One can also select parents in part depending upon how similar they are. In addition, rather than using just one population, it is standard to use more than one with different criteria for choosing the next generation of children, what is called niche specialization. With careful use of these additional steps the algorithm is remarkably powerful, performing parallel optimization to give

a representative sampling of the feasible-set solution space; in fact finding multiple minima is one its strengths. A weakness is that it may only find points close to the local minima rather than their exact locations, but this is not an issue, since we only want approximate solutions. Rather importantly, in general the size of the population that one needs scales approximately linearly with the number of variables, not as the square as in most optimization methods. This makes this approach very useful for larger problems.

The particular algorithm that we currently use involves the following steps.

- (a) Produce three different initial random populations. There are some advantages to making these initial populations cover the solution space as evenly as possible. One can do this in some cases using an error-correcting code to chose the initial population, or by pruning the population of members which are too similar. Unfortunately it is very expensive in computer time to generate a truly well dispersed initial population.
- (b) Cycling between the three different populations, calculate the FOM for each member.
- (c) Produce new children by a different method for each population. The three we use are the following.
 - (1) Choose the first parent with a probability weighted by its FOM times a cross-correlation metric of how similar it is to the other members of the population. Choose the second parent by how similar it is to the first parent as well as its weighted FOM.
 - (2) Choose the first parent as in (1), but with a different weighting for how similar it is for the second parent.
 - (3) Choose both parents using just their weighted FOMs.
- (d) Mutate the bits in a few of the children (mutation is not an important operator).
- (e) Every other generation, reintroduce a small number of the best solutions found to date using the initial phases.
- (f) In the alternating generations to (e) above, reintroduce a small number of the best solutions to date using the final phases.

The end result will be a large list of FOMs versus initial phases. One generally obtains bunches of solutions which are very similar; a cross-correlation test can be used to find the ones which are genuinely different. These are then used as the ‘models’ for the approximate location of atoms, and refinements are then performed. (Very often the smallest FOM is a very good approximation to the true surface charge density (or electrostatic potential), although this cannot be guaranteed.

4. Solving surfaces in three dimensions

Three-dimensional surface structures are in some respects harder, and in others easier to solve than a two dimensional projection. Although the data are less complete (a small fraction of the total possible measurements are available), atoms never overlap in three dimensions (they can in projection) and there are some special aspects of surfaces that can be exploited.

A three-dimensional surface has no periodicity in the direction normal to the surface, rather a continuous intensity distribution in reciprocal space. This is generally sampled at some fixed set of values with a given increment of $\Delta\ell$. The first step is to create an artificial supercell in real space corresponding to this sampling, i.e. a cell size of $1/\Delta\ell$.

We can also generally estimate the extent normal to the surface where there are substantial displacements away from the bulk positions, for instance by using the Patterson function. Let us call this distance L . We therefore have as an additional constraint in the problem [16] that

$$\rho(\mathbf{r}) \sim 0 \quad -L/2 < z < L/2. \quad (24)$$

Table 1. List of published surface structures solved at Northwestern in projection (two dimensions) by direct methods. * marks structures analysed as calibration tests.

Structure	Data	Reference	Atoms
Si(111) $-(\sqrt{3} \times \sqrt{3})R30^\circ$ Au*	electron	[1]	2
Si(111) $-(5 \times 2)$ Au*	electron	[1]	5
Si(111) $-(7 \times 7)^*$	electron	[2]	14
Si(111) $-(6 \times 6)^*$ Au	x-ray	[3]	15
TiO ₂ (100) -1×3	x-ray	[4]	4
Si(111) $-(3 \times 1)$ Ag	electron	[5]	5
Si(111) $-(\sqrt{3} \times \sqrt{3})R30^\circ$ Ag*	electron	[6]	2
Ge(111) $-(4 \times 4)$ Ag	x-ray	[7]	10
MgO(111) $-(\sqrt{3} \times \sqrt{3})R30^\circ$	electron	[8]	1
MgO(111) $-(2 \times 2)$	electron	[8]	2
MgO(111) $-(2\sqrt{3} \times 2\sqrt{3})R30^\circ$	electron	[8]	3
Ni(111) $-(5\sqrt{3} \times 2)$ S*	x-ray	[9]	5
Cu(111) $-\begin{vmatrix} 4 & 1 \\ -1 & 4 \end{vmatrix} S^*$	x-ray	[9]	3
Cu(110) $-p(4 \times 1)$ Bi*	x-ray	[9]	3
Ge(111) $-(3 \times 1)$ Ag	electron	[10]	5

Table 2. List of surface structures solved in three dimensions at Northwestern by direct methods. * marks structures analysed as calibration tests.

Structure	Reference	Atoms
Ge(111) $-(3 \times 1)$ Rb*	—	4
Si(111) $-(4 \times 1)$ In*	—	4
Si(001) $-(4 \times 3)$ In	—	10
NiO(111) $-(2 \times 2)$	[11]	4
InSb(001) $-c(2 \times 8)$	[12]	12
GaAs(001) $-c(2 \times 8)$	[12]	12
InAs(001) $-c(2 \times 8)$	[12]	12

This is a strong constraint (known as a ‘support constraint’), which has been extensively analysed in the image processing literature [29]. One additional aspect is that we do not need to know all the phases, just a subset of them. In particular, from the sampling theorem,

$$F(\mathbf{h}, \mathbf{k}, \mathbf{l}) = \sum_n F(\mathbf{h}, \mathbf{k}, n/L) \sin(\pi[l - n/L]) / \pi[l - n/L] \quad (25)$$

where \mathbf{h} and \mathbf{k} are reciprocal lattice vectors in the surface plane. Therefore the ‘true’ variables in the problem are not the phases of all the reflections, only a much smaller subset spaced by $1/L$ in reciprocal space. This reduces the dimensionality of the problem, and one can also use equation (25) in the first cycle to expand from the smaller subset to the full set of values in the first iteration.

This raises one issue special to the case of surface data—how many independent measurements are there? The Patterson function spans a width of $2L$ Å; outside this region it will be zero. The sampling theorem tells us that if we have some function which is only non-zero for a width of $2L$, it is fully defined by a sampling in reciprocal space of $1/(2L)$. However, in a typical surface x-ray diffraction experiment measurements are made with much finer sampling than this, i.e. the data are oversampled. A good case can be made that the number of truly independent measurements is given by the sampling theorem step $1/(2L)$, which is generally much smaller than the number of measurements. It may well be that some published

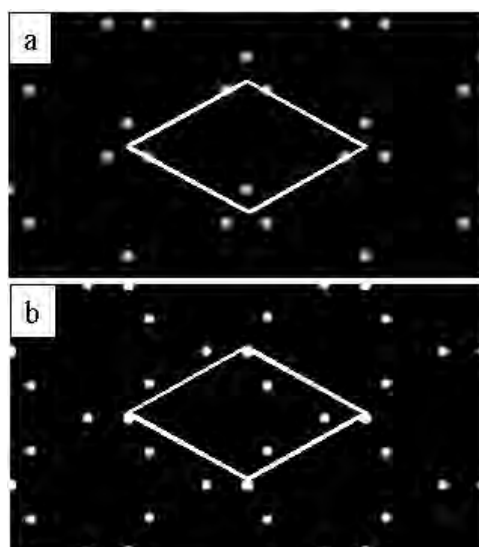


Figure 1. Charge density maps calculated using the 2D code corresponding to (a) $p3$ and (b) $p3m1$ symmetries. The primitive unit cell is indicated with solid lines ($a = b = 5.8970 \text{ \AA}$, $g = 120$). White regions correspond to high charge density, the probable locations of atoms.

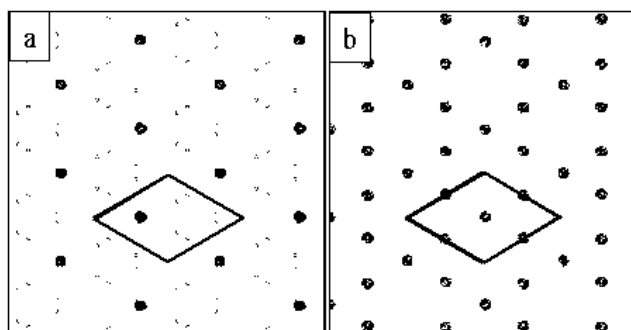


Figure 2. Contour maps obtained from 3D code for $P3m1$ symmetry. Shown are two out of 39 'slices' along the z -axis, separated by 2.257 \AA ; (a) shows the top (first) Ni layer with three-quarters of the Ni atoms missing, and (b) shows the underlying (third) Ni layer with one-quarter of the Ni atoms missing. The primitive unit cell is indicated with solid lines ($a = b = 5.8970 \text{ \AA}$, $g = 120$). The dense contour regions (essentially black) are the sites of the Ni atoms.

three-dimensional surface structures have had far too few independent measurements and are therefore suspect (or wrong).

5. Examples

Solving structures in projection with two dimensional data has been fairly well covered in the literature and will not be discussed in more detail here. Table 1 gives an overview of the cases that we have analysed to date. Three dimensions are easier in principle since atoms can overlap in projection (which breaks the assumptions behind the generalized Sayre equation, although not the statistical methods). However, in practice they are much harder because a

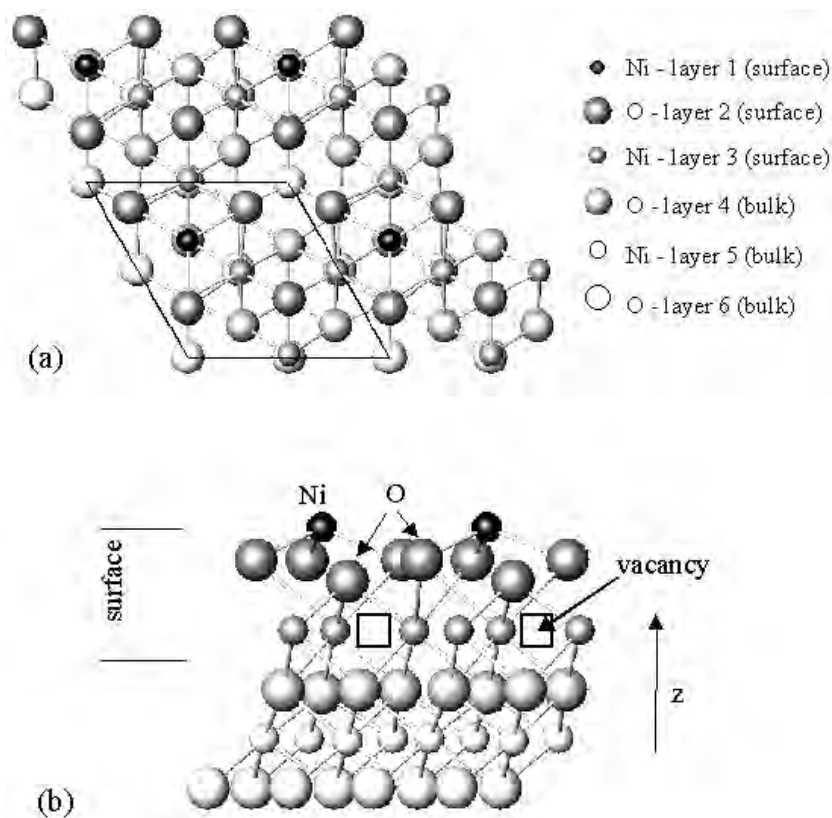


Figure 3. (a) Top view of the oxidized $p(2 \times 2)\text{NiO}(111)$ surface structure obtained by direct methods analysis. (b) Side view of the structure. The solution is characterized by one-quarter of the Ni atoms removed from the second layer and put on the surface in a 2×2 configuration (Schottky defect formation).

smaller fraction of the intensities have been measured.

Table 2 summarizes the three-dimensional surfaces that we have examined to date. A simple example is the $\text{NiO}(111)2 \times 2$ surface [11]. The two dimensional projection showed two possible solutions (see figure 1), one with a trimer of strong scatterers (candidate Ni sites), the other with four. The three-dimensional analysis (figure 2) indicated that the trimer solution was incorrect, and gave two main Ni sites separated normal to the electron beam. The sites of the oxygen atoms were not apparent in the maps, but plausible locations could be inferred and then tested by refinements. For reference, figure 3 shows the final structure after refinement against the experimental data.

6. Discussion

Crystallographic direct methods are very powerful model-independent methods for solving surface structures. Provided the experimental data are good enough, it is relatively easy to obtain a good initial model for the structure in either a two dimensional projection or in three dimensions. More complicated structures with many atoms are not, in principle, a problem, and in fact the statistical relationships work better, hence the algorithms should work better.

Our experience is that the complexity of the calculations scales roughly linearly with the complexity of the structure, so one does not need access to very fast computers.

The principal current limitation is the quality of the experimental data that is available, particularly in three dimensions. A standard benchmark in crystallography is that one should have 10–20 measurements per variable to have any confidence in the final results. Using three positional variables per atom and one temperature factor to properly determine a structure with ten independent sites requires at least 400 unique measurements (with the appropriate $1/2L$ sampling as mentioned above). This is more measurements than is currently available in some cases. As a consequence, one may not be able to uniquely determine all the sites, particularly those of lighter elements. In many cases a reasonable guess based upon chemical rules can be used, but this can be dangerous. Hopefully with the increased use of brighter x-ray sources more complete data will become available and complicated structures, e.g. proteins at surfaces, will become amenable to study.

References

- [1] Marks L D, Plass R and Dorset D L 1997 *Surf. Rev. Lett.* **4** 1
- [2] Gilmore C J, Marks L D, Grozea D, Collazo C, Landree E and Twisten R D 1997 *Surf. Sci.* **381** 77
- [3] Marks L D, Grozea D, Feidenhans'l R, Nielsen M and Johnson R L 1997 *Surf. Rev. Lett.* **5** 459
Marks L D, Grozea D, Feidenhans'l R, Nielsen M and Johnson R L 1998 *Surf. Sci.* **418** 32
- [4] Landree E, Marks L D, Zschack P and Gilmore C J 1998 *Surf. Sci.* **408** 300
- [5] Collazo-Davila C, Grozea D and Marks L D 1998 *Phys. Rev. Lett.* **80** 1678
- [6] Grozea D, Landree E, Collazo-Davila C, Bengu E, Plass R and Marks L D 1999 *Micron* **30** 41
- [7] Collazo-Davila C, Grozea D, Marks L D, Feidenhans'l R, Nielsen M, Seehofer L, Lottermoser L, Falkenberg G and Johnson R L 1998 *Surf. Sci.* **418** 395
- [8] Plass R, Egan K, Gajdardziska-Josifovska M, Collazo-Davila C, Grozea D, Landree E and Marks L D 1998 *Phys. Rev. Lett.* **81** 4891
- [9] Marks L D, Bengu E, Collazo-Davila C, Grozea D, Landree E, Leslie C and Sinkler W 1998 *Surf. Rev. Lett.* **5** 1087
- [10] Grozea D, Bengu E, Collazo-Davila C and Marks L D 1999 *Surf. Rev. Lett.* **6** 1061
- [11] Erdman N, Warschkow O, Ellis D E and Marks L D 2000 *Surf. Sci.* **470** 1–2
- [12] Kumpf C *et al* 2001 *Phys. Rev. Lett.* **86** 3586
- [13] Marks L D and Landree E 1998 *Acta Crystallogr. A* **54** 296
- [14] Landree E, Collazo-Davila C and Marks L D 1997 *Acta Crystallogr. B* **53** 916
- [15] Marks L D, Sinkler W and Landree E 1999 *Acta Crystallogr. A* **55** 601
- [16] Marks L D 1999 *Phys. Rev. B* **60** 2771
- [17] Torrelles X, Rius J, Boscherini F, Heun S, Mueller B H, Ferrer S, Alvarez J and Miravittles C 1998 *Phys. Rev. B* **57** R4281
- [18] Torrelles X, Rius J, Miravittles C and Ferrer S 1999 *Surf. Sci.* **423** 338
- [19] Torrelles X, Rius J, Pedio M, Felici R, Rudolf P, Alvarez J, Ferrer S and Miravittles C 1999 *Phys. Status Solidi B* **215** 773
- [20] Pedio M, Felici R, Torrelles X, Rudolf P, Capozzi M, Rius J and Ferrer S 2000 *Phys. Rev. Lett.* **85** 1040
- [21] Woolfson M M 1987 *Acta Crystallogr. A* **43** 593
- [22] Giacovazzo C 1980 *Direct Methods in Crystallography* (New York: Plenum)
- [23] Woolfson M M and Fan H 1995 *Physical and Non-Physical Methods of Solving Crystal Structures* (Cambridge: Cambridge University Press)
- [24] Wilson A J C 1942 *Nature* **150** 152
Wilson A J C 1949 *Acta Crystallogr.* **2** 318
- [25] Sayre D 1952 *Acta Crystallogr.* **5** 60
- [26] Cochran W 1955 *Acta Crystallogr.* **11** 579
- [27] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley)
- [28] Goldberg D E 1989 *Genetic Algorithms in Search, Optimization, and Machine Language* (Reading, MA: Addison-Wesley)
- [29] See e.g. the articles in Stark H (ed) 1987 *Image Recovery: Theory and Application* (Orlando, FL: Academic)