Journal of Microscopy, Vol. 195, Pt 1, July 1999, pp. 44–57.
Received 31 January 1995; accepted 20 February 1999

Crystal structure retrieval by maximum entropy analysis of atomic resolution incoherent images

A. J. McGIBBON,* S. J. PENNYCOOK & D. E. JESSON
Solid State Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6031, U.S.A.

Key words. Atomic resolution, crystal structure retrieval, incoherent imaging, maximum entropy, STEM, Z-contrast imaging.

Summary

In this paper, we discuss the application of the maximum entropy method to atomic resolution Z-contrast images acquired in a scanning transmission electron microscope. Z-contrast is an incoherent imaging technique, and can be described as a convolution between an object function (the real-space map of the columnar scattering cross-section to high angles) and a point spread function (the effective electron probe). As such, we show that the technique is ideally suited to maximum entropy analysis which can, given an electron probe distribution, retrieve the 'most likely' Z-contrast object function. Using both simulated and experimentally acquired data, we explore the capabilities of maximum entropy analysis when applied to atomic resolution Z-contrast images, drawing conclusions on both the range of applicability of the technique and the nature of the retrieved crystal structures. Ultimately, we show the way in which the combination of Z-contrast imaging with maximum entropy analysis can be used to yield important information on unexpected atomic structures.

Introduction

The physical properties of many important materials rely on the nature of atomic arrangements at defects, grain boundaries and interfaces. The ultimate aim of structural analyses of such systems is to produce, solely from experimental data, a reconstruction of the region of interest on an atomic scale. The necessary experimental basis for the realization of this goal is a technique which can provide directly interpretable information on crystal structures on a column-by-column level. From these data, it is equally important that, in addition to retrieving as much information as possible, inferences on the nature of the atomic structure must be made in a logical, consistent and statistically sound manner. In this paper, we show that the aims and criteria stated can be fulfilled employing a maximum entropy analysis of atomic resolution incoherent images of crystals acquired with a scanning transmission electron microscope. By this route, we show that structural information on the angstrom scale can be retrieved from the experimental evidence alone.

Previous investigations of crystalline specimens orientated along principal zone axes in a scanning transmission electron microscope (e.g. Pennycook & Jesson, 1991) have shown that, by forming images principally from the high-angle thermally diffuse scattered component of the transmitted signal, it is possible to produce high spatial resolution images with variations in intensity that can be related to the mean squared atomic number of the region irradiated by the electron probe. Furthermore, the recorded signal in such Z-contrast images is highly localized to the atomic column positions and can be interpreted as a map of the high-angle columnar scattering intensity, the resolution of which is primarily limited by the size of the electron probe. Z-contrast is an incoherent imaging technique in that no reversals in image contrast occur as a function of probe defocus or column separation. Consequently, unlike high resolution electron microscopy (HREM), in which contrast reversal does occur (e.g. Smith *et al.*, 1985), Z-contrast images are intrinsically intuitive.

Z-contrast images possess a wealth of quantifiable information on column position and composition at or below the probe-defined resolution limit. In this paper, we address the important issue of how much information can be reliably retrieved from the evidence. Since the technique is incoherent (Pennycook & Jesson, 1991; Loane *et al.* 1992), image intensity can be described in terms of a convolution between a point spread function (the effective electron probe) and an object function (the map of the high-angle columnar scattering cross-section). In this way, the task of retrieving the initial crystal structure given the experimental evidence becomes a classic inverse problem.

Correspondence to: S. J. Pennycook. Tel: +1 423 574 5504; fax: +1 423 574 4143; e-mail: pyk@ornl.gov

* Present address: Analogue Process Development, National Semiconductor (UK) Ltd, Larkfield Industrial Estate, Greenock PA16 0AQ, U.K.

The method employed here in the retrieval of the object function is that of maximum entropy (Gull & Daniell, 1978; Gull & Skilling, 1984). This technique is based on Bayesian probability theory and centres on the selection of a 'most likely' object function (given a particular point spread function) by choosing that with the highest entropy which is consistent with the evidence of the available data. Maximum entropy has previously been applied to a number of inverse problems in experimental analysis in both one dimension (e.g. Kuzuo & Tanaka, 1993; Sivia *et al.*, 1993) and two dimensions (e.g. Bonavito *et al.*, 1993). Previous applications to image analysis in electron microscopy were made by, for example, Farrow & Ottensmayer (1989) and Hu & Li (1991), whilst preliminary discussions of the technique in the context described here were made by Pennycook *et al.* (1992).

We begin this paper by describing the principal features of both incoherent imaging in scanning transmission electron microscopy (STEM) and in maximum entropy. Emphasis is given to the compatibility of the two techniques and the way in which this approach can be used to retrieve the Z-contrast object function. By applying maximum entropy to a number of simulated Z-contrast images, we define the scope of applicability of the technique and gain an understanding of the output data. We then proceed to apply maximum entropy to the study of as-acquired Z-contrast images, drawing conclusions both on the validity of our approach and on the nature of the observed crystal structure.

Incoherent imaging in STEM: an inverse problem

The key consideration in the general philosophy of applying maximum entropy to the study of atomic resolution Z-contrast images is that the recorded image is incoherent in the transverse plane of the specimen (**R**) perpendicular to the beam direction (z). A schematic diagram outlining Z-contrast imaging of crystals in STEM is given in Fig. 1. A finely focused electron probe is scanned across a thin-film specimen orientated along a principal zone axis and the transmitted high-angle scattered signal recorded by means of an annular detector. All images described in this paper are associated with experimental and theoretical data from a VG Microscopes HB603 STEM operating at 300 kV, possessing an optimum electron probe size at Scherzer defocus of 1·3 Å.

As explained by Jesson & Pennycook (1993), transverse coherence is broken primarily as a result of detector geometry. By considering two neighbouring atoms in the transverse plane they showed that, when integrating over an annular detector of sufficiently high inner angle θ_i, destructive interference effectively cancels out coherent scattering contributions. As a result, the detected signal approaches that for two atoms scattering incoherently. In

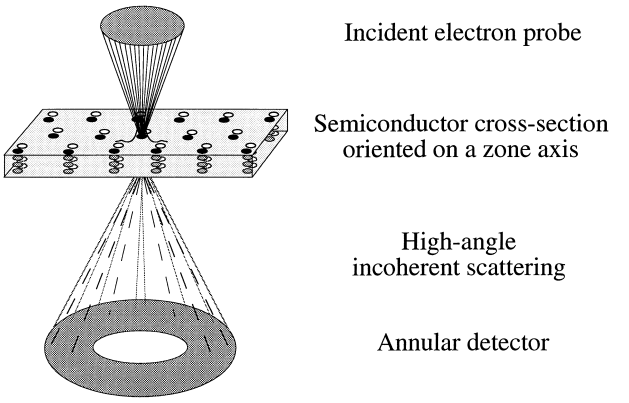

Fig. 1. Schematic diagram depicting Z-contrast (high angle annular dark-field) imaging in STEM.

practice, the high-angle coherent signal is further damped by the presence of phonon vibrations, which largely destroys coherence along the beam direction, ensuring that a column of atoms also scatters incoherently. From their calculations, they suggested a general criterion for the minimum inner angle θ_i^{inc} required for incoherent imaging of neighbouring atoms or columns separated by a distance ρ:

$$\theta_i^{inc} = \frac{1{\cdot}22\lambda}{\rho} \tag{1}$$

where λ is the electron wavelength. For incoherent imaging at probe-limited resolution (1·3 Å at 300 kV) $\theta_i^{inc} \approx 20$ mrad, giving a coherent contribution to image intensity of $< 5\%$. The smallest collection angle used in experiments described here was 30 mrad for the analysis of <110> GaAs.

The practical result of transverse incoherence is that, as probe defocus is changed, there are no reversals in image contrast, only a reduction in spatial resolution at values below Scherzer defocus, and an increase in resolution above Scherzer defocus accompanied by a loss in direct interpretability due to enhanced probe tails (Pennycook & Jesson, 1991). Consequently, the imaging process can be described in terms of a (normally positive) object transfer function as shown in Fig. 2. The corollary of this is that the recorded image I(**R**) can be simply expressed as a convolution:

$$\mathrm{I}(\mathbf{R}) = \mathrm{O}(\mathbf{R}) * \mathrm{P}_{\mathrm{eff}}{}^2(\mathbf{R}) + \mathrm{N}(\mathbf{R}) \tag{2}$$

where O is the object, $P_{eff}{}^2$ is the effective electron probe and N is image noise. Provided the criterion for transverse incoherent imaging is satisfied, the object function can be written as

$$\mathrm{O}(\mathbf{R}) = \frac{2}{\mathrm{h}v} V^{\mathrm{HA}}(\mathbf{R}) \tag{3}$$

where V^{HA} is the two-dimensional high-angle potential associated with scattering to the annular detector. In the

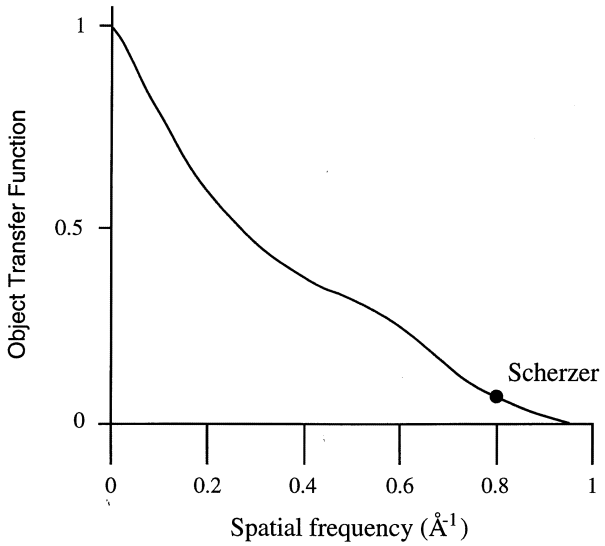

Fig. 2. Graph showing the (always positive) object transfer function for Z-contrast imaging in a 300-kV STEM, under optimum Scherzer conditions for incoherent imaging (defocus −44 nm, objective semiangle 9·4 mrad, spherical aberration coefficient 1 mm). The Scherzer resolution limit of 0·13 nm is marked.

limit of complete localization, Eq. (3) becomes

$$O(\mathbf{R}) = \sum_i \sigma_i \delta(\mathbf{R} - \mathbf{R}_i) \tag{4}$$

Where σ_i is the cross-section for high-angle scattering of the ith column. Furthermore, Pennycook & Jesson (1992) have shown that the object function for Z-contrast imaging of crystals can best be described using a Bloch-wave formulation. By collecting the high-angle signal, the primary contribution to detected intensity is from the non-dispersive, tightly bound s-states. Contributions from higher-order and more delocalized states are averaged out by the process of angular integration. Using the s-state formulation, the object function (which includes the effect of signal absorption as a function of film thickness t through the s-state absorption coefficient μ^{1s}) can be described in the following terms:

$$O(\mathbf{R}_i, t) = \sum_i \sigma_i \left(\epsilon^{1s}(0)\right)^2 \left(\tau^{1s}(\mathbf{R}_i, 0)\right)^2 \left(\frac{1-e^{-2\mu^{1s}(0)t}}{2\mu^{1s}(0)}\right) \delta(\mathbf{R} - \mathbf{R}_i) \tag{5}$$

where ϵ^{1s} is the excitation amplitude of state τ^{1s}. Using the same formulation, the effective probe is:

$$P^2_{\text{eff}}(\mathbf{R}) = \left| \frac{1}{\epsilon^{1s}(0)} \int_{\substack{\text{objective} \\ \text{aperture}}} \epsilon^{1s}(\mathbf{K}) e^{i[\mathbf{K}.(\mathbf{R}-\mathbf{R}_0)+\gamma(\mathbf{K})]} d\mathbf{K} \right|^2 \tag{6}$$

The primary assumption of Eqs. (4) to (6) is that the lateral extent of the projected high-angle potential at the column sites is small when compared to the size of the electron probe scanning the specimen. It should be noted that this approximation holds even in the case of imaging at 300 kV where the finite distribution of the projected potential is $<10\%$ that of the probe size. Consequently, under the assumption of complete localization, the effective probe contains a small but insignificant specimen-dependent contribution. Equation (6) includes a slow angular fall-off in s-state excitation across the objective aperture which reaches a value of (typically) $0{\cdot}6–0{\cdot}8\epsilon(0)$ at the aperture cut-off. In real space, this results in a small ($<10\%$) symmetric broadening of the effective probe. When imaging samples that are tilted slightly off the appropriate zone axis, this may result in an effective shift of the aperture position across the dispersion plane, which may in turn lead to asymmetries in the effective probe distribution. The consequences of such an effect on the interpretation of experimentally acquired images will be addressed in a later section.

The arguments to this point have centred on the assumption that the specimen of interest is a perfect crystal. However, when imaging imperfect regions of high strain such as grain boundaries or dislocations, dechannelling may affect image interpretation. Discussion of such effects in the light of experimental data will be discussed in later sections.

It should also be noted that no attempt has been made to evaluate individual columnar scattering cross-sections. Although primarily incoherent, Jesson & Pennycook (1995) showed that rather than representing each column as an array of independently vibrating atoms, scattering can be visualized in terms of an assembly of independent 'packets' of atoms. By retrieving the object function directly from acquired data, it may be possible to gain a greater understanding of effects such as partial coherence in the z-direction directly from the image data.

Retrieval of the object function: a maximum entropy approach

As illustrated in Fig. 3, we are now at a stage where the Z-contrast image can, in general, be described as an inversion problem (Eq. 2): a convolution between an object function (the real-space map of the s-state-dominated columnar scattering cross-section) and a point spread function (the effective electron probe). There remains, however, the non-trivial problem of retrieving the object function in a logical and consistent manner. To date, when confronted with a problem that is incoherent in nature, the most common method employed by electron microscopists in one- or two-dimensional problems is that of simple Fourier deconvolution (e.g. Batson *et al.*, 1992). However, such techniques are highly sensitive to the presence of noise, and as such are precluded from the majority of problems posed in Z-contrast imaging. In addition, no basis exists for judging the

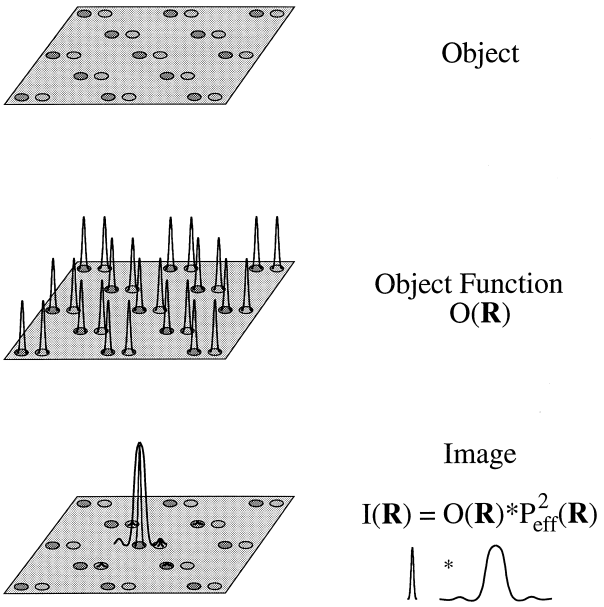

Fig. 3. Schematic diagram showing the way in which a Z-contrast image can be interpreted as a convolution between a point spread function (the effective electron probe) and an object function (the map of the high-angle columnar scattering cross-section).

'likelihood' of the deconvoluted result given the evidence of the original image. In this paper, we base our choice of the 'most likely' object function given the image evidence using the technique of maximum entropy as described by Gull & Skilling (1984).

The following discussion of the maximum entropy technique will be given in terms of a Z-contrast image consisting of N pixels (an N-dimensional problem) from which the aim is to begin with an initial estimate of the object function, **m**, and proceed to the 'most likely' solution, **h**. Here, **m** and **n** are N-dimensional vectors representing the N image pixels. Now, given a particular Z-contrast image, there are a large number of possible object functions which, when convoluted with the effective electron probe, agree with the evidence to within experimental error. The conventional way in which to choose the most likely object function is by measuring the 'likelihood', L, where

$$L(\mathbf{h}) = \frac{\chi^2(\mathbf{h})}{2} = \frac{1}{2}\sum_{i=1}^{N}\frac{(\mathrm{M}(\mathbf{h})_i - \mathrm{I}_i)^2}{\mathrm{s}_i^2}. \qquad (7)$$

Here, **M** represents 'mock' data, the convolution of the object function being tested with the given probe intensity profile, and **I** represents the original image intensity values. Both are N-dimensional data sets, and s_i is the standard deviation of the intensity in the ith pixel. The conventional measure of fit is through χ^2, which differs from L only by a factor of two, and has a 'best' value N for a perfect fit.

As a trivial two-dimensional example, we will refer to Fig. 4, which illustrates the maximum entropy approach towards the solution of the problem of two simultaneous equations in x and y. Values of the two 'pixels' are plotted in a two-dimensional plot to illustrate how the maximum entropy method moves from the initial assumption $\mathbf{m} = (m_x, m_y)$ to a final answer of maximum likelihood $\mathbf{h} = (h_x, h_y)$. In Fig. 4, the global minimum of L is unique, and contours of constant L (the shape of which depend on the associated errors) radiate outward from this point. It can be seen from this figure that, at any point other than the final **h**, many potential solutions have equal likelihood. It is therefore necessary to establish a criterion for choosing the 'most likely' object function from those located on the contour of equal likelihood. Using Bayesian probability theory as a basis, it has been established both mathematically and conceptually (e.g. Gull & Skilling, 1984) that a consistent choice can be made by maximizing the value of entropy:

$$S = -\sum_i p_i \log(p_i), \qquad (8)$$

where p_i is the probable intensity of pixel i and $\sum p_i = 1$. The process of selecting a 'best' image based on optimising some criterion in this way is known as regularization. Perhaps the best way to visualize this choice of regularization function is to imagine what the most likely object function would be in the limit of an image consisting solely of random noise. The only logical choice in this case, given the lack of experimental evidence, is an object function of uniform intensity, i.e. that of maximum entropy. A more generally applicable version of Eq. (9) which is necessary when dealing with a positive additive distribution (such as a Z-contrast image)

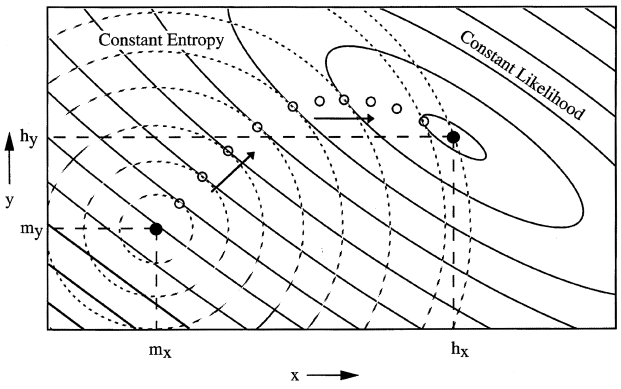

Fig. 4. Schematic graph depicting the 'maximum entropy trajectory' (marked with circles) towards the solution of a two-dimensional problem in x and y. Lines of constant entropy (broken) and constant likelihood (full) are marked. The maximum entropy routine proceeds from (m_x, m_y), the global maximum of entropy, to (h_x, h_y), the point of maximum constrained entropy.

is to maximize entropy in the form:

$$S(\mathbf{h}) = \sum_{i=1}^{N}\left(\mathrm{h}_i - \mathrm{m}_i - \mathrm{h}_i \log\left(\frac{\mathrm{h}_i}{\mathrm{m}_i}\right)\right) \qquad (9)$$

where **m** is the initial or 'default' model for the object function defined previously, and the maximum possible value for S is zero. The value of **m** can, if required, include known information on the nature of the object function (Gull *et al.*, 1986). It should be stressed, however, that in all analyses described here, no preconceived models are employed and so all m_i are of equal intensity. Consequently, on a plot such as Fig. 4, **m** is at a global maximum, and contours of equal entropy radiate outward from **m**.

Having defined criteria for both likelihood and regularization, it is now possible to combine the two to provide a route via which the most likely object function can be evaluated in the light of the available evidence. The maximum entropy condition gives the safest possible image, i.e. the image with the least structure necessary to fit the data. It is a technique of constrained deconvolution, and proceeds by maximizing the value of constrained entropy $Q(\mathbf{h})$:

$$Q(\mathbf{h}) = \alpha S(\mathbf{h}) - L(\mathbf{h}) \qquad (10)$$

where α is known as the regularization constant. Evaluation of Q is carried out by an iterative algorithm, the details of which are described by Gull & Skilling (1990). In this paper, all analyses of Z-contrast images based on the Gull & Skilling algorithm were carried out using a VG Microscopes MaxEnt software routine on an IBM PC equipped with an accelerator card.

In qualitative terms, the iterative process begins with the assumption of maximum entropy in the object function, i.e. $\mathbf{h} = \mathbf{m}$, where all m_i are set to the same value. Evaluation of constrained entropy begins with α set to a high (effectively infinite) value, with the aim being to move towards the minimum numerical value of likelihood consistent with the experimental data. For each α, the subset of likely object functions is determined on the basis of the likelihood criterion. From this subset, that object function with maximum entropy is chosen by iteration. This is illustrated in Fig. 4 by the point on each contour of constant L which possesses the highest entropy. Using the new **h**, α is decreased and the process is repeated. The set of final values of **h** for each α is known as the maximum entropy trajectory (marked on Fig. 4), and it is the updated **h** on this trajectory which is displayed by the MaxEnt program at the end of each iteration. To ensure control over the iterative process, changes in α and in the iteration to find maximum $S(\mathbf{h})$ for a given α are made within a 'trust radius', r, which is inversely related to the value of entropy. Consequently, as entropy increases towards zero, steps in α become progressively smaller until there reaches a point or 'stopping criterion' (where α is close to zero) beyond which little new information is obtained about the object function. In the MaxEnt algorithm, the stopping criterion is related to the expectation value of the cross entropy between the exact solution and that actually obtained.

As a qualitative illustration of the way in which the MaxEnt program proceeds along the maximum entropy trajectory towards the most likely object function, we will use as an example the experimentally acquired 300 kV Z-contrast image of a stacking fault in CdTe(001) viewed along [110] shown in Fig. 5(a). The intrinsically intuitive nature of Z-contrast imaging is obvious: both the column-by-column atomic arrangements (dumbbell pairs) and compositions (structural polarity of the lattice) can be deduced from the as-acquired image alone. However, due in large part to high noise levels which result from discarding low-angle signal, measurement of (for example) the mean separation of the dumbbell pairs or of the mean intensities of each column is problematical (A. J. McGibbon *et al.*, 1994).

Calculations of the object function from Fig. 5(a) using a Lorentzian probe distribution are shown in Fig. 5(b) to (g) for 0 iterations (uniform intensity) and 1,3, 5, 10 and 15 iterations, respectively, along the maximum entropy trajectory using the MaxEnt routine. Justification for the choice of probe is made in the next section. The final object function at which the stopping criterion is reached (25 iterations) is shown in Fig. 5(h). In later iterations, gamma functions >1 have been used to aid observation of the intensity distribution. A typical feature of the iteration process is that, when beginning with a uniform object function, gross changes in image intensity (due primarily to changes in film thickness over the region of interest) are accounted for first. As the program proceeds, localization of the object function at column centres becomes increasingly severe until, at the conclusion of the process, the object function resembles an array of delta functions located at the column sites. It can also be seen that, towards the end of the program, changes in the estimated object function become very small. Another noticeable feature of the entire process is that short-range artefacts in the original image (such as scan effects) are quickly discarded by the program, simply because they are too small to have been created by convolution of the object function with the effective probe. It should be stressed that maximum entropy yields only the minimum amount of information that can be reasonably discerned from the evidence.

From the discussion in this section, the power of maximum entropy in the analysis of Z-contrast images is apparent. In future sections we will dwell on interpretation of the calculated object function in a number of experimental and theoretical scenarios. However, in addition to the possible benefits of maximum entropy in image quantification, the nature of the output data is such that a striking enhancement of image quality can be obtained.

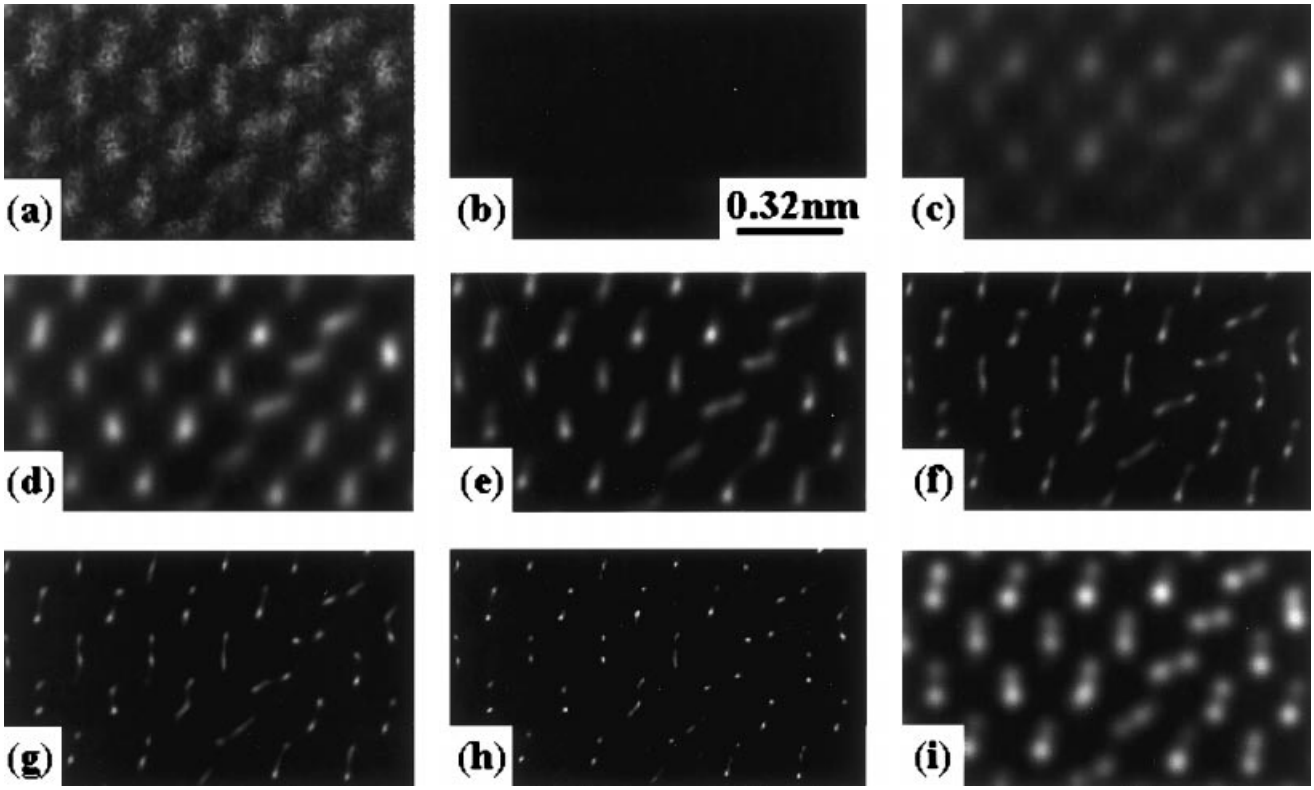

Fig. 5. Maximum entropy analysis of an atomic resolution Z-contrast image (a) of a stacking fault in (001) CdTe viewed along [110]. Output data at 0 (uniform intensity), 1, 3, 5, 10 and 15 iterations of the maximum entropy routine are shown in (b) to (g), respectively, with the maximum entropy object function (25 iterations) shown in (h). In later iterations, gamma functions > 1 have been applied to aid observation of the object function. (i), the maximum entropy image, is a convolution of (g) with a small Gaussian probe.

For example, Fig. 5(i) shows a convolution of the output object function with a small Gaussian distribution. This image, which we shall henceforth describe as a maximum entropy image, clearly demonstrates the salient features of the acquired data such as the atomic arrangements and the structural polarity, whilst retaining quantitative object function information. At a number of points in future sections, we will use a maximum entropy image as part of our discussion.

Development of the maximum entropy technique: simulated image data

Before it is possible to draw conclusions on the accuracy of maximum entropy analyses of experimentally acquired images, it is first necessary to gain an understanding of the capabilities of the analysis technique itself. The primary consideration in this regard is the inherent error involved in the determination of column positions and intensities as a function of probe intensity distribution and image noise.

In the analysis of experimentally acquired images, it is generally not possible to determine the precise current density distribution of the incident electron probe. It is therefore desirable to use a standard probe distribution that can be applied with reasonable certainty to most images. In the search for such a distribution, we will concentrate on the analysis of a simulated Z-contrast image of Si < 110 > at 300 kV that was created using an idealized object function (Fig. 6a); an array of delta functions of equal intensity corresponding to the column positions in a perfect lattice. The Z-contrast image (Fig. 6b) has been formed simply by the convolution of a theoretically calculated probe (Fig. 6a inset) at Scherzer defocus (Colliex & Mory, 1983; Pennycook & Jesson, 1991) and the simulated object function. As an initial 'control experiment' we first applied maximum entropy using the same probe distribution as that employed in the simulation. As would be expected (Fig. 6c), the maximum entropy routine retrieves the original crystal structure with no measurable deviation in either column position or intensity (see also Nellist & Pennycook, 1997). The maximum entropy image resulting from these analyses, together with the probe employed (inset) is shown in Fig. 7(a). The main problem in probe selection for maximum entropy analysis is that in practice (primarily as a result of image noise), secondary maxima in image intensity that are not located at column sites, such as those

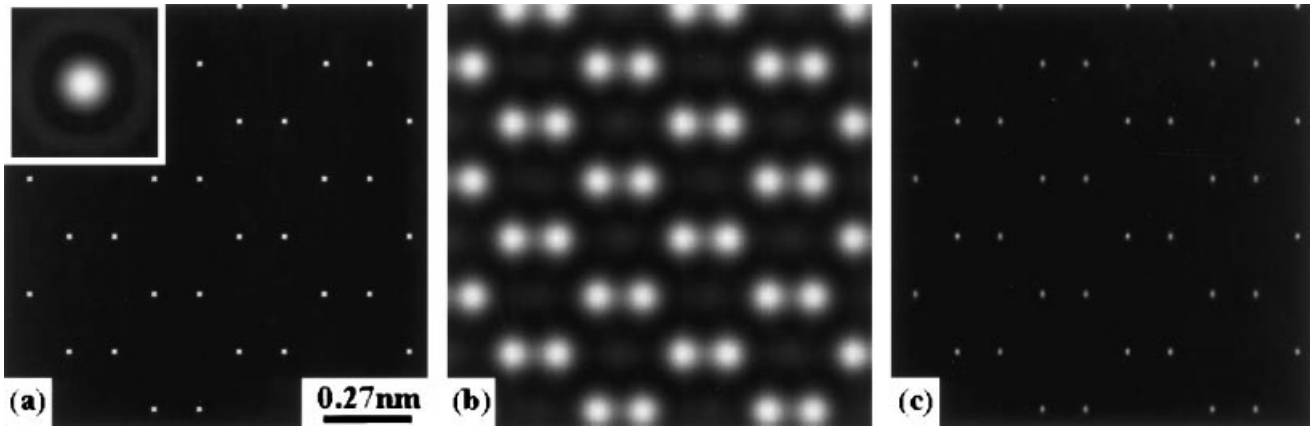

Fig. 6. (a) Simulated Z-contrast object function for Si<110>: an array of delta functions of equal intensity at positions corresponding to the column coordinates in a perfect lattice. Convolution of (a) with the theoretically calculated electron probe distribution at Scherzer defocus (inset and contrast-enhanced to enable observation of the secondary maximum) yields the simulated image shown in (b). (c) Maximum entropy object function of (b). Note the close similarity between (a) and (c).

clearly seen in the simulation, are not always observed. Instead, there exists a diffuse, specimen-dependent background caused by probe 'tails', or by scattering due to an amorphous surface layer on the sample. In such cases, although a Gaussian probe shape (Fig. 7b inset) is a suitable description of the primary peak in probe intensity, maximum entropy analysis using this probe produces anomalous intensity at interstitial positions (Fig. 7b). It is therefore necessary (Fig. 7c) to use a Lorentzian (inset) distribution possessing 'tails' in intensity as a more robust probe approximation. A further advantage of this approach is that the calculated Z-contrast object function is relatively insensitive to Lorentzian probe size. This is illustrated in Fig. 8, in which data from probe sizes of ± 30% of the optimum are compared directly with that of optimum size. In this case, $3\sigma_d$ (where σ_d is the standard deviation of measured column positions in the maximum entropy object function with respect to those of the original object function) lies within 0·04 Å. Full data on the variation between probe size and positional error are shown in Fig. 9. No significant variation in column intensity was found. An important consequence of these results in the general application of maximum entropy to Z-contrast images is that the range of probe sizes with which the routine can operate effectively is sufficiently broad to enable quick and easy selection of probes without significantly reducing the quality of retrieved object function data.

A second potential source of error in the application of maximum entropy is image noise. The nature of STEM is such that it is difficult to produce a standard by which noise

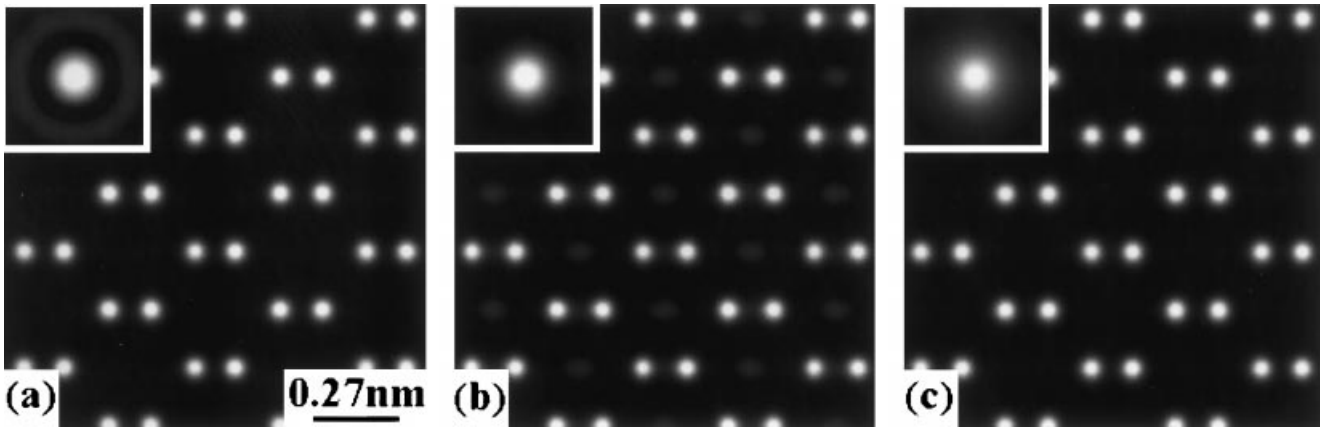

Fig. 7. An illustration of the effect of probe shape on object function retrieval by maximum entropy. (a) Maximum entropy image of the simulated Z-contrast image shown in Fig. 6(b) obtained using the theoretically calculated electron probe distribution at Scherzer defocus (inset). (b) As (a), but obtained using a Gaussian probe approximation (inset). (c) As (a), but obtained using a Lorentzian probe approximation (inset). All images of probe distributions have been contrast-enhanced.

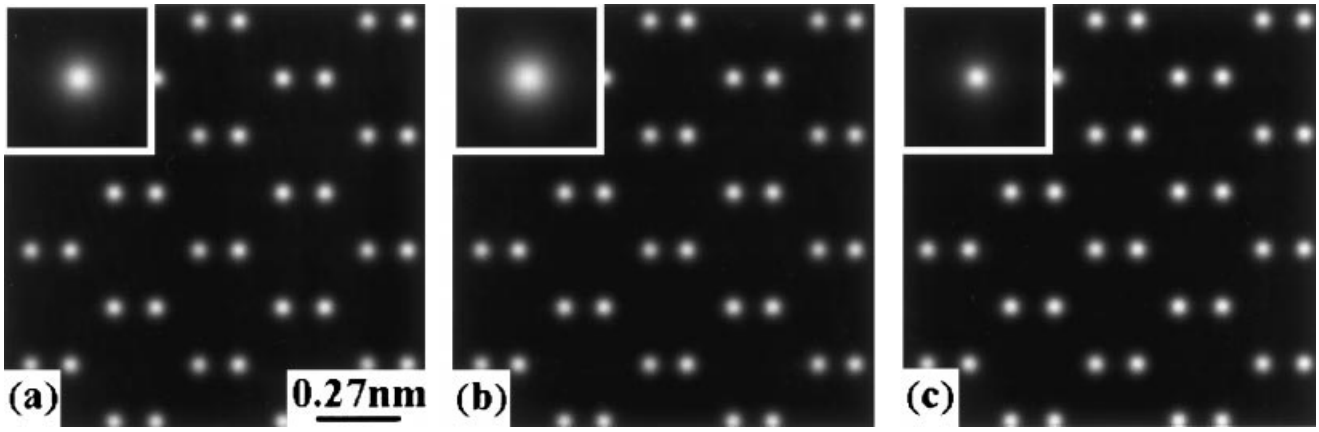

Fig. 8. An illustration of the effect of probe size on object function retrieval by maximum entropy. (a) Maximum entropy image of the simulated Z-contrast image shown in Fig. 6(b) obtained using the optimum Lorentzian probe approximation (inset). (b) As (a), but obtained using a Lorentzian probe (inset) 30% larger than optimum. (c) As (a), but obtained using a Lorentzian probe (inset) 30% smaller than optimum.

can be measured. Loane *et al.* (1988) suggested that, as the standard deviation of the background noise, which is assumed to be entirely shot noise, is proportional to the square root of the background intensity (I_{av}), the visibility V ($0<V<1$) of a particular feature can be expressed as:

$$V = (I - I_{av})/(I_{av})^{1/2} \qquad (11)$$

Where I is the mean intensity of the feature of interest and $I - I_{av}$ is image contrast. If a series of images are normalized to a particular background intensity, then V is proportional to image contrast. Such is the case for the simulated images shown in Fig. 10 to which shot noise (following a Poisson intensity distribution) has been added. Simulations of low ($V=0{\cdot}55$), intermediate ($V=0{\cdot}30$) and high ($V=0{\cdot}15$) noise levels are shown in Fig. 10(a) (b) and (c), respectively, with corresponding maximum entropy images in (d) (e) and (f). It can be seen from this figure that, even with very poor imaging statistics, maximum entropy can retrieve the Z-contrast object function.

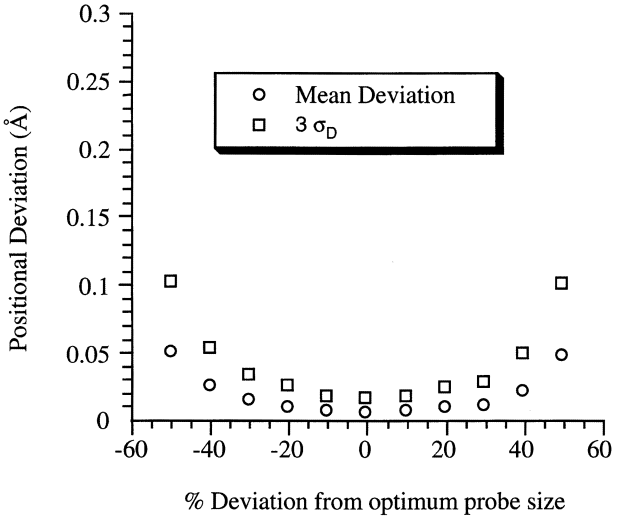

Fig. 9. Graph showing the changes in positional deviation (both mean and $3\sigma_D$) as a function of electron probe (Lorentzian) size employed in the maximum entropy routine. Positional deviation is defined as the difference in column position between the original object function and that retrieved by maximum entropy. Probe sizes are expressed as a percentage of the optimum value. For comparison, the vertical scale in this diagram is the same as that shown in Fig. 11.

Typically, the experimentally acquired images shown in this paper are in the intermediate ($\approx 0{\cdot}2$–$0{\cdot}5$) visibility range. This is subject to the particular imaging conditions employed and material observed. A full summary of the deviations between column positions measured by maximum entropy and those of the original object function as a function of visibility are shown in Fig. 11. With the exception of the most extreme imaging conditions (i.e. $V<<0{\cdot}2$), $3\sigma_D$ lies within $0{\cdot}1$ Å of the original column coordinates. These values, although significantly less than the nearest neighbour separation of columns, are considerably higher than those attributable to variations in probe size. In a manner similar to Fig. 11, a graph showing the deviations, $3\sigma_I$ (expressed as a percentage of the mean value) in measured column intensity from the mean as a function of visibility is shown in Fig. 12. Here, it can be seen that, rather than there being a 'plateau' at which retrieved object functions have similar accuracy, errors decrease as a function of increasing visibility. An interesting consequence of this observation in experimental analysis is that an improvement in the accuracy of column intensity information in the retrieved object function requires a less demanding increase in image quality than that required for an increase in positional accuracy.

The clear conclusion that can be drawn from this section is that, providing imaging statistics are in the regime investigated here, maximum entropy is robust, highly

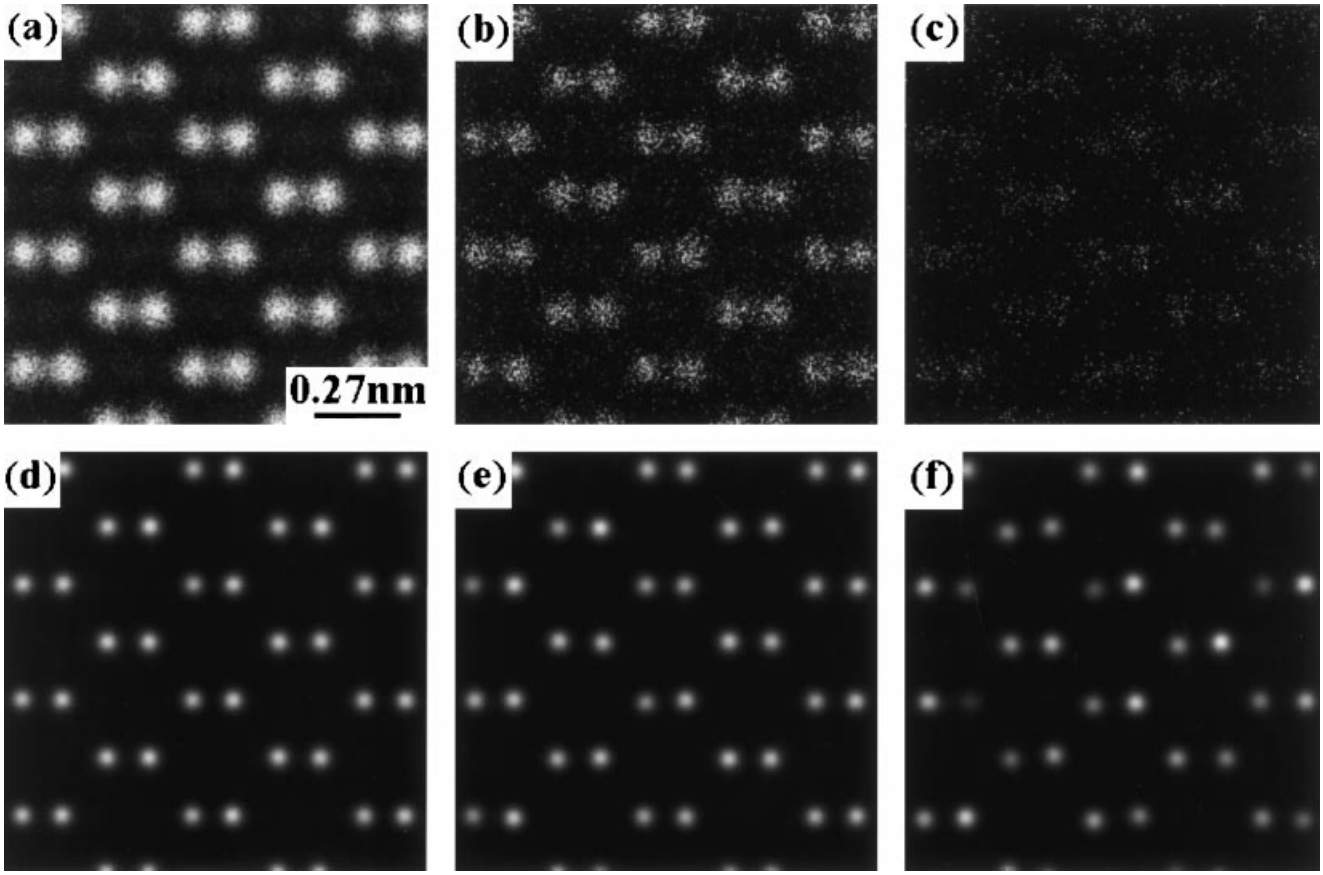

Fig. 10. An illustration of the effect of shot noise on object function retrieval by maximum entropy. Shown in (a) (b) and (c), respectively, are simulations of high ($V = 0{\cdot}55$), intermediate ($V = 0{\cdot}30$) and low ($V = 0{\cdot}15$) visibility, with the corresponding maximum entropy object functions shown in (d) (e) and (f).

accurate and relatively insensitive to image noise and probe size errors.

Development of the maximum entropy technique: experimentally acquired data

As was discussed in previous sections, small misorientations or tilts of the crystal with respect to the incident probe direction may induce a (possibly asymmetric) change in the effective electron probe. However, the effect of such a change on the accuracy of object function retrieval when a symmetric probe is employed is difficult to simulate. In an attempt to address this problem, we have applied maximum entropy to experimentally acquired images of a known crystal structure, the column spacing of which is well within the spatial resolution capabilities of the 300 kV scanning transmission electron microscope: shown in Fig. 13(a) is an as-acquired Z-contrast image of $SrTiO_3$, viewed in the [001] projection, with the corresponding maximum entropy image in 13(e). $SrTiO_3$ has a cubic perovskite structure at room temperature with a lattice parameter of 3·905 Å. In a Z-contrast image, the brighter columns correspond to those of the strongly scattering Sr ($Z = 38$) while the less bright columns correspond to those of Ti–O ($Z = 22$ for Ti, 8 for O). As a result of the very low Z-contrast cross-section, the columns that consist entirely of O are not visible. An additional advantage of employing this material in the investigation of tilt effects is that the effective probe may differ as a function of Z for similar tilt values. Consequently, by recording the positions of both Sr and Ti columns in the maximum entropy object function, such an effect can be explored.

From the retrieved object function of Fig. 13(a), $3\sigma_D$ for the Sr columns was found to be 0·19 Å, whilst that for the less 'visible' Ti columns is 0·24 Å. In the case of experimentally acquired images, the value of σ_D quoted is calculated by comparing the maximum entropy object function with a rigid, idealized lattice that has been moved to a position at which σ_D is minimized. By observation of the selected area diffraction pattern from the region of interest, it was possible to tilt the specimen and acquire Z-contrast images at each tilt value. Shown in Fig. 13(b), (c) and (d),

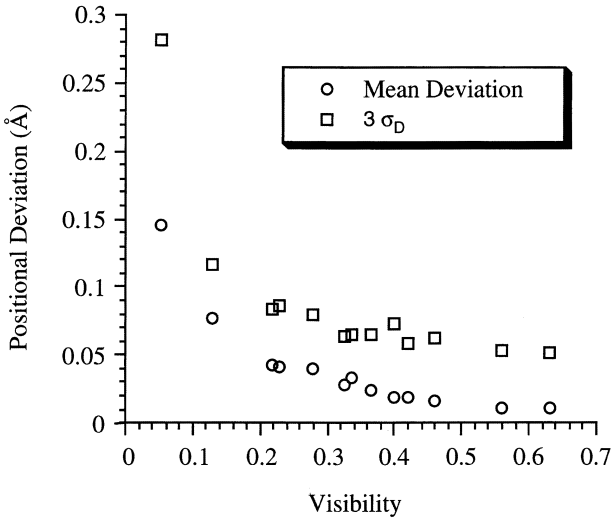

Fig. 11. Graph showing the changes in positional deviation (both mean and $3\sigma_D$) as a function of Z-contrast image visibility when maximum entropy is applied. Positional deviation is defined as the difference in column position between the original object function and that retrieved by maximum entropy. For comparison, the vertical scale in this diagram is the same as that shown in Fig. 9.

respectively, are images taken at tilts of 13·4 mrad along <110>(tilt A), 13·4 mrad along <1$\bar{1}$0>(tilt B) and 19 mrad along <100>(tilt C), with corresponding maximum entropy images shown in (f) (g) and (h). All four

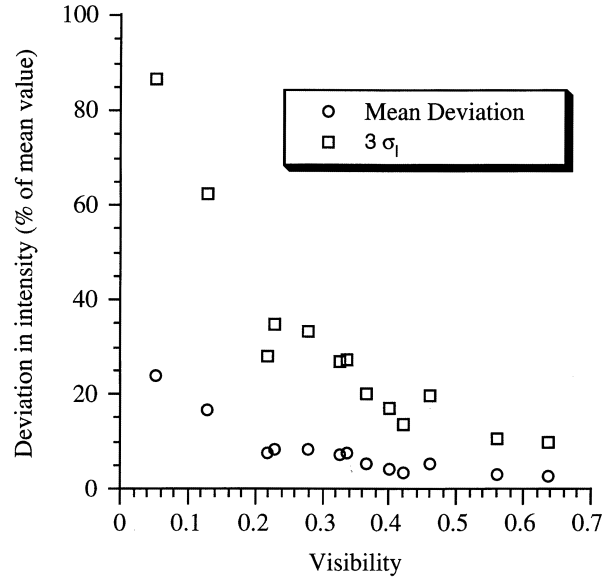

Fig. 12. Graph showing the changes in intensity deviation (both mean and $3\sigma_I$) as a function of Z-contrast image visibility when maximum entropy is applied. Intensity deviation is defined as the difference between the intensity of a particular column and the mean value for all columns in the retrieved maximum entropy object function. For comparison, the vertical scale in this diagram is the same as that shown in Fig. 9.

Z-contrast images in Fig. 13 are of comparable visibility. In the scanning transmission electron microscope, tilting of a specimen onto a particular zone axis can be carried out with an accuracy of approximately ± 2 mrad and so the tilt values used here are considerably more pronounced than those encountered in the study of a specimen orientated along a principal zone axis. The most readily observable feature of Fig. 13 is that, even at a relatively high tilt values, it is possible at tilts A and B to discriminate between columns of Sr and Ti. The extent of dechannelling at tilt C is such that Z-sensitivity is lost. From the object function data, $3\sigma_D$ for Sr for tilts A and B was 0·20 Å and 0·17 Å, respectively, whilst that for Ti was 0·21 Å and 0·25 Å, respectively. $3\sigma_D$ for all columns at tilt C was 0·24 Å. Consequently, the only conclusion that can be drawn from these values is that, to within experimental accuracy, there is no significant increase in positional deviation as a function of tilt for either Sr or Ti columns. It should be noted that such an investigation can only explore anisotropic changes in column position induced by tilt effects. If, for example, tilt produces a systematic shift in all measured column positions, such an effect will not be detected by the methods employed here.

Another phenomenon that may affect the interpretation of the maximum entropy Z-contrast object function is the effect of local strain in the observed specimen. The primary cause of strain is elastic relaxation of the thin film specimen at, for example, heteroepitaxial interfaces, dislocations and grain boundaries. In terms of Z-contrast imaging, local, strain-induced 'bending' and misorientation of atomic columns will result in dechannelling of the incident electron beam, so reducing the measured column intensity. As an example of this effect, a 300-kV Z-contrast image of a $\sum = 5$ [001]{310} grain boundary in $SrTiO_3$ is shown in Fig. 14(a) with the corresponding maximum entropy image in (b). The most readily observable feature in the as-acquired image is the reduction in signal intensity at the grain boundary, primarily attributable to dechannelling of the electron beam as a result of local strain (no changes in film thickness were observed). The effective reduction of column 'visibility' in this region is apparent in the maximum entropy image where it is more difficult to discriminate between columns of Sr and columns of Ti. Such an observation is consistent with the findings of the previous section, suggesting that, although individual identification of columnar composition is problematic, the accuracy with which columnar positions can be measured is similar to that in regions away from the grain boundary. However, the intuitive nature of Z-contrast imaging remains; previous work by M. M. McGibbon *et al.* (1994) on this material carried out in a 100-kV scanning transmission electron microscopy (in conjunction with maximum entropy analysis) has shown that, by

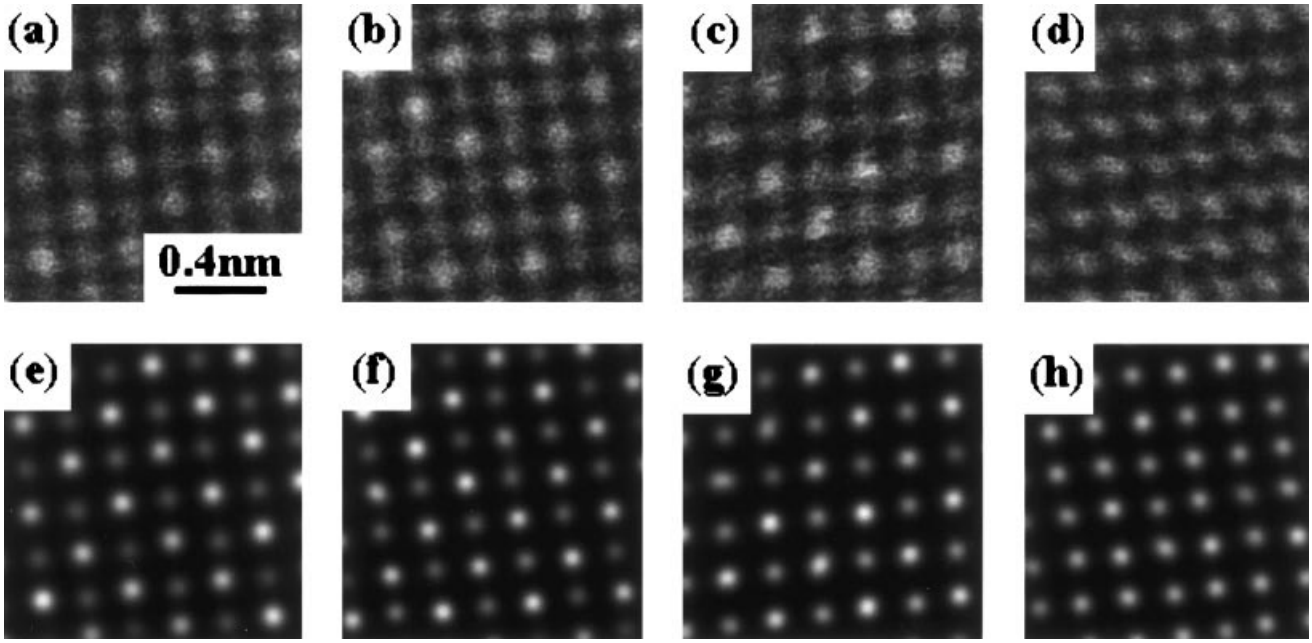

Fig. 13. An illustration of the effect of specimen tilt on object function retrieval by maximum entropy. Shown in (a) to (d) are experimentally acquired Z-contrast images of $SrTiO_3$ viewed in the [001] projection acquired (a) on axis and after tilts of (b) 13·4 mrad along <110> (tilt A); (c) 13·4 mrad along <1$\bar{1}$0> (tilt B); and (d) 19 mrad along <100> (tilt C). Maximum entropy images of (a) to (d) are shown in (e) to (h), respectively.

extrapolating Sr and Ti column positions from the bulk to points in the grain boundary, it is possible to yield anion grain boundary structures directly from the image data.

Maximum entropy analysis of tetrahedral compound semiconductors

Having explored the capabilities of maximum entropy analysis through object function retrieval from both simulated and experimentally acquired data, we now apply the technique to solve unknown structures in III-V and II-VI materials. The most stringent test of Z-contrast imaging at 300 kV is in the analysis of GaAs orientated along the <110> zone axis. Previous work by A. J. McGibbon *et al.* (1994) has demonstrated that, despite the small nearest neighbour separation in the [110] projection (1·41 Å) and small difference in atomic number (Z = 33 for As, 31 for Ga) it is possible to directly observe sublattice polarity using Z-contrast imaging. Such an image is shown in Fig. 15(a). However, as a result of image noise, quantitative analysis of this image is highly problematical. The mean visibility of columns in the Z-contrast image is

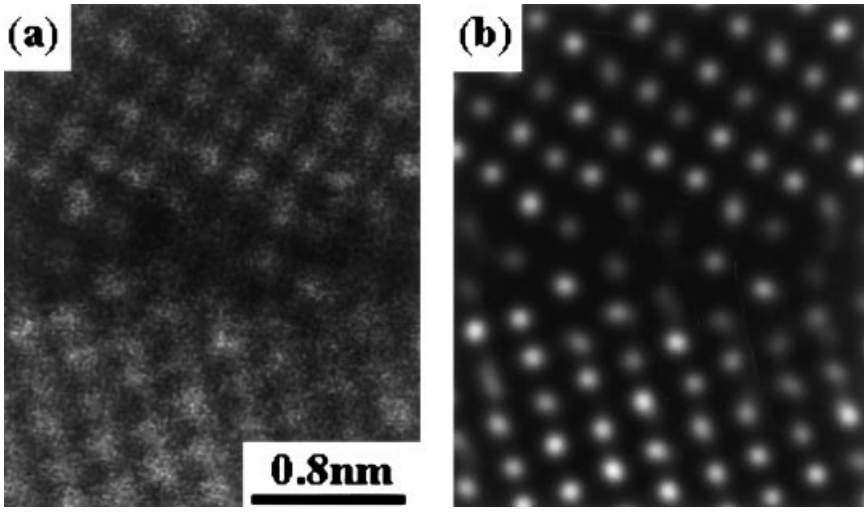

Fig. 14. An illustration of the effect of strain on object function retrieval by maximum entropy. (a) An atomic resolution Z-contrast image of a $\sum = 5$ [001]{310} grain boundary in $SrTiO_3$ taken in the 300 kV scanning transmission electron microscope. Note the decrease in signal intensity at the boundary, primarily attributable to the presence of strain. (b) Maximum entropy image of (a).

0·36, a value that is comparable to the intermediate visibility range in the simulated images of Si. By applying maximum entropy (Fig. 15b) sublattice polarity is clear. Both this image and that in (a) are shown with enhanced contrast. From the retrieved object function, $3\sigma_D$ and $3\sigma_I$ are 0·25 Å and 15% (mean of 3·5%), respectively, for all columns. The mean percentage contrast between columns of Ga and As was 27%. It should be noted that the original image used in these analyses was acquired using a nonzero value of black level (i.e. enhanced contrast) to aid observation of the region of interest. As a result, the measured percentage contrast between As and Ga is considerably higher than the intrinsic Z-contrast ($\sim Z^2$ assuming Rutherford scattering) of 13%. The most likely explanation for the increase in $3\sigma_D$ when compared to that from the simulated images is that, although maximum entropy-induced errors (primarily a result of shot noise) contribute to positional error, errors that are directly attributable to experimental operating conditions such as mechanical and acoustical vibrations and specimen drift are of at least comparable magnitude. Analysis of the measured intensities reveals that although in most cases it is likely that, in a particular dumbbell, the measured As column intensity will be higher than that of Ga, there is a finite probability that a small number of dumbbells will exhibit an anomalous polarity 'inversion'. Thus, to discern local polarity, it generally requires information from more than an individual dumbbell.

By combining a knowledge of the capabilities of the maximum entropy analysis routine with the observations made in the previous paragraph, it is now possible to use the maximum entropy data at atomic resolution in the investigation of unknown atomic structures. Shown in Fig. 16(a) is a maximum entropy image of a Lomer dislocation core viewed along [110] and situated near the interface in the CdTe(001)/GaAs(001) system (Angelo *et al.*, 1993; McGibbon *et al.* 1995a). From the columnar positions alone, it is clear that the core structure is unexpected in that it is asymmetric (best described as a four-fold ring surrounded by five distorted six-fold rings) and unlike the Hornstra model (Hornstra, 1958), a symmetric structure consisting of a seven-fold ring adjacent to a five-fold ring. Shown in Fig. 16(b) is a schematic diagram representing the image in (a) in which column positions have been taken directly from the Z-contrast object function. In addition, atomic species have been assigned to each. The basis for assignation was as follows. Firstly, the interface between GaAs and CdTe could be clearly distinguished by the increase in signal intensity over one monolayer. Secondly, the direction of structural polarity on both sides of the interface was, as in the analysis of Fig. 15, directly observable. As discussed above, there are measured polarity inversions in the maximum entropy image, which, although possibly a real effect, cannot be assumed to be so with any degree of certainty and so are labelled with the same polarity as the surrounding matrix. Finally, at the dislocation core, where it is not possible to assign a particular polarity, the column compositions chosen were those which are intuitively consistent with the atomic arrangements in the surrounding matrix. Although the predicted column compositions from the maximum entropy image may not entirely represent that actually occurring in the specimen, the retrieval of atomic coordinates directly from experimentally acquired data with estimated elemental distributions does represent a firm starting point for further theoretical and experimental investigations into the reasons for such a dislocation core, and the ultimate effect it has on material properties. A full discussion of this and other

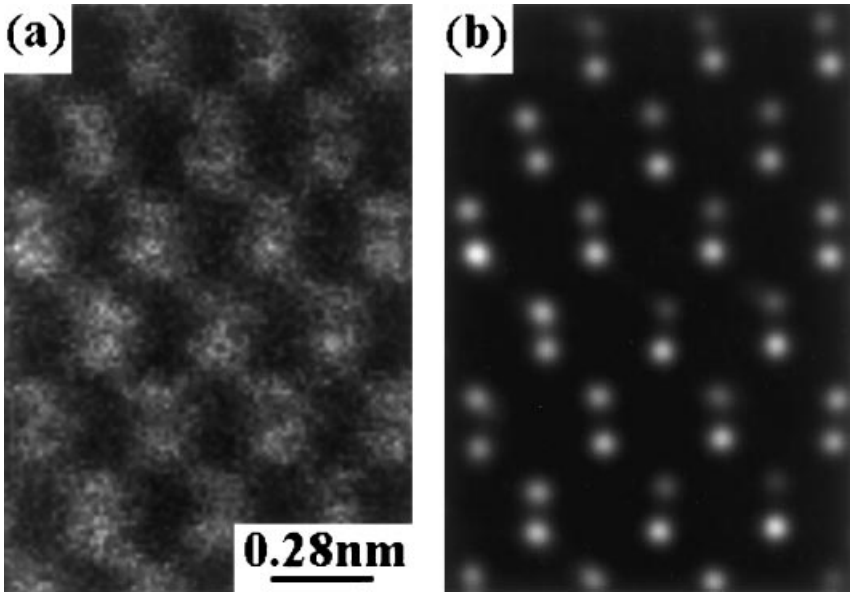

Fig. 15. (a): Atomic resolution Z-contrast image of GaAs viewed in the [110] projection. (b): Maximum entropy image of (a). Note that structural polarity (As columns brighter than Ga) is observable in both images. Both images are shown with enhanced contrast.

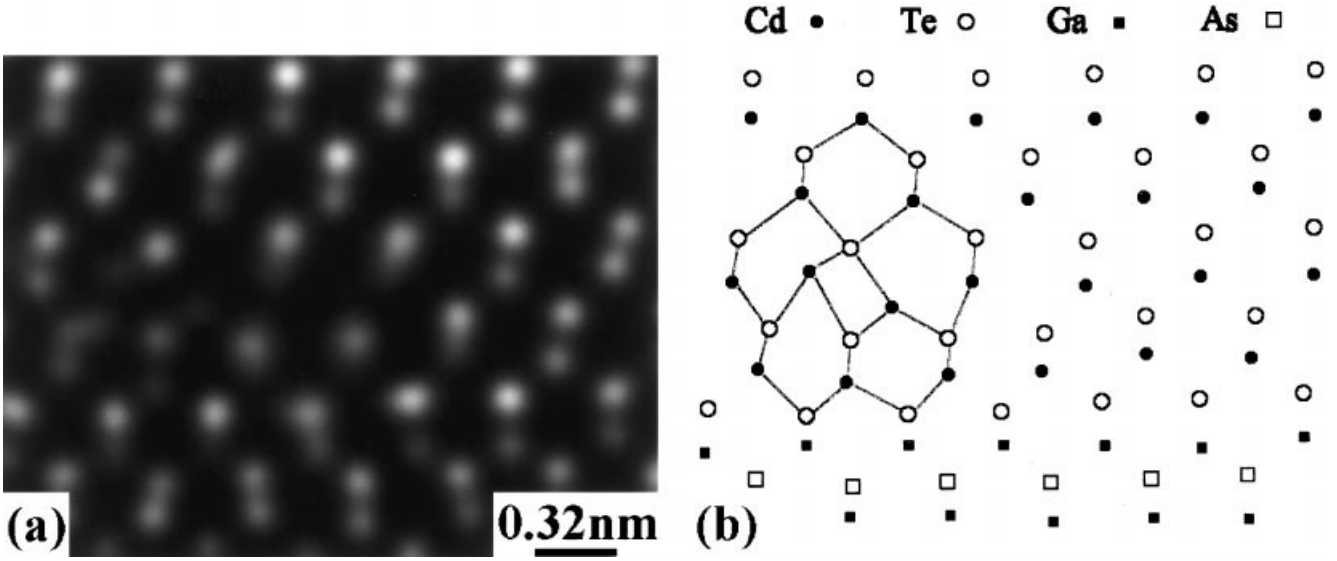

Fig. 16. (a): Atomic resolution Z-contrast image of a Lomer (perfect edge) dislocation near the interface in CdTe(001)/GaAs(001) viewed in the [110] projection. Note the asymmetric core structure. (b) Schematic representation of (a) reached using the information given in the retrieved maximum entropy object function.

structures retrieved from the CdTe(001)/GaAs/(001) system is given by McGibbon *et al.* (1995b).

Conclusions

In this paper, we have demonstrated that the maximum entropy method, when applied to the analysis of incoherent atomic resolution Z-contrast images, is a powerful and robust analytical tool. By retrieving the 'most likely' Z-contrast object function from a particular image, it is possible to reconstruct the projected atomic structure to a level of spatial resolution and compositional sensitivity that cannot be achieved solely from the as-acquired data. Such information is essential in the detailed investigation of the physics and chemistry of specific atomic arrangements at interfaces, grain boundaries and dislocations. Direct structure retrieval can reveal unanticipated structures, leading to new insights into materials properties.

In addition to providing important information on material structures, a detailed understanding of the nature of the maximum entropy object function will, in future, enable investigations into the physics of incoherent image formation in STEM on an experimental level that has not hitherto been possible. Ultimately, maximum entropy analysis may provide a route by which the probe-limited resolution of Z-contrast imaging can be surpassed.

Acknowledgements

We would like to thank T. C. Estes and J. T. Luck for technical assistance. This research was sponsored by the Division of Materials Sciences, US Department of Energy, under contract No. DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp., and supported in part by an appointment to the Oak Ridge National Laboratory Postdoctoral Research Program administered jointly by Oak Ridge National Laboratory the Oak Ridge Institute for Science and Education.

References

Angelo, J.E., Gerberich, W.W., Bratina, G., Sorba, L. & Franciosi, A. (1993) Effects of surface reconstruction on CdTe/GaAs (001) interface structure. *J. Cryst. Growth*, **130,** 459–465.

Batson, P.E., Johnson, D.W. & Spence, J.C.H. (1992) Resolution enhancement by deconvolution using a field emission source in electron energy loss spectroscopy. *Ultramicroscopy*, **41,** 137–145.

Bonavito, N.L., Dorbrand, J.E. & Busse, T. (1993) Maximum entropy restoration of blurred and oversaturated Hubble Space Telescope Imagery. *Appl. Optics*, **32,** 5768–5774.

Colliex, C. & Mory, C. (1983) Quantitative aspects of scanning transmission electron microscopy. *Quantitative Electron Microscopy* (ed. by J.N. Chapman and A.J. Craven), pp. 149–216. SUSSP Publications, University of Edinburgh, U.K.

Farrow, N.A. & Ottensmayer, F.P. (1989) Maximum entropy methods and dark-field microscopy images. *Ultramicroscopy*, **31,** 275–284.

Gull, S.F. & Daniell, G.J. (1978) Image reconstruction from incomplete and noisy data. *Nature*, **272,** 686–690.

Gull, S.F., Livesey, A.K. & Sivia, D.S. (1986) Maximum entropy solution of a small centrosymmetric crystal structure. *Acta Crystallogr.* **A43,** 112–117.

Gull, S.F. & Skilling, J. (1984) Maximum entropy method in image processing. *IEE Proc.* **131F,** 646–659.

Gull, S.F. & Skilling, J. (1990) *Quantified Maximum Entropy MemSys5 User's Manual*, Version 1.1. Maximum Entropy Data Consultants, Meldreth, England.

Hornstra, J. (1958) Dislocations in the diamond lattice. *J. Phys. Chem. Solids*, **5,** 129–141.

Hu, J.J. & Li, F.H. (1991) Maximum entropy image deconvolution in high resolution electron microscopy. *Ultramicroscopy*, **35,** 339–350.

Jesson, D.E. & Pennycook, S.J. (1993) Incoherent imaging of thin specimens using coherently scattered electrons. *Proc. R. Soc. London Ser. A*, **441,** 261–281.

Jesson, D.E. & Pennycook, S.J. (1995) Incoherent imaging of thin specimens using thermally scattered electrons. *Proc. R. Soc. London A*, **449,** 273–293.

Kuzuo, R. & Tanaka, M. (1993) Resolution enhancement of electron energy-loss spectra using the maximum entropy method. *J. Electron Microsc.* **42,** 240–243.

Loane, R.F., Kirkland, E.J. & Silcox, J. (1988) Visibility of single heavy atoms on thin crystalline silicon in simulated annular dark-field images. *Acta Crystallogr.* **A44,** 912–927.

Loane, R.F., Xu, P. & Silcox, J. (1992) Incoherent imaging of zone axis crystals with ADF STEM. *Ultramicroscopy*, **40,** 121–138.

McGibbon, A.J., Chisholm, M.F. & Pennycook, S.J. (1995a) Direct sublattice imaging of semiconductor materials. *J. Vac. Sci. Technol.* **B13,** 1751–1754.

McGibbon, A.J., Pennycook, S.J. & Angelo, J.E. (1995b) Direct observation of dislocation core structures in CdTe/GaAs (001). *Science*, **269,** 519–521.

McGibbon, A.J., Pennycook, S.J. & Wasilewski, Z. (1994) Structural characterization of semiconductor heterostructures by atomic resolution Z-contrast imaging at 300 kV. *Mater. Res. Soc. Symp. Proc.* **326,** 299–304.

McGibbon, M.M., Browning, N.D., Chisholm, M.F., McGibbon, A.J. & Pennycook, S.J.. (1994) Direct determination of grain boundary atomic structure in $SrTiO_3$. *Science*, **266,** 102–104.

Nellist, P.D. & Pennycook, S.J. (1997) Accurate structure determination from image reconstructions in ADF STEM. *J. Microsc.* **190,** 159–170.

Pennycook, S.J. & Jesson, D.E. (1991) High-resolution Z-contrast imaging of crystals. *Ultramicroscopy*, **37,** 14–38.

Pennycook, S.J. & Jesson, D.E. (1992) Atomic resolution Z-contrast imaging of interfaces. *Acta metall. Mater.* **40** (Suppl.), S149–S159.

Pennycook, S.J., Jesson, D.E., Chisholm, M.F., Ferridge, A.G. & Seddon, M.J. (1992) Sub-angstrom microscopy through incoherent imaging and image reconstruction. *Scanning Microscopy*, Suppl. **6,** 233–243.

Sivia, D.S., David, W.I.F., Knight, K.S. & Gull, S.F. (1993) An introduction to Bayesian model selection. *Physica D*, **66,** 234–242.

Smith, D.J., Bursill, L.A. & Wood, G.J. (1985) Non-anomalous high resolution imaging of crystalline materials. *Ultramicroscopy*, **16,** 19–32.