Z-CONTRAST SCANNING TRANSMISSION ELECTRON MICROSCOPY

S. J. PENNYCOOK[1] AND P. D. NELLIST[2]
[1]*Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831-6030, USA.*
[2]*Nanoscale Physics Research Laboratory, School of Physics and Astronomy, The University of Birmingham, Birmingham, B15 2TT, UK.*

1. Introduction

Historically, the development of the transmission electron microscope has followed the path of continually increasing the degree of coherence of the imaging process. This is despite the fact that coherent high resolution images suffer from the phase problem which means they cannot be directly inverted to give the object. Interpretation must necessarily rely on simulation of images of trial objects. Even with the prospect of spherical aberration correction, coherent images will still take many forms depending on objective lens defocus and specimen thickness, and the inversion problem will remain.

In the last decade, electron probes of atomic dimensions have become available in commercial electron microscopes, and make possible the efficient realization of incoherent imaging. Incoherent images have no phase problem, and therefore can be directly inverted to the object without the need for image simulations. Furthermore, the directly interpretable Scherzer resolution limit is significantly higher for incoherent imaging than for coherent imaging, in fact, the information limit is double that of bright field coherent imaging, and in addition the sensitivity to instabilities in defocus and energy is greatly reduced. The incoherent image uses high angle scattering which leads to strong atomic number (*Z*) contrast, and also makes simultaneous electron energy loss spectroscopy (EELS) possible from single atomic columns selected from the image. In the quest for higher resolution to understand the atomic origins of materials properties, incoherent imaging would appear to hold substantial advantages.

In this review we first describe incoherent imaging in light optics, and then the special considerations for incoherent imaging with electrons, which are not absorbed in the sample like photons, but only scattered. It is seen that an annular detector in STEM with a high inner angle provides an almost perfect approximation to incoherent imaging. The image is given by a straightforward convolution of the probe intensity profile with the scattering power of the object. High angle scattering occurs close to the atomic sites, so that the image resolution is dominated by the probe intensity

profile. These characteristics are retained even in the presence of strong dynamical diffraction. Such multiple scattering effects do not alter the *total* intensity falling on the detector, with the result that the image does not show the strong thickness oscillations characteristic of coherent images, and is still given by a simple convolution. The brightness of atomic columns seen in the image is still directly related to their mean square atomic number and the image can still be directly inverted. It is this behavior that makes *Z*-contrast imaging so powerful for structure determination of interfaces and defects such as dislocation cores.

2. Incoherent Imaging with Photons

A perfect incoherent image results from a self-luminous object. As shown in Fig. 1a, each part of the object emits independently, so there are no permanent phase relationships between rays emerging from neighboring parts of the object. After passing through the imaging system, Abbé theory tells us that each point is blurred into an Airy disc, but discs from neighboring points are uncorrelated in phase and no

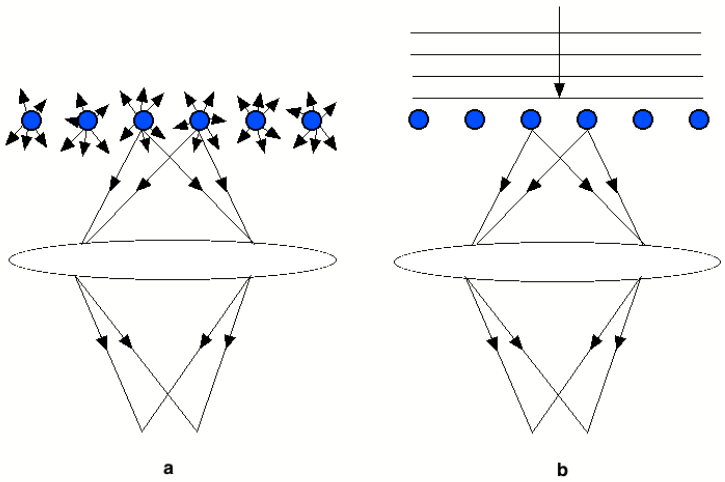

Figure 1. (a) Perfect incoherent imaging of a self-luminous object; (b) perfect coherent imaging of the same object illuminated by a plane wave.

interference is observable on a long time scale. We simply square each Airy disc amplitude $A(\mathbf{R})$ into its intensity $A^2(\mathbf{R})$, and the image is given by the convolution of the object intensity function $O^2(\mathbf{R})$ with $A^2(\mathbf{R})$

$$I(\mathbf{R}) = O(\mathbf{R})^2 * A(\mathbf{R})^2 . \qquad (1)$$

Now if there are permanent phase relationships between nearby parts of the object the situation is completely different. Fig. 1b shows a plane wave illuminating the same

object. Now, we cannot square up the Airy discs from each point because the two Airy discs have a permanent phase relationship which we must know to determine the intensity. The phase problem has appeared; the image is now given by

$$I(\mathbf{R}) = \left(O(\mathbf{R}) * A(\mathbf{R})\right)^2 . \tag{2}$$

Because of the interference we can no longer interpret the image directly in terms of object properties. This was realized over one hundred years ago by Lord Rayleigh (1896). He further realized that by illuminating the object with a large range of angles the phase factors are averaged and an approximation to incoherent imaging can be achieved. In his words "the function of the condenser in microscopic practice is to cause the object to behave, at any rate in some degree, as if it were self-luminous, and thus to obviate the sharply-marked interference bands which arise when permanent and definite phase relationships are permitted to exist between the radiations which issue from various points of the object".

He also realized that incoherent imaging gives significantly better resolution than coherent imaging. For the situation in Fig. 1, clearly the incoherent resolution function $A^2(\mathbf{R})$ must be sharper than $A(\mathbf{R})$ for the coherent mode. Fig. 2 shows Lord Rayleigh's classic result comparing the observation of two point objects illuminated coherently and incoherently. The two point objects are separated so that the first zero in the Airy disc of one coincides with the central maximum of the other, a condition that has become known as the Rayleigh resolution criterion. With incoherent illumination there are clearly two peaks in the intensity distribution and a distinct dip between them; the two objects are just resolved, and the peaks in the image intensity correspond closely with the positions of the two objects. With coherent illumination by a plane wave source the two objects are unresolved.

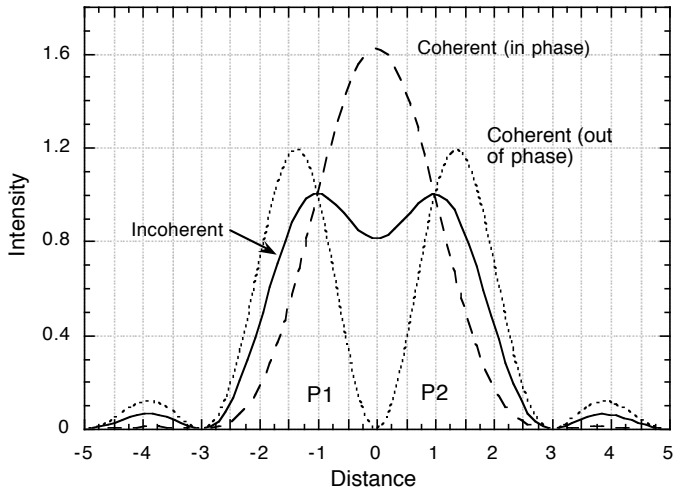

Figure 2. Image intensity for two point objects P1 and P2 illuminated incoherently, coherently in phase, and coherently 180° out of phase, after Lord Rayleigh (1896).

However, if the two objects are illuminated coherently but 180° out of phase, the intensity drops to zero half way between them; they are always resolved whatever their spacing. Unfortunately it is impossible to achieve out of phase illumination for more than one specific image spacing. For example, illuminating from one particular angle will give out of phase illumination for one spatial frequency, but other spatial frequencies will have different phase relationships and show different contrast. Also the two peaks in the image intensity are significantly displaced from their true positions.

Incoherent imaging gives the optimum combination of high resolution with a faithful representation of the object. It allows direct image interpretation, as we do in our everyday lives with our eyes. Our task is to achieve the same ideal of incoherent imaging with electrons.

3. Incoherent Imaging with Electrons

3.1. CTEM OR STEM?

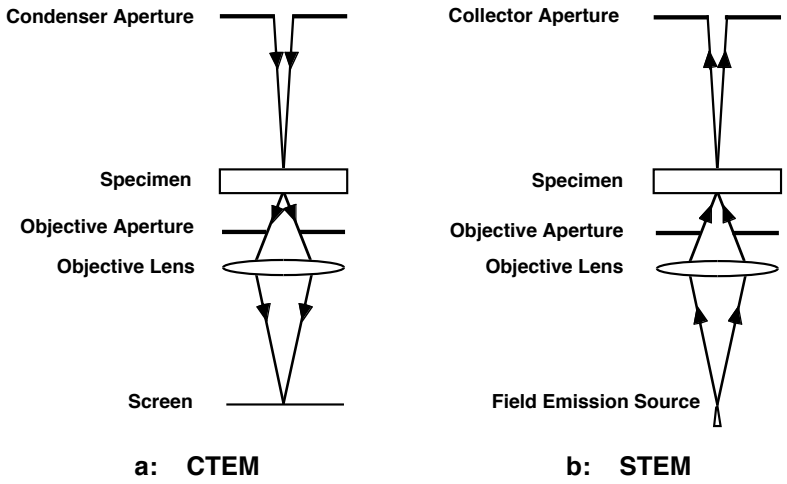

Figure 3. Ray diagrams for coherent bright field imaging in (a) the CTEM and (b) the STEM.

As seen in the ray diagrams of Fig. 3, the essential difference between the conventional TEM and the STEM is the position of the objective lens relative to the specimen. In the CTEM it is used to gather diffracted beams which are brought to a focus on the microscope screen where they interfere to produce the image contrast. The electrons travel from top to bottom in the figure. Not shown are additional projector lenses to provide higher magnification. In Fig. 3b, the optical path of the STEM is shown, with the electrons travelling from bottom to top. A point source is focussed into a small probe by the objective lens, which is placed before the specimen. Not shown are the condenser lenses (equivalent to the CTEM projector lenses) between the source and the

objective lens to provide additional demagnification of the source, and the scan coils to move the probe sequentially from point to point across the specimen and form the image. Transmitted electrons are detected through an angular range defined by the collector aperture. For the small axial collector aperture shown, the two microscopes have identical optics, apart from the fact that the direction of electron propagation is reversed. Since image contrast in the electron microscope is dominated by elastic scattering, no energy loss is involved and time reversal symmetry applies. With equivalent apertures, the image contrast is independent of the direction of electron propagation and the two microscopes are optically equivalent: the STEM bright field image will be the same image, and described by the same imaging theory, as that of a conventional TEM with axial illumination. This is the principle of reciprocity, the original basis for understanding the formation of high resolution lattice images in the STEM (Cowley, 1969, Zeitler and Thomson, 1970).

Now suppose our objects consist of crystals oriented along low index directions, i.e. discrete columns of atoms. Applying the concepts of Lord Rayleigh, whether we have coherent or incoherent imaging depends on whether the Airy discs of neighboring atomic columns have permanent and definite phase relationships between them, which depends on the transverse coherence length at the object. With an axial illumination (or collection) aperture much smaller than a typical Bragg angle, the transverse coherence length will be much longer than interatomic spacings, as shown in Fig. 4a. This is a coherent imaging condition. If however the illumination (or collection) aperture is

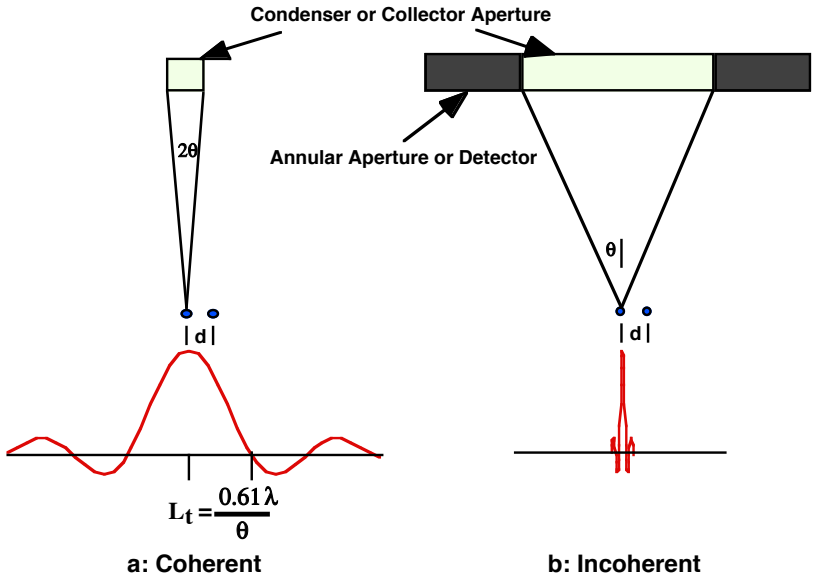

Figure 4. Schematic showing how the transverse coherence length L_t of the condenser aperture (CTEM) or collector aperture (STEM) determines if neighboring atoms are imaged coherently (a) or incoherently (b).

opened up much wider than a typical Bragg angle, then the transverse coherence length at the specimen is much smaller than the atomic separation. No permanent definite phase relationship exists between the electrons illuminating (or detected from) the two columns. This is incoherent imaging.

It would seem that all we need to do is to open up the illumination (or collection) aperture in the CTEM (or STEM) to achieve incoherent imaging, in a way exactly analogous to Lord Rayleigh's condenser lens. A useful criterion for the minimum aperture semiangle θ to achieve incoherent imaging of two objects separated by d is

$$\theta = 1.22\,\lambda/d, \tag{3}$$

where the image intensity varies by less than 5% from the incoherent expectation (Jesson and Pennycook, 1993). In this case the Airy disc coherence envelope of the illumination (collection) aperture is half the width of that of the objective aperture. This condition therefore corresponds to separating the coherence envelopes by double the Rayleigh resolution criterion.

Although optically equivalent, there is a large difference between the CTEM and STEM as regards image efficiency. The objective apertures in both microscopes are similar in size. For coherent imaging the illumination (collection) aperture must be much smaller than the objective aperture, but conversely much larger for incoherent imaging. It creates much less damage in the specimen to illuminate with the small aperture and collect with the large aperture. Illuminating with the large aperture means many more electrons pass through the sample than are collected for imaging, and beam damage is much greater. The CTEM is therefore the natural choice for coherent imaging, whereas the STEM is the instrument of choice for incoherent imaging.

For the incoherent image, better contrast results from the complementary annular dark field detector, because the unscattered beam is then removed from the image. This is particularly important for thin weakly scattering objects. The concept of the annular detector was introduced by Crewe, Wall and Langmore, 1970, and spectacular images of single heavy atoms were obtained (see for example Isaacson, Ohtsuki and Utlaut, 1979). In the field of materials, despite annular detector images showing improved resolution (Cowley, 1986) and theoretical predictions of the lack of contrast reversals (Engel, Wiggins, and Woodruff, 1974), it was generally believed impossible to achieve an incoherent image at atomic resolution. A crucial element in realizing incoherent imaging was the development of the *high angle* annular detector, suggested first as a means of improving the contrast of small catalyst clusters on amorphous or diffracting supports (Treacy, Howie and Wilson, 1978, Howie, 1979 and Treacy, Howie and Pennycook, 1980). The mechanism here was the reduction of coherent diffraction due to the Debye-Waller factor and increased Z-contrast at higher scattering angles.

Only much later was it realized how the detector imposes a narrow coherence envelope which results in almost perfect incoherent imaging. Incoherent images of thick crystalline materials were first reported by Pennycook and Boatner (1988), and the explanation for the incoherent characteristics despite the strong dynamical diffraction followed (Pennycook and Jesson, 1990). The narrow coherence envelope preferentially selects highly localized s-type Bloch states, and reduces the contribution

of less localised states. Inter-column interference is eliminated even in thicker crystals. A detailed discussion of transverse incoherence was given by Jesson and Pennycook (1993) and in Jesson and Pennycook (1995) it was shown how phonon scattering also destroys coherence through the thickness of the sample. Recently a complete mathematical derivation of the dynamical object function has been given by Nellist and Pennycook (1998a).

Finally we note that because the detector aperture can be made much greater than the objective aperture size, the imaging can be made to approach the ideal of perfect incoherent imaging as closely as desired. In contrast, Lord Rayleigh's best condenser lens was equal to the objective lens, and improved incoherence could only be achieved by reducing the objective aperture and degrading the image resolution.

3.2. COMPARISON OF COHERENT AND INCOHERENT IMAGES

The difference between coherent and incoherent characteristics is apparent in the contrast transfer functions of Fig. 5, shown for the VG Microscopes HB603U operating at 300kV with an objective lens C_s of 1mm and optimum conditions for each mode as defined originally by Scherzer (1949). All images presented here were taken with this microscope. For the coherent imaging mode, the contrast oscillates rapidly at high spatial frequencies, and is zero at zero spatial frequency. Because of the reversals, atoms in distorted regions such as grain boundaries may reverse contrast or be absent from a phase contrast image. The incoherent mode has a smoothly decaying positive transfer function which avoids contrast reversals and extends to higher resolution. Transfer is unity at zero spatial frequency, so that image intensities are proportional to thickness (in the absence of multiple scattering).

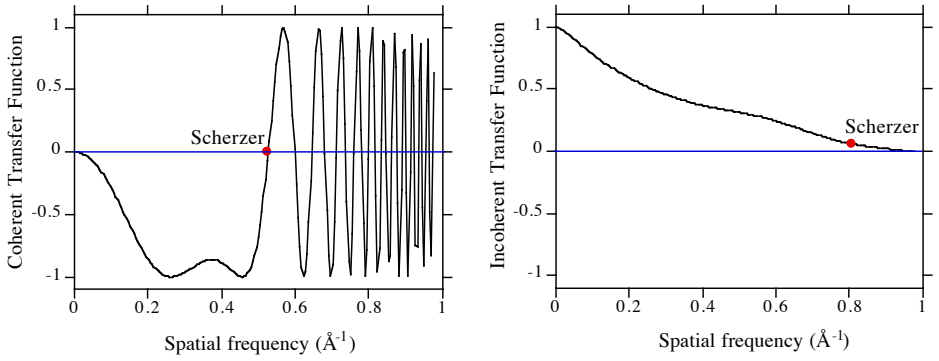

Figure 5. Contrast transfer functions for a 300 kV microscope with an objective lens of 1mm C_s; (a) coherent imaging conditions; (b) incoherent imaging conditions. Curves assume the Scherzer (1949) optimum conditions shown in table 1, (a) defocus -505 Å, (b) defocus -438 Å, aperture cutoff 0.935 $Å^{-1}$.

Both the axial bright field phase contrast image and the *Z*-contrast image can be recorded simultaneously on a STEM, and carry very different information. As an

example, Fig. 6 compares images of an Al-Ni-Co decagonal quasicrystal. This decagonal phase is quasiperiodic in two dimensions but periodic along the third axis, which makes it ideal for electron microscopy. Viewing along the periodic direction, the ten-fold clusters arranged in their quasiperiodic tiling are easily observed in the bright field image. However, the *Z*-contrast image clearly shows improved resolution. Now the central regions of the ten-fold rings are seen to have two different structures, either a disordered ring of almost uniform intensity or much more localised bright columns. This is evidence of chemical ordering. From higher magnification images of the clusters the Al columns could be located and the structure determined (Yan, Pennycook and Tsai, 1998).

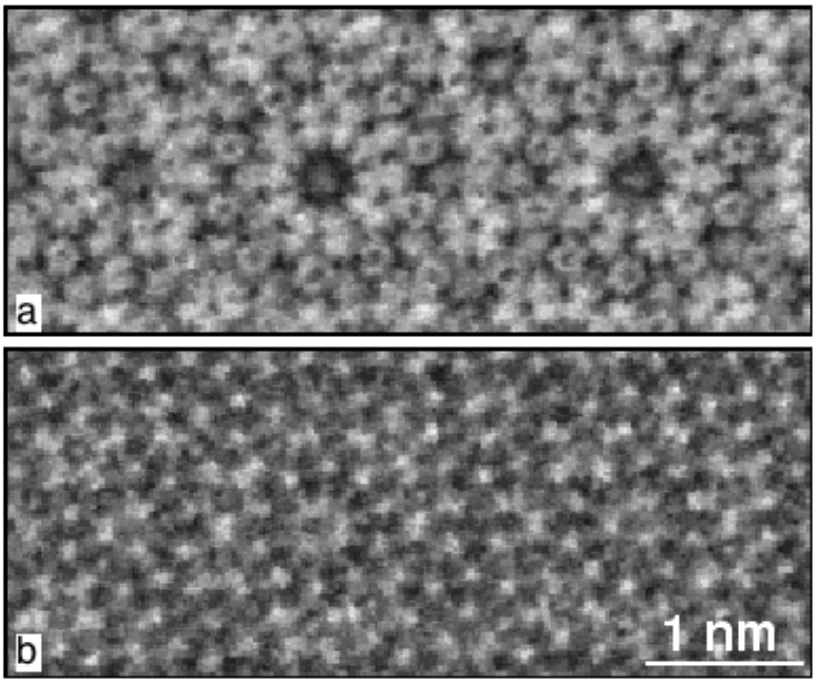

Figure 6. (a) Bright field and (b) *Z*-contrast images of an Al-Ni-Co decagonal quasicrystal taken simultaneously on the STEM showing the increased resolution in the incoherent mode.

The electrical resistivity of amorphous alloys of $Si_{1-x}V_x$ reveal a metal-insulator transition near $x \sim 0.18$ with corresponding changes in electronic structure and atomic structure seen by X-ray and neutron diffraction. Bright field and *Z*-contrast images of two samples above and below the transition point are shown in Fig. 7 (Tanaka et al., 1997). The bright field images show the speckle pattern of the amorphous phase with little structural information, but the *Z*-contrast images show very distinct differences. For $x = 0.12$, small bright clusters ~1-2 nm in size are seen, whereas for $x = 0.30$, a continuous bright region is observed. This striking difference in connectivity was not

apparent from the X-ray or neutron diffraction data, but is clearly the origin of the transition from the insulating to the metallic state.

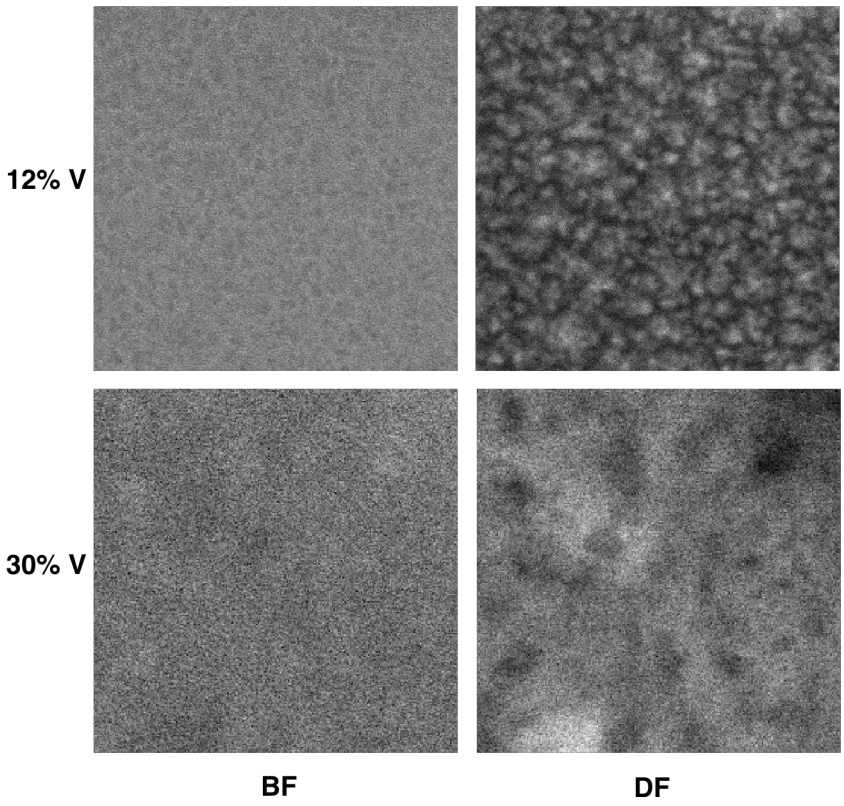

Figure 7. Bright field (left) and dark field *Z*-contrast images (right) of amorphous Si-V alloys. The bright field images show little contrast but the *Z*-contrast images show a change in structure from isolated V-rich clusters to a continuous distribution on increasing the V concentration from 12% (upper) to 30% (lower).

Figure 8 shows a bundle of iodine-intercalated carbon nanotubes (Grigorian et al., 1998). Because of their cylindrical form, only a few atomic layers are parallel to the electron beam. Nevertheless, lattice fringes from the tubes are seen clearly in the phase contrast image because it is tuned to the spacing expected, and filters out the uniform background near zero spatial frequency due to all other atoms. The *Z*-contrast image on the other hand is sensitive to the absolute numbers of atoms under the beam. Where there is no significant dynamical diffraction, it can be considered as an image of projected mass thickness. There is no detectable contrast from the tubes themselves, but the iodine intercalation is clearly visible. Thus the two kinds of images are highly complementary in this case.

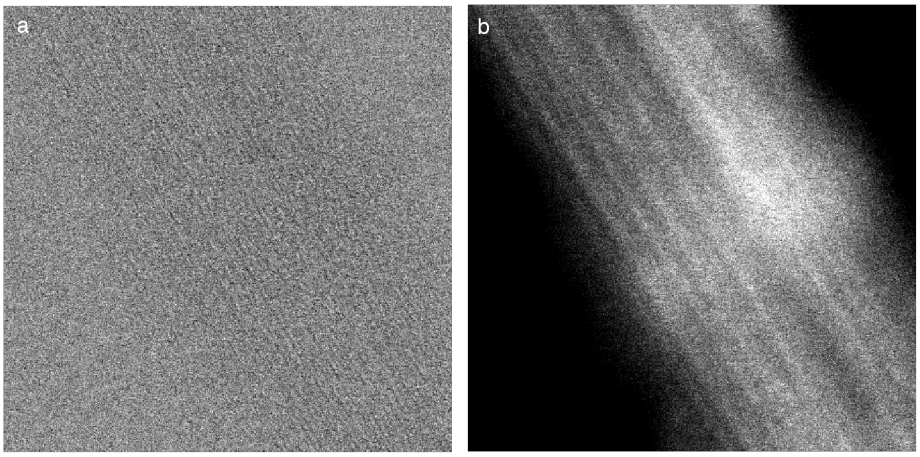

Figure 8. Bright field (a) and *Z*-contrast (b) images of iodine intercalated carbon nanotubes

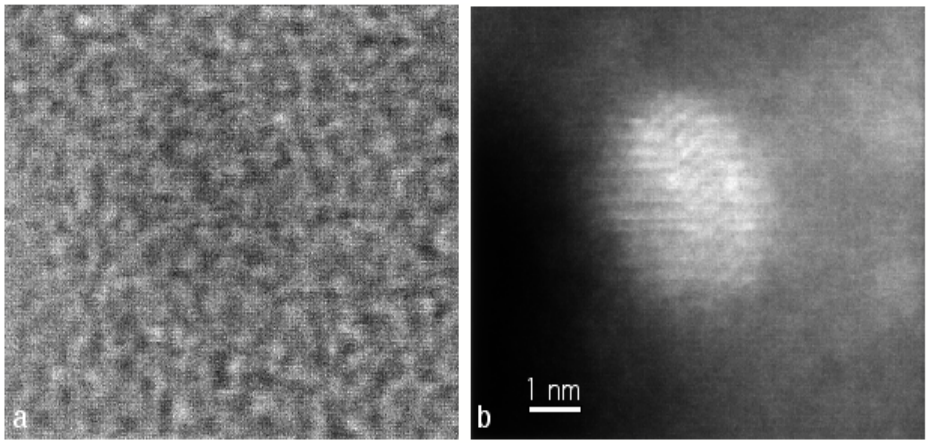

Figure 9. (a) Bright field and (b) *Z*-contrast images of a Rh catalyst particle on γ-alumina.

Fig. 9 shows a Rh catalyst cluster supported on γ-alumina (Pennycook et al., 1996). In this case, the bright field image is dominated by phase contrast from the carbon film ($Z = 6$) used to support the sample, whereas the Rh particle ($Z = 45$) is clearly visible in the *Z*-contrast image. Bright field imaging of small metal clusters becomes very difficult for particles less than about 1 nm in size, due to the inevitable coherent interference effects from the support and the lack of *Z*-contrast (Datye and Smith, 1992).

The ultimate example of *Z*-contrast imaging is the detection of single Pt atoms on γ-alumina shown in Fig. 10 (Nellist and Pennycook, 1996). Here again the two images are very complementary. The orientation of the γ-alumina support can be deduced

from the bright field image, while single atoms, dimers and trimers are detectable in the *Z*-contrast image. Spacings and angles between the Pt atoms are constrained to match the atomic spacings in the γ-alumina surface, suggesting the possible adsorption sites shown in the schematic. The ability to work with insulating, rough substrates of this kind represents an important advantage of STEM compared to atomic force microscopy or scanning tunneling microscopy.

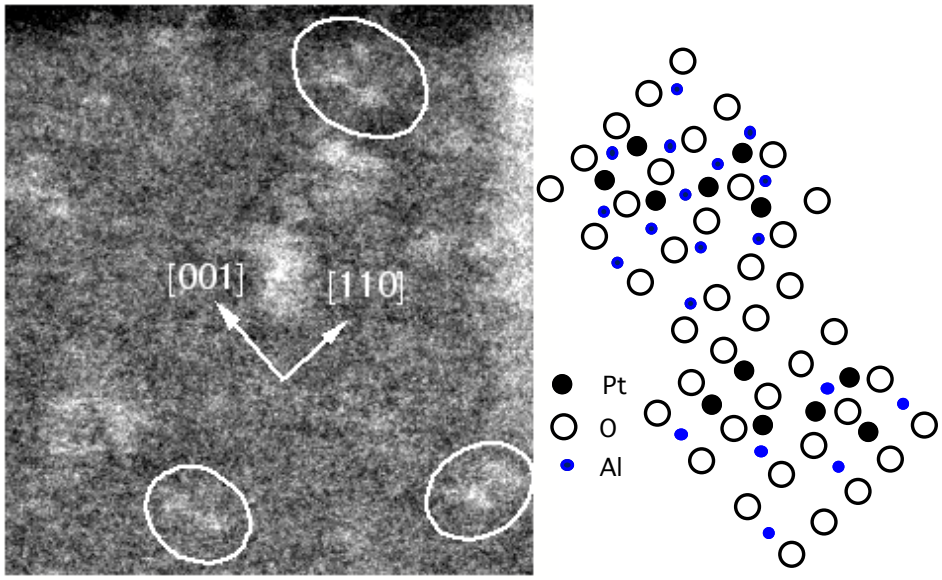

Figure 10. *Z*-contrast image of a Pt catalyst supported on γ-alumina. The spacings and angles between Pt atoms match the orientation of the support, suggesting the possible configurations shown in the schematic. A Pt trimer and two dimers are circled.

3.3. PROBE FORMATION IN STEM

Consider a plane wave filling the objective aperture in the STEM. The probe profile is just the intensity distribution in the object plane, the square of the Airy disc amplitude distribution. Due to spherical aberration and defocus, the phase of each point in the wavefront suffers an aberration $e^{i\gamma}$ as it propagates to the object plane. The probe is therefore best described as a coherent, converging, phase-aberrated spherical wave.

We will denote transverse coordinates by upper case letters, and the beam direction as z, so that positions in the objective back focal plane are labeled by the two-dimensional vector **K**, and positions in real space by (**R**,*z*). Assuming the objective aperture is centered on the optic axis of the lens, $|\mathbf{K}| = \chi\theta$ is the transverse component of the incident electron wavevector $\chi = 2\pi/\lambda$, where λ is the electron wavelength. We will write the objective lens transfer function as

$$A(\mathbf{K}) = H(\mathbf{K})\, e^{i\gamma(K)} \qquad (4)$$

where the amplitude H(**K**) is unity inside the aperture and zero elsewhere. The transfer function phase factor γ is given by

$$\gamma = \frac{\pi}{\lambda}\left(\Delta f\theta^2 + \frac{1}{2}C_s\theta^4\right) = \frac{1}{2\chi}\left(\Delta fK^2 + \frac{1}{2}C_s\frac{K^4}{\chi^2}\right), \tag{5}$$

where C_s is the objective lens spherical aberration coefficient and Δf is the defocus. The amplitude distribution P(**R**) of the STEM probe is obtained by integrating the transfer function over the objective aperture,

$$P(\mathbf{R}) = \int A(\mathbf{K})\, e^{i(\mathbf{K}\cdot\mathbf{R})}\, d\mathbf{K}. \tag{6}$$

Since a shift in real space by **R** is equivalent to multiplication in reciprocal space by $e^{i(\mathbf{K}\cdot\mathbf{R})}$, if the probe is translated to a position $\mathbf{R}_O$ the probe amplitude distribution is given by

$$P(\mathbf{R} - \mathbf{R}_0) = \int A(\mathbf{K})\, e^{i\,\mathbf{K}\cdot(\mathbf{R}-\mathbf{R}_0)}\, d\mathbf{K}. \tag{7}$$

The probe intensity distribution is given by $P^2(\mathbf{R})$. A focal series is shown in Fig. 11 for a 300 kV STEM with C_s = 1mm and an objective aperture of 9.4 mrad. Notice how the intensity profiles are not symmetric with defocus. At low values of defocus the probe is close to Gaussian in nature, while at high defocus values the probe develops a sharper central peak but also a substantial "tail", a subsidiary maximum

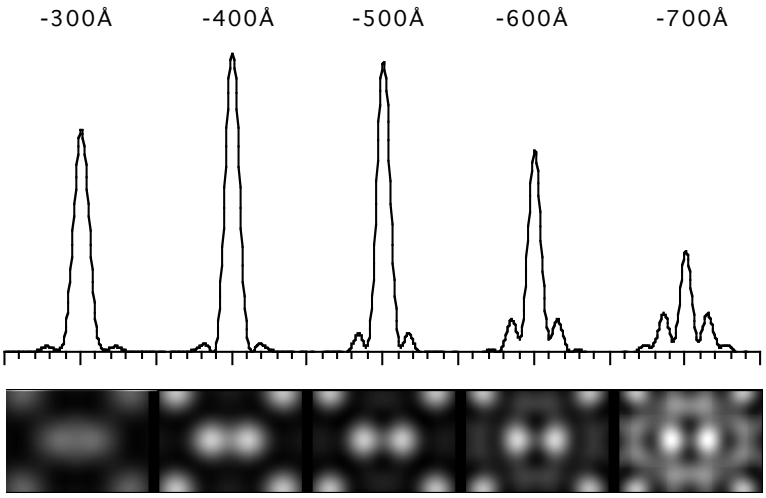

Figure 11. Probe intensity profiles for a 300 kV STEM with C_s = 1mm and a Scherzer optimum objective aperture of 9.4 mrad, with corresponding simulated images of Si(110).

around the central peak. Although the width of the central peak is significantly reduced, actually dropping to below 1 Å at a defocus of -500 Å, now over half the total intensity is in the tail. This gives rise to significant false detail in the image which makes intuitive interpretation no longer straightforward, as shown by the corresponding simulated images of Si⟨110⟩ shown in Fig. 11.

For intuitive image interpretation, the optimum probe is that which gives the narrowest central peak but without significant tails. The situation was analysed by Scherzer in his classic 1949 paper. Although this paper is well known for the so-called Scherzer optimum conditions for bright field phase contrast imaging, it also gives optimum conditions for a wide illumination aperture. They are somewhat different from the coherent conditions, as seen in Table 1, primarily through the use of a slightly smaller aperture to eliminate probe tails. The resulting image resolution is significantly higher than for the coherent case, as expected. The question of the optimum probe has also been discussed in detail by Crewe and Salzman, 1982.

TABLE 1. Comparison of Scherzer optimum conditions for coherent and incoherent imaging

	Coherent Imaging	Incoherent Imaging
Resolution limit	$0.66\ C_s^{1/4}\ \lambda^{3/4}$	$0.43\ C_s^{1/4}\ \lambda^{3/4}$
Optimum aperture	$1.51\ (\lambda/C_s)^{1/4}$	$1.41\ (\lambda/C_s)^{1/4}$
Optimum defocus	$-1.15\ (C_s\lambda)^{1/2}$	$-(C_s\lambda)^{1/2}$

This ideal electron-optical limit is achieved when the objective aperture is illuminated by a plane wave, which requires the size of the geometric image of the source to be zero. This in turn requires infinite demagnification of the source by the condenser and objective lenses resulting in zero current. To compromise we typically work with a source contribution of a few tenths of an angstrom, much smaller than the probe profile. This incoherent broadening of the probe results in less image contrast than would be calculated for the theoretical probe. The size of the broadening can be estimated by convoluting the theoretical probe by a narrow Gaussian, then convoluting with the object function and comparing to the experimental image intensity.

3.4. INCOHERENT IMAGING OF THIN WEAKLY SCATTERING OBJECTS

3.4.1. Detection of all Scattered Electrons

For a thin weakly scattering specimen, we can assume the probe to be a wave packet with amplitude $P(\mathbf{R})$ throughout the thickness of the specimen, and calculate the amplitude Ψ_s scattered into the direction $\mathbf{K}_f$ from the first Born approximation, (Pennycook et al., 1997)

$$\psi_s(\mathbf{K}_f) = c \int e^{-i\mathbf{K}_f \cdot \mathbf{R}} V(\mathbf{R})\, P(\mathbf{R} - \mathbf{R}_0)\, d\mathbf{R}, \tag{8}$$

where $V(\mathbf{R})$ is the projected potential and $c = \frac{m}{2\pi\hbar^2}$. Integrating the scattered intensity $|\psi_s|^2$ over all final states $\mathbf{K}_f$, and using the identity,

$$\int e^{-i\mathbf{K}_f \cdot (\mathbf{R} - \mathbf{R}')}\, d\mathbf{K}_f = (2\pi)^2\, \delta(\mathbf{R} - \mathbf{R}') \tag{9}$$

gives the total scattered intensity as

$$I(\mathbf{R}_0) = \int O(\mathbf{R})\, P^2(\mathbf{R} - \mathbf{R}_0)\, d\mathbf{R}, \tag{10}$$

$$= O(\mathbf{R}) * P^2(\mathbf{R}_0), \tag{11}$$

a convolution of the probe intensity profile $P^2(\mathbf{R})$ with an object function $O(\mathbf{R})$ given by

$$O(\mathbf{R}) = \sigma^2\, V^2(\mathbf{R}), \tag{12}$$

where $\sigma = 2\pi c = \chi/2E$ is the interaction constant. Therefore, provided *all* scattered electrons could be collected, we would have incoherent imaging of the square of the projected potential with a resolution controlled by the incident probe intensity profile. The problem lies in the fact that many of the scattered electrons lie within the cone of the incident beam and cannot be distinguished from the unscattered electrons. Detecting the total transmitted intensity would give no contrast, as electrons are not absorbed just scattered. This led to the so-called "hole-in-the-detector" problem; by cutting a hole in the annular detector large enough for the incident beam to pass through, not all the scattered electrons could be collected. The problem is minimized by using a small objective aperture, but then only low resolution would be obtained. So it was commonly believed that incoherent imaging was impossible at atomic resolution (Cowley, 1976, Ade, 1977). The resolution of the problem is to use a hole in the detector that is *large* compared to the objective aperture, as discussed in section 3.1 above. The theory for this is presented next.

3.4.2. Detection of High Angle Scattered Electrons

We include the detector through the function $D(\mathbf{K}_f)$, which is unity over the detector and zero elsewhere, so that the total detected intensity is now

$$I(\mathbf{R}_0) = c^2 \iiint e^{-i\mathbf{K}_f \cdot (\mathbf{R}-\mathbf{R}')} V(\mathbf{R}) V^*(\mathbf{R}') P(\mathbf{R}-\mathbf{R}_0) P^*(\mathbf{R}'-\mathbf{R}_0) D(\mathbf{K}_f)\, d\mathbf{R}\, d\mathbf{R}'\, d\mathbf{K}_f, \tag{13}$$

which can be integrated over $\mathbf{K}_f$ to give the image in terms of a real space detector function $D(\mathbf{R}'')$, where $\mathbf{R}'' = \mathbf{R} - \mathbf{R}'$,

$$I(\mathbf{R}_0) = \sigma^2 \iint V(\mathbf{R}''+\mathbf{R}')V^*(\mathbf{R}')P(\mathbf{R}''+\mathbf{R}'-\mathbf{R}_0)P^*(\mathbf{R}'-\mathbf{R}_0)D(\mathbf{R}'')d\mathbf{R}'d\mathbf{R}''. \quad (14)$$

Now it can be seen that if the detector function in real space is narrow on the scale of the probe, as it is for a high angle annular detector, then $P(\mathbf{R}''+\mathbf{R}'-\mathbf{R}_0)$ will be practically constant during the $\mathbf{R}''$ integration and the integral can be separated to give

$$I(\mathbf{R}_0) = \sigma^2 \int V(\mathbf{R}''+\mathbf{R}')D(\mathbf{R}'')d\mathbf{R}'' \int V^*(\mathbf{R}')P(\mathbf{R}'-\mathbf{R}_0)P^*(\mathbf{R}'-\mathbf{R}_0)d\mathbf{R}'. \quad (15)$$

Performing the $\mathbf{R}''$ integration convolutes the detector function with the potential in real space. As this is equivalent to multiplication in reciprocal space, it is clear that the detector selects only the high order Fourier components of the potential, i.e. it acts as a high pass filter. The image is now given by

$$I(\mathbf{R}_0) = \sigma^2 \int [V^2(\mathbf{R}')*D(\mathbf{R}')]P^2(\mathbf{R}'-\mathbf{R}_0)d\mathbf{R}', \quad (16)$$

which is again in the form of a convolution, so we regain incoherent imaging with

$$I(\mathbf{R}_0) = O(\mathbf{R}) * P^2(\mathbf{R}_0), \quad (17)$$

but the object function is now high pass filtered,

$$O(\mathbf{R}) = \sigma^2 [V^2(\mathbf{R})*D(\mathbf{R})]. \quad (18)$$

As before, if we have a bright field detector, $D(\mathbf{K}_f) = \delta(0)$, then $D(\mathbf{R}'') = 1$, and we cannot separate Eqn. 14. The bright field image is given by

$$I_{BF}(\mathbf{R}_0) = \sigma^2 |V(\mathbf{R}) * P(\mathbf{R}_0)|^2, \quad (19)$$

and we now have a phase problem.

3.4.3. Image Contrast

The origin of the image contrast can be seen by inserting Eqn. 7 for the probe into Eqn. 13 for the intensity on the detector,

$$I(\mathbf{R}_0) = c^2 \iiint e^{-i\mathbf{K}_f\cdot(\mathbf{R}-\mathbf{R}')}V(\mathbf{R})V^*(\mathbf{R}')\int A(\mathbf{K})\, e^{i\,\mathbf{K}\cdot(\mathbf{R}-\mathbf{R}_0)}\, d\mathbf{K}$$
$$\times \int A^*(\mathbf{K}')e^{-i\,\mathbf{K}'\cdot(\mathbf{R}'-\mathbf{R}_0)}d\mathbf{K}'\, D(\mathbf{K}_f)d\mathbf{R}d\mathbf{R}'d\mathbf{K}_f. \quad (20)$$

Now the variables $\mathbf{K}$ and $\mathbf{K}'$ refer to the incident cone of illumination. Collecting terms in $\mathbf{R}$ and $\mathbf{R}'$ we have

$$I(\mathbf{R}_0) = c^2 \iint e^{-i(\mathbf{K}_f-\mathbf{K})\cdot\mathbf{R}} V(\mathbf{R})\, d\mathbf{R} \int e^{i(\mathbf{K}_f-\mathbf{K}')\cdot\mathbf{R}'} V^*(\mathbf{R}')d\mathbf{R}'$$
$$\times \int A(\mathbf{K})e^{-i\mathbf{K}\cdot\mathbf{R}_0}d\mathbf{K} \int A^*(\mathbf{K}')e^{i\mathbf{K}'\cdot\mathbf{R}_0}d\mathbf{K}'D(\mathbf{K}_f)d\mathbf{K}_f. \quad (21)$$

Now the $\mathbf{R}$ and $\mathbf{R}'$ integrations select the $\mathbf{K}_f$-$\mathbf{K}$ and $\mathbf{K}_f$-$\mathbf{K}'$ Fourier components of the potential, and the image is given by

$$I(\mathbf{R}_0) = \sigma^2 \int V(\mathbf{K}_f-\mathbf{K})V^*(\mathbf{K}'-\mathbf{K}_f) \int A(\mathbf{K})e^{-i\mathbf{K}\cdot\mathbf{R}_0}d\mathbf{K} \int A^*(\mathbf{K}')e^{i\mathbf{K}'\cdot\mathbf{R}_0}d\mathbf{K}'D(\mathbf{K}_f)d\mathbf{K}_f. \quad (22)$$

This expression is simplified considerably by taking the Fourier transform with respect to $\mathbf{R}_0$ (Nellist and Pennycook, 1998b). Defining a spatial frequency $\boldsymbol{\rho} = \mathbf{K}'-\mathbf{K}$, we obtain

$$I(\boldsymbol{\rho}) = \sigma^2 \iint V(\mathbf{K}_f-\mathbf{K})V^*(\boldsymbol{\rho}-\mathbf{K}_f+\mathbf{K})A(\mathbf{K})A^*(\mathbf{K}+\boldsymbol{\rho})d\mathbf{K}D(\mathbf{K}_f)d\mathbf{K}_f. \quad (23)$$

We can see that the intensity of the image spatial frequency $\boldsymbol{\rho}$ depends on a product of the probe-forming aperture functions with incident direction separated by $\boldsymbol{\rho}$. Clearly the contrast falls to zero if $\boldsymbol{\rho}$ exceeds the objective aperture diameter. Contrast comes only from regions of overlapping discs on the detector, as shown in Fig. 12a (Spence and Cowley, 1978). This is also the origin of the factor of two improvement in resolution for incoherent imaging compared to axial bright field imaging. In the bright field case no disc overlaps are detected if $\boldsymbol{\rho}$ exceeds the objective aperture *radius*, as shown in Fig 12a. The bright field STEM resolution limit is therefore the same as for bright field imaging in CTEM, as expected by reciprocity.

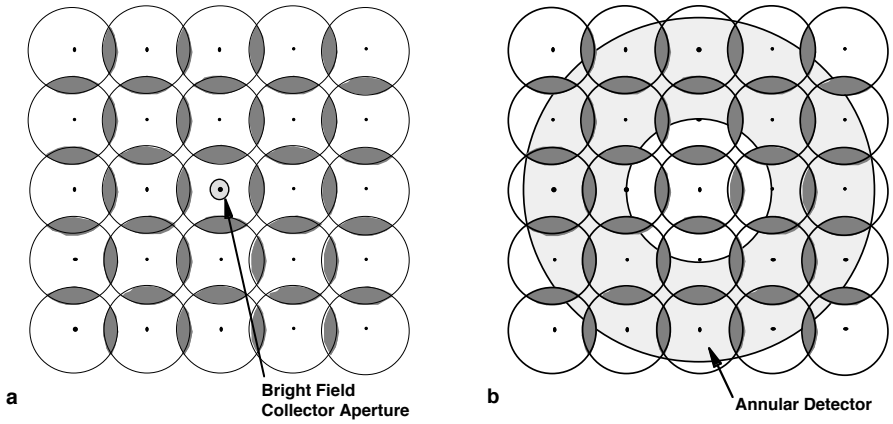

Figure 12. Diffraction pattern in the detector plane from a simple cubic crystal of spacing such that the angle between diffracted beams is greater than the objective aperture radius. (a) An axial bright field detector shows no contrast, while in (b), regions of overlapping discs on the annular detector produce atomic resolution in the incoherent image.

Now if the detector is large compared to the range of incident wavevectors **K**, we can ignore the **K** dependence of the potential coefficients in Eqn. 16 and separate it into

$$I(\boldsymbol{\rho}) = \sigma^2 \int V(\mathbf{K}_f)V^*(\boldsymbol{\rho} - \mathbf{K}_f)D(\mathbf{K}_f)d\mathbf{K}_f \int A(\mathbf{K})A^*(\mathbf{K}+\boldsymbol{\rho})d\mathbf{K}. \quad (24)$$

Now we have again separated the image into a contribution dependent only on the specimen, and one dependent on the probe, and achieved incoherent imaging. It is described in reciprocal space as,

$$I(\boldsymbol{\rho}) = O(\boldsymbol{\rho})\, t(\boldsymbol{\rho}) \quad (25)$$

where

$$O(\boldsymbol{\rho}) = \sigma^2 \int V(\mathbf{K}_f)V^*(\boldsymbol{\rho} - \mathbf{K}_f)D(\mathbf{K}_f)d\mathbf{K}_f \quad (26)$$

is the Fourier transform of the object function $O(\mathbf{R})$ in Eqn. 18, and

$$t(\boldsymbol{\rho}) = \int A(\mathbf{K})A^*(\mathbf{K}+\boldsymbol{\rho})d\mathbf{K} \quad (27)$$

is the Fourier transform of the probe intensity profile $P^2(\mathbf{R})$, which acts as the transfer function for incoherent imaging. Transfer functions for the optimum Scherzer conditions are shown in Fig. 13, corresponding to the profiles shown in Fig. 11. Also shown is the ideal transfer function without any aberration. The approximately triangular shape reflects the decreasing disc overlap with increasing spatial frequency. Increasing the defocus is seen to enhance the high spatial frequencies but reduces the transfer at lower frequencies. In all cases the transfer reaches zero at the cutoff defined by the aperture. If the aperture size is increased, the transfer function can be extended; Fig. 13b shows transfer functions obtained with a 13 mrad objective aperture, corresponding to an aperture cutoff of 0.74 Å. At high defocus values, significantly improved transfer is obtained at spatial frequencies near and beyond the cutoff of the smaller aperture. The curves are shown on the same vertical axis. Note that the absolute transfer at zero spatial frequency scales as the square of the aperture diameter, i.e. is proportional to beam intensity. At lower defocus values transfer is very similar to that with the optimum aperture and an intuitive image is expected. Contrast will be reduced however compared to the optimum aperture because of the much larger uniform background.

Experimental verification of enhanced transfer under such conditions is seen in Fig. 14, which shows images of Si ⟨110⟩ obtained with a 17 mrad objective aperture (Nellist and Pennycook, 1998c). Under optimum defocus the dumbbells are well resolved (Fig. 14a) with maxima close to the atomic positions, the expected intuitive image. The Fourier transform of the image intensity, Fig. 14b shows the spatial frequencies being transferred, which includes the {004} reflection at a spacing of 1.36 Å. On increasing the objective lens defocus, Fig. 14c, the image is no longer intuitive but many additional spots are seen in the Fourier transform (Fig. 14d). The {444} spot at 0.78 Å is the highest resolution so far achieved in an electron microscope. This result

demonstrates very clearly the improved resolution available with incoherent imaging, as first pointed out by Lord Rayleigh.

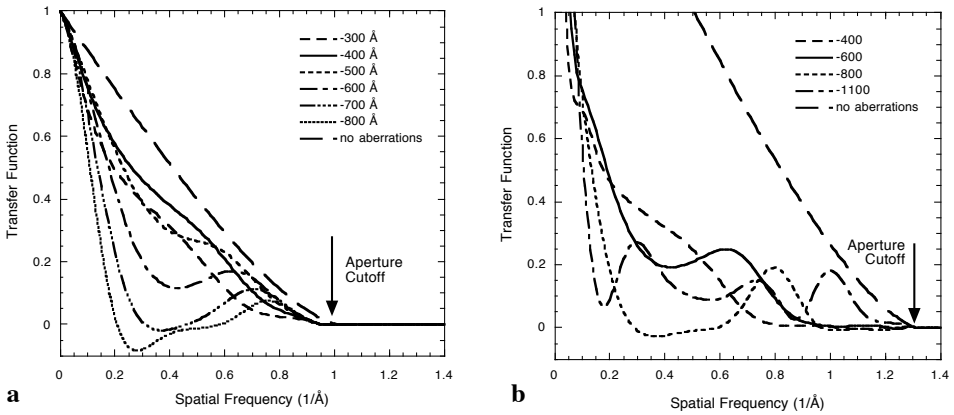

Figure 13. (a) Transfer functions for a 300 kV STEM with C_s = 1mm and a Scherzer optimum objective aperture of 9.4 mrad; (b) extended transfer obtained with an oversized 13 mrad objective aperture.

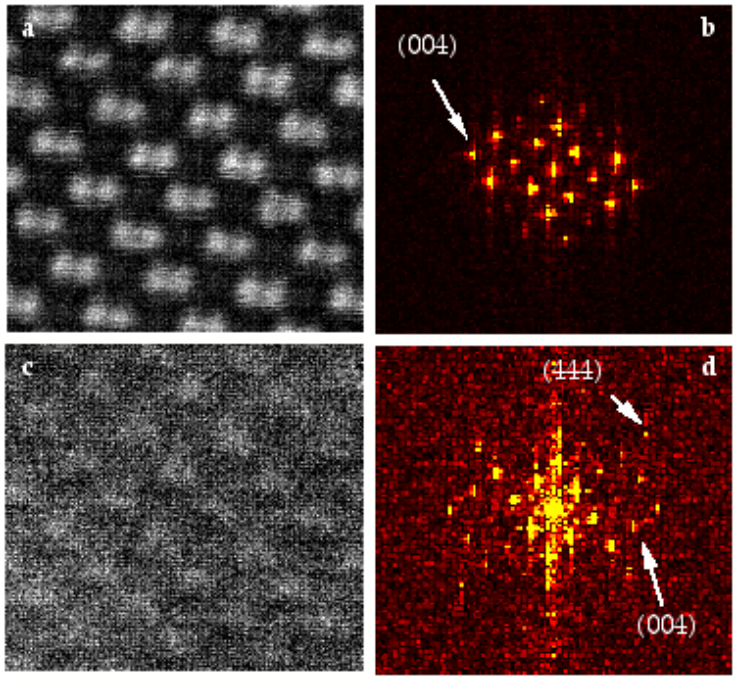

Figure 14. Images and corresponding Fourier transforms of Si ⟨110⟩ obtained with a 17 mrad objective aperture; (a,b) at Scherzer defocus passing the {004} spacing and resolving the dumbbell; (c,d) using a high underfocus giving increased transfer at high spatial frequencies, including the {444} spacing at 0.78 Å.

3.5. INCOHERENT IMAGING OF THICK WEAKLY SCATTERING OBJECTS – z-COHERENCE

Because of the short wavelength of high energy electrons, a short transverse coherence length requires an inner detector angle of only a few degrees. The phase differences between atoms separated along the beam direction will be much shorter than between atoms separated laterally, as shown in Fig 15. Atoms spaced by x and z in the transverse and longitudinal directions respectively, when viewed from direction θ have phase differences of $x\sin\theta$ and $z(1-\cos\theta)$ respectively. For small scattering angles these are approximately $x\theta$ and $z\theta^2/2$, suggesting that as atoms are formed into columns, although columns separated laterally will be imaged incoherently, all atoms in an individual column will scatter coherently. Useful insights into the physics of this situation can be seen if we ignore temporarily the complications of dynamical scattering and treat an entire column as a weakly scattering object. We will see that it is thermal vibrations that break coherence in the z-direction, by inducing *transverse* displacements comparable to the detector coherence envelope.

3.5.1. Kinematical scattering from a column of atoms

We now need to include the z dimension in calculating the amplitude scattered to each final state $\mathbf{K}_f$. Ignoring probe dispersion, Eqn. 8 is replaced by

$$\psi_s(\mathbf{R}_0,\mathbf{K}_f) = c \int e^{-i\mathbf{k}_f\cdot\mathbf{r}} V(\mathbf{r}) P(\mathbf{R}-\mathbf{R}_0) d\mathbf{r}. \quad (28)$$

Separating the transverse and longitudinal directions,

$$\psi_s(\mathbf{R}_0,\mathbf{K}_f) = c \int e^{-i\mathbf{K}_f\cdot\mathbf{R}} V(\mathbf{R}) P(\mathbf{R}-\mathbf{R}_0) d\mathbf{R} \times \int e^{-ik_z z} dz, \quad (29)$$

where k_z is the difference in z components of the incident and scattered wave vectors due to the curvature of the Ewald sphere, and $V(\mathbf{R})$ is the projected potential per unit thickness. The integration over z gives the usual kinematic shape factor for a crystal of thickness t,

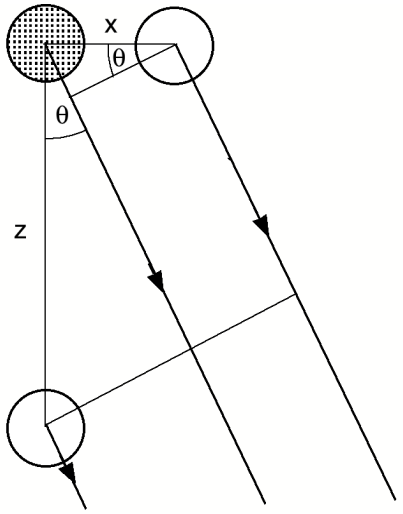

Figure 15. Path differences between scattering from atoms separated in the transverse and longitudinal directions.

$$\int e^{-ik_z z} dz = \frac{\sin(k_z t/2)}{k_z t/2}, \quad (30)$$

where

$$k_z = \chi(1-\cos\theta) \approx K_f^2/2\chi. \quad (31)$$

Because of this shape factor the detector will now be covered with a system of thickness fringes, becoming denser with increasing scattering angle and with thicker crystals. The detected intensity is now oscillatory with thickness, as expected with coherent scattering (Jesson and Pennycook, 1993).

3.5.2. The Role of Thermal Vibrations

We now look at the role of phonons in breaking the coherent integration over thickness in Eqn. 21. We consider each atom to have an instantaneous displacement due to thermal motion of $\mathbf{u} = (\mathbf{U}, u_z)$ and take the time average to find the scattered intensity. The general expression for the intensity detected from a vibrating crystal is

$$I(\mathbf{R}_0) = c^2 \iiint e^{-i\mathbf{K}_f\cdot(\mathbf{R}-\mathbf{R}')} V(\mathbf{R}) V^*(\mathbf{R}') P(\mathbf{R}-\mathbf{R}_0) P^*(\mathbf{R}'-\mathbf{R}_0) d\mathbf{R} d\mathbf{R}' \times \iint \left\langle e^{-i\mathbf{k}\cdot(\mathbf{u}-\mathbf{u}')} \right\rangle e^{-ik_z(z-z')} dz dz' d\mathbf{K}_f. \quad (32)$$

Here the angled brackets denote the time-averaged phase factors due to the thermal displacements. To evaluate the effect of these thermal displacements in breaking the coherence in the z integration we need an explicit model for a vibrating crystal. The convenient Einstein model assumes independently vibrating atoms, and will automatically break the z-coherence. To see physically how the coherence is broken it is necessary to go to a phonon model, in which neighboring atoms are vibrating in phase (Jesson and Pennycook, 1995). Such a model has been developed by Warren (1990), using a Debye dispersion relation and assuming equipartition. It reduces the

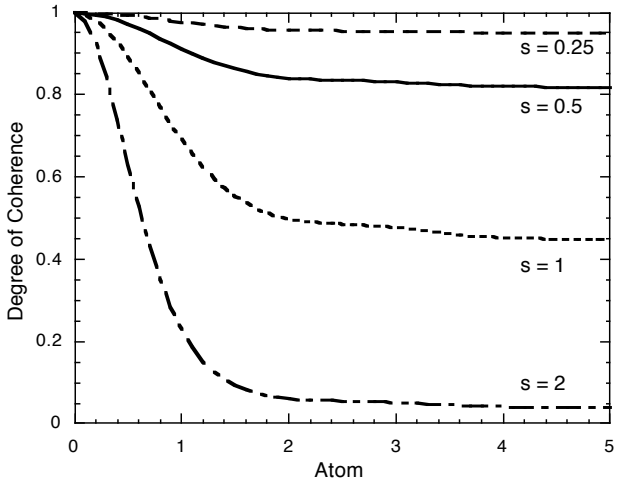

Figure 16. Degree of coherence between an atom at the origin and neighboring atoms along a column using a phonon model of thermal vibrations.

time average to a convenient analytical form, given by

$$W = \left\langle e^{-i\mathbf{k}\cdot(\mathbf{u}-\mathbf{u}')} \right\rangle = \exp\left\{ 2M \left[\frac{Si(q_B|\mathbf{u}-\mathbf{u}'|)}{q_B|\mathbf{u}-\mathbf{u}'|} - 1 \right] \right\}. \quad (33)$$

Here Si(x) is the sine integral function, $M = Bs^2$ is the usual Debye Waller factor with $s = (\sin\theta_B/\lambda)$, where $\theta_B = \theta/2$ is the Bragg angle, and q_B is the Brillouin zone boundary in the z-direction. Figure 16 shows plots of this function for different values of s showing how the coherence rapidly reduces as the separation of atoms along the column increases. For large separations the degree of coherence approaches the limiting value of e^{-2M}, which is the Einstein value for the strength of coherent reflections in his model of independently vibrating atoms. The phonon model shows clearly that atoms close together scatter with greater coherence than those far apart, leading to the concept of a longitudinal coherence volume.

Therefore, the need for large detector angles to ensure *inter*column incoherence (transverse incoherence) will automatically break the *intra*column coherence leading to longitudinal incoherence also, which is extremely convenient. To see which phonon modes are involved we can expand the time average as

$$\langle e^{-i\mathbf{k}\cdot(\mathbf{u}-\mathbf{u}')} \rangle = \langle e^{-i[\mathbf{K}_f\cdot(\mathbf{U}-\mathbf{U}') + k_z(u_z-u_z')]} \rangle. \quad (34)$$

For a high angle detector, the transverse momentum K_f is much greater than k_z, and so amplifies the effect of the transverse displacements U. Since it is the z integration we are trying to break, it is phonon wavevectors in the z direction that are needed (to

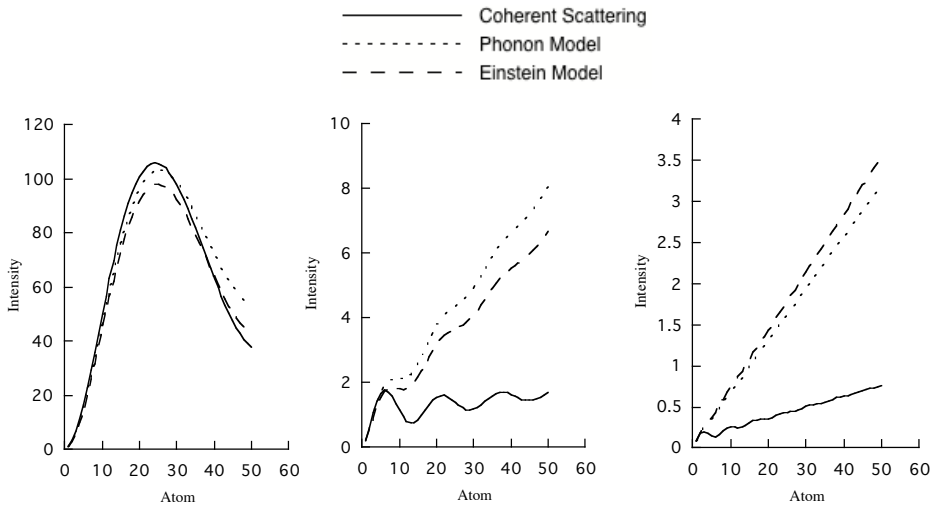

Figure 17. Thickness dependence of the scattering from a column of Rh atoms spaced 2.7 Å apart along the beam direction for low, medium and high scattering angles. The coherent thickness oscillations at low angles are largely suppressed at high angles.

compensate for the curvature of the Ewald sphere). So transverse modes of phonons travelling in the z direction are required to break the z-coherence.

The change in the thickness dependence of the image intensity with increasing annular detector angle is largely governed by the changing longitudinal coherence length. The thickness behavior changes from oscillatory at low angles reflecting the long coherence length to more linear at large angles where the coherence length is much shorter than the specimen thickness. This is illustrated in Fig. 17 for a column of Rh atoms 2.7 Å apart illuminated by 300 kV electrons. With low detection angles the scattering is almost entirely coherent. With increasing detector angle the scattered intensity exhibits an initial coherent dependence with thickness, changing to an incoherent dependence as the column becomes significantly longer than the coherence length. At higher angles the initial coherent oscillation occurs more quickly, and the thickness dependence is mostly linear. Also, a significant fraction of the thermal displacements of the atoms are due to zero point fluctuations which will not disappear on cooling the sample. Thus it should not be assumed that the scattering will become coherent on cooling, but the required detection angles will be increased somewhat.

Although this kinematic scattering model is inappropriate for thick crystals with strong dynamical diffraction, the physical insights remain valid; it is the phonons that break the coherence in the z direction and remove the strong oscillatory thickness behavior. This gives us an image that is effectively integrated through the specimen thickness rather than an image based on the exit face wave function.

3.6. INCOHERENT IMAGING OF THICK CRYSTALS: DYNAMICAL EFFECTS

3.6.1. Bloch States

Bloch states are the quantum mechanical stationary states of a fast electron in a crystal, and therefore the natural basis for examining the effects of dynamical diffraction. If it were possible to excite a single Bloch state at the entrance surface of a crystal, it would propagate to the exit surface unchanged, except for a depletion in amplitude due to processes such as phonon excitation and inelastic scattering. Such processes are usually modeled by a phenomenological absorption coefficient. However, Bloch states are not the stationary states of the fast electron in the vacuum, plane waves are, and so it is necessary to couple the two sets of states at the crystal entrance surface. For example, a single incident plane wave can be expanded into a complete set of two-dimensional Bloch states $b^j(\mathbf{K},\mathbf{R})$ as

$$\psi(\mathbf{R},z) = \sum_j \varepsilon^j(\mathbf{K}) b^j(\mathbf{K},\mathbf{R}) e^{i\mathbf{K}\cdot\mathbf{R}} e^{-ik_z^j(\mathbf{K})z} e^{-\mu^j(\mathbf{K})z} \qquad (35)$$

with excitation coefficients $\varepsilon^j(\mathbf{K})$ and absorption coefficient $\mu^j(\mathbf{K})$ propagating along the z axis with wavevector $k_z^j(\mathbf{K})$. The first six states for Si(110) are shown in Fig. 18, and take on the form of molecular orbitals about the atomic strings. Usually, the wave function inside the crystal can be well-represented with just a few strongly

excited Bloch states, and it is their propagation with different wavevectors k_z^j that leads to the depth dependent dynamical diffraction effects. The 1s states are located over the deepest part of the projected potential and consequently have the highest kinetic energy and the largest k_z^j. As they overlap little with neighboring columns, their $k_z^j(\mathbf{K})$ is independent of **K**. Such states are said to be non-dispersive. These states that are the most localized in real space are the broadest in reciprocal space. States that are less localized in real space will overlap and perhaps hybridize with states on neighboring columns. It is these states that are responsible for the non-local effects in phase contrast imaging. In reciprocal space such states are narrow. Now it begins to be clear how a small axial detector will detect the interference between all highly excited Bloch states whereas the high angle detector will be sensitive only to the 1s states.

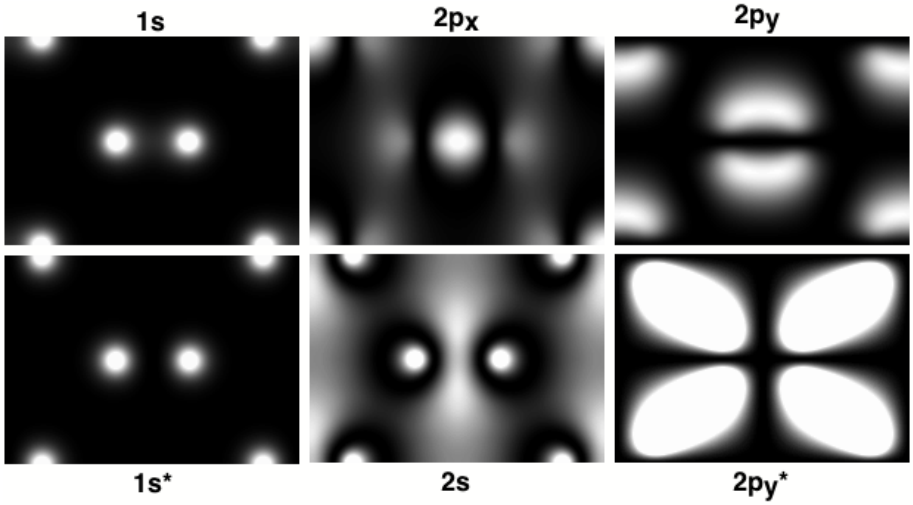

Figure 18. Intensities of the first six Bloch states in Si(110) with their molecular orbital assignments. The 1s states are located around the Si atomic columns.

3.6.2. The Dynamical Object Function

Dynamical diffraction can be included in the expressions for detected intensity given in the previous sections simply by replacing the aperture function $H(\mathbf{K})$ by $H(\mathbf{K})\psi(\mathbf{R},z)$. For example, Eqn. 6 for the probe becomes

$$P(\mathbf{R}-\mathbf{R}_0,z) = \int A(\mathbf{K}) \sum_j \varepsilon^j(\mathbf{K}) b^j(\mathbf{K},\mathbf{R}) e^{i\mathbf{K}\cdot(\mathbf{R}-\mathbf{R}_0)} e^{-ik_z^j(\mathbf{K})z} e^{-\mu^j(\mathbf{K})z} d\mathbf{K}. \qquad (36)$$

And the probe intensity inside the crystal is now given by $P^2(\mathbf{R}-\mathbf{R}_0,z)$. This is now the total wave function inside the crystal (incident wave plus scattered waves). As shown by Nellist and Pennycook (1998a) it can be integrated over the detector and Fourier transformed with respect to $\mathbf{R}_o$ to give an expression analagous to Eqn. 16 for the image spatial frequency ρ

$$I(\boldsymbol{\rho},z) = \int D(\mathbf{K}_f)d\mathbf{K}_f \int A(\mathbf{K})A^*(\mathbf{K}+\boldsymbol{\rho})\sum_{j,k} \varepsilon^j(\mathbf{K})\varepsilon^{k*}(\mathbf{K}+\boldsymbol{\rho})b_{\mathbf{K}_f}{}^j(\mathbf{K})b_{\mathbf{K}_f}{}^{k*}(\mathbf{K})$$

$$\times e^{-i[k_z^j(\mathbf{K})-k_z^k(\mathbf{K})]z}d\mathbf{K}, \qquad (37)$$

where $b_{\mathbf{K}_f}{}^j(\mathbf{K})$ represents the $\mathbf{K}_f$ Fourier component of the Bloch state j, and we have ignored absorption for the time being. The enormous advantage of this fully reciprocal space representation is that the integral over the detector can be performed immediately to see which Bloch states give important contributions to the image intensity. This results in vast savings in computer time compared to multi-slice approaches where all the beams reaching the detector are calculated even if they eventually sum to zero. The detector sum is given by

$$C_{jk}(\mathbf{K}) = \int D(\mathbf{K}_f)\, b_{\mathbf{K}_f}{}^j(\mathbf{K})b_{\mathbf{K}_f}{}^{k*}(\mathbf{K})\, d\mathbf{K}_f \qquad (38)$$

and acts on the Bloch states as a high pass filter in the same way as it selected the high frequency components of the atomic potential in section 3.4.2.

3.6.3. The High Thickness Limit

The efficiency of the Bloch wave filtering is seen very clearly in the limit of high thickness where the cross terms C_{jk} become insignificant compared to the independent terms C_{jj} because of the exponential factor in Eqn. 37. Figure 19 shows the contribution to the object function of the two 1s states in GaAs compared to the object function calculated with all 265 states. These other states add just a small, almost uniform background intensity to the object function. A similar conclusion had been reached by Pennycook and Jesson (1990, 1991, 1992) assuming the image to be given by the intensity at the atom sites. Although the intensity at the sites is dominated by the 1s states, this approach is equivalent to setting $D(\mathbf{K}_f) = 1$ above. Using the full detector function results in even more perfect 1s state filtering.

The image intensity, Eqn. 37, is now given by

$$I(\boldsymbol{\rho},z) = \int A(\mathbf{K})A^*(\mathbf{K}+\boldsymbol{\rho})\sum_{j,k} C_{jk}(\mathbf{K})\varepsilon^j(\mathbf{K})\varepsilon^{k*}(\mathbf{K}+\boldsymbol{\rho})e^{-i[k_z^j(\mathbf{K})-k_z^k(\mathbf{K})]z}d\mathbf{K}, \qquad (39)$$

which, since the dominant contribution at high thickness is the 1s states, we can approximate as

$$I(\boldsymbol{\rho}) = \int A(\mathbf{K})A^*(\mathbf{K}+\boldsymbol{\rho})\sum_{1s} C_{jj}\varepsilon^j(\mathbf{K})\varepsilon^{j*}(\mathbf{K}+\boldsymbol{\rho})d\mathbf{K}. \qquad (40)$$

Here we have removed the $\mathbf{K}$ dependence of the C_{jj} term, as the s states are non-dispersive. Also, since the excitation of a Bloch state is its Fourier transform, for the

highly localized s states the excitation is also quite a slowly varying function over the objective aperture range, and can therefore be replaced by an average excitation

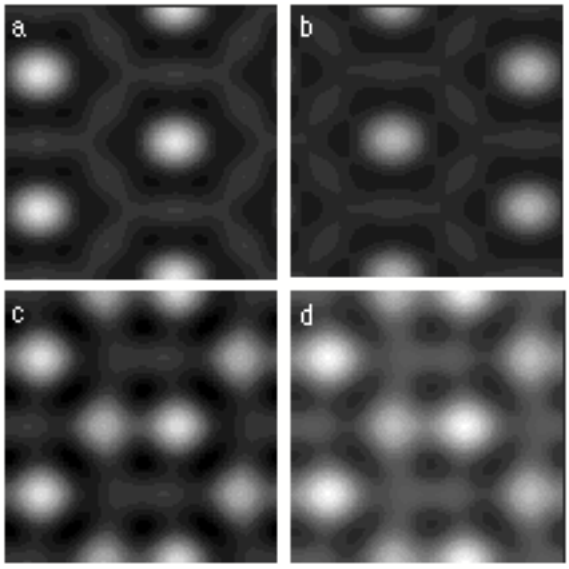

Figure 19. Contributions to the image intensity in GaAs (110) from (a) the 1s Bloch state on the As column, (b) the Ga 1s state, (c) both Ga and As 1s states, (d) all 265 states, showing the effectiveness of the high angle detector as a Bloch state filter.

$$\varepsilon^{av} = \int \varepsilon^{1s}(\mathbf{K})\, d\mathbf{K} \tag{41}$$

Now the only K dependence is in the aperture function, and we again have incoherent imaging. Transforming back to real space, the Fourier transform of the aperture functions gives the probe intensity profile. With the excitation slowly varying, we can approximate the dynamical object function in the high thickness limit as a set of weighted δ-functions. The image is then given by

$$I(\mathbf{R}_0) = C_{jj}\, \varepsilon^{av^2} \delta(\mathbf{R}_j) * P^2(\mathbf{R}_0), \tag{42}$$

where $\mathbf{R}_j$ is the coordinate of column j. For constant atom separation along a column, the high angle components of the 1s states (C_{jj}) scale as Z^2, as would be expected on the basis of Rutherford scattering. The excitation on axis, $\varepsilon^{1s^2}(\mathbf{0})$, goes as 1/Z, but as states from lighter columns are broader, their excitation falls off faster with K. The two factors largely cancel, and so, in the high thickness limit, we find the dynamical object function is only slightly less than Z^2. In this limit we therefore see practically the same contrast in the presence of dynamical diffraction as we would for single atoms.

3.6.4. Channeling Approximation

If we are not at the high thickness limit described above, we expect dynamical oscillations with depth z due to the exponential term in Eqn. 29. Although the 1s states have by far the largest single C_{jj} value, not all of the incident beam can be coupled into the s states. Depth dependent oscillations come from the beating of the 1s states with the sum of all the other states. As noted before, the 1s states are located over the deepest part of the projected potential and so have the highest kinetic energy and the largest k_z^j. All the other less localised states have very similar k_z^j values, and so in thin crystals they all propagate approximately in phase through the thickness z. Therefore it is a good approximation to consider just two components to the electron wave function, the 1s state propagating with wavevector k_z^{1s}, and a term $1- b^{1s}(\mathbf{R})$ propagating at an average k_z^0. The beating between these two components occurs with an extinction distance $\xi = 2\pi/(k_z^{1s} - k_z^0)$. Replacing the c_{jj} term by Z^2, the depth dependent object function for a column is now given by

$$O_C(\mathbf{R},z) = Z^2 \varepsilon^{av^2} \left[1 - \cos(2\pi z/\xi)\right] e^{-2\mu^{1s}(\mathbf{0})z} \tag{43}$$

where we have included absorption, as the 1s states are the most highly absorbed Bloch states. Figure 20 shows a plot of this function for Si and Ge in the ⟨110⟩ orientation, compared to the 1s state intensities alone.

This expression assumes all scattering is coherent. Experimental images do not show these strong depth oscillations because most of the scattering reaching the detector is thermal diffuse scattering. We are detecting the absorption out of the coherent wavefield described by Eqn. 43. The phonons are again breaking the z-coherence and allowing us to integrate the generation of high angle diffuse scattering

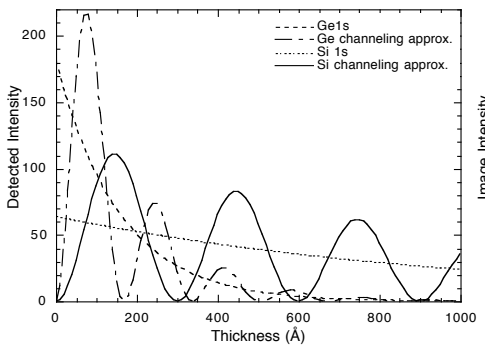

Figure 20. Intensity of coherent scattering reaching the annular detector from Si and Ge ⟨110⟩ using the channeling approximation, Eqn. 42. Parameters are: Si, ξ=300Å, μ^{1s} = 0.00048, Ge, ξ = 169 Å, μ^{1s} = 0.00032.

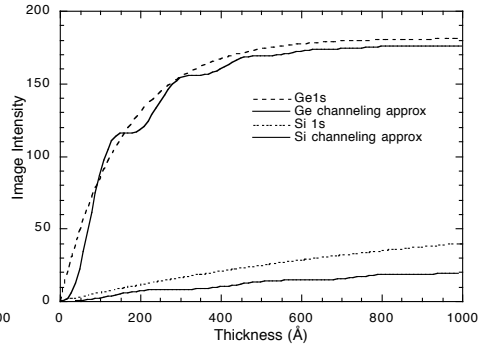

Figure 21. Intensity of incoherent scattering reaching the annular detector from Si and Ge ⟨110⟩ using the channeling approximation, Eqn. 43. Parameters are the same as Fig. 20.

over the thickness t of the crystal. Integrating Eqn. 43 gives

$$O_{TDS}(\mathbf{R},t) = \frac{Z^2\varepsilon^{av^2}}{\left(\xi^2\mu^{1s^2}+\pi^2\right)}\left[\pi^2\left(1-e^{-2\mu^{1s}t}\right)-\xi^2\mu^{1s^2}e^{-2\mu^{1s}t}\left(1-\cos\frac{2\pi t}{\xi}\right)-\pi\xi\mu^{1s}e^{-2\mu^{1s}t}\sin\frac{2\pi t}{\xi}\right] \tag{44}$$

which is plotted in Fig. 21. The thickness integration has removed the strong dynamical oscillations and the form of the curve is in good agreement with both experimental observations and with multislice simulations for a high detection angle (Loane, Xu and Silcox, 1992, Anderson, et. al., 1997, Hartel, Rose and Dinges 1996). Under these conditions, the channeling approximation is very useful for simulating images, and saves enormously on computer time. For lower detection angles the other states become more important, and a fraction of the coherent exit face wave function $O_C(\mathbf{R},t)$ must be added. The image contrast begins to show more oscillatory dependence on thickness and the incoherent characteristics are progressively lost. Also, single heavy impurity atoms will sample the oscillating wavefield of Fig. 20, and so show depth dependent contrast (Loane, Kirkland and Silcox, 1988, Nakamura et al., 1997).

Another important situation where other states become significant is if the atomic columns are no longer straight, but bent due to the presence of a defect or impurity. Then transitions occur between Bloch states which is the origin of diffraction contrast imaging in the CTEM, and diffraction contrast effects will also be seen in the annular detector signal, as discussed below.

3.7. STRAIN CONTRAST

Elastic strain fields due to impurities, point defects or extended defects such as dislocations and stacking faults will induce transitions into or out of the 1s Bloch states. This is the usual mechanism of diffraction contrast in CTEM images, and clearly, will also be a source of contrast in the annular detector image. In general, transitions will again depend on depth in the crystal, giving oscillatory contrast from

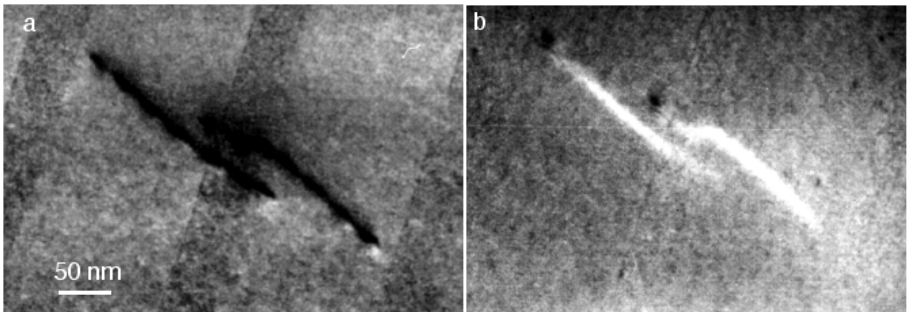

Figure 22. (a) Bright field and (b) annular dark field STEM images of inclined dislocations in a thick Si/Si(B) superlattice

inclined dislocations, as seen in Fig. 22 (Perovic, Rossouw and Howie, 1993, Perovic, Howie and Rossouw, 1993). For zone axis imaging, the periodicity will be the extinction length ξ as seen in Fig. 20. Even dislocations that are viewed end-on may show strain contrast due to the transverse relaxations of the atomic positions that occur near the sample surface. For this reason, grain boundaries, which are closely spaced arrays of dislocations, often appear brighter or darker than the matrix. Strain contrast is relatively long range compared to the lattice parameter and can be removed by Fourier filtering if desired. Strain contrast also depends strongly on detector angle, and can therefore be distinguished from compositional changes (*Z*-contrast) by comparing images taken with different inner detector angles, as shown below.

From our discussion of z-coherence, and its destruction by phonons, it is clear that static transverse displacements comparable to the atomic vibration amplitude will also significantly affect the scattering at high angles. In the Einstein model of thermal vibrations, atomic scattering cross sections can be defined for coherent and incoherent scattering. The coherent scattering is reduced by the Debye-Waller factor $M = 8\pi^2\overline{u_T^2}$, where $\overline{u_T^2}$ is the mean square thermal vibration amplitude of the atom (Hall and Hirsch, 1965), and is given by

$$\sigma_c = f^2 e^{-2Ms^2}, \tag{45}$$

where f is the atomic form factor and $s = (\sin\theta_B/\lambda)$ is the scattering angle. The incoherent cross section is then

$$\sigma_{TDS} = f^2 \left(1 - e^{-2Ms^2}\right). \tag{46}$$

In the presence of static random atomic displacements, assuming a Gaussian distribution of strain with a mean square static displacement of $\overline{u_S^2}$, the atomic scattering cross section will be modified to (Hall, Hirsch and Booker, 1966)

$$\sigma_S = f^2 \left(1 - e^{-2(M+M_S)s^2}\right). \tag{47}$$

where $M_s = 8\pi^2\overline{u_S^2}$. It is clear from the form of these expressions that at a sufficiently high scattering angle σ_c tends to zero and both σ_{TDS} and σ_S tend to the full atomic scattering cross section f^2. In other words no additional scattering due to strain will occur. At lower angles where the Debye-Waller factor is significant, static strains comparable to the thermal vibration amplitude may lead to a significantly enhanced scattering cross section.

Figure 23 shows images of a thermally grown Si/SiO_2 interface. The bright field phase contrast image, Fig. 23a, shows dark contrast that could be due to a number of effects, such as strain, thickness variation, bending of the crystal or a combination of these mechanisms. Figure 23b shows an incoherent dark field image collected simultaneously using a low (25 mrad) inner radius for the annular detector. Now there

is a bright band near the interface indicating additional scattering. With this image alone, this additional scattering could be due either to strain or to the presence of some heavy impurity atoms. However, when the inner detector angle is increased further to 45 mrad the bright line disappears (Fig. 23c), showing that the contrast cannot be due to the presence of heavy impurity atoms which would still give increased scattering. The contrast must therefore be due to a static strain effect. Intensity profiles across the two dark field images are shown in Fig. 24. Taking their ratio normalizes any change in the s states due to the strain and allows the additional static displacement to be calculated, as shown in Fig. 25 (Duscher et al., 1998). This is seen to decrease exponentially from the incoherent interface as would be expected for a uniform array of misfit dislocations.

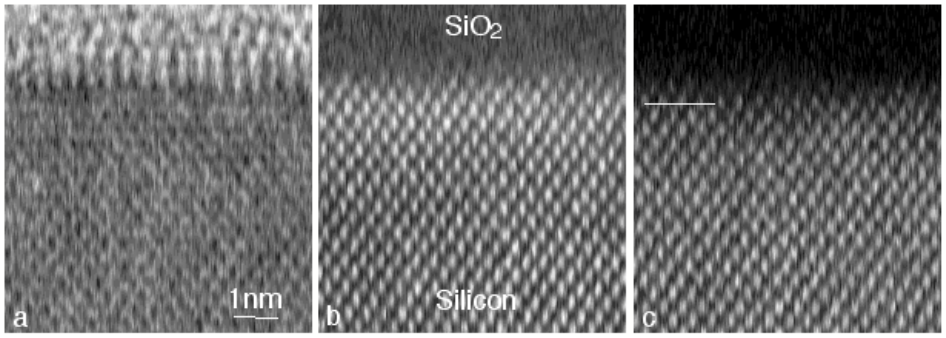

Figure 23. STEM images of a Si/SiO_2 interface; (a) bright field phase contrast image, (b) *Z*-contrast image with 25 mrad inner detector angle showing strain contrast, (c) *Z*-contrast image with 45 mrad inner detector angle. The horizontal line marks the last Si plane used for strain profiling.

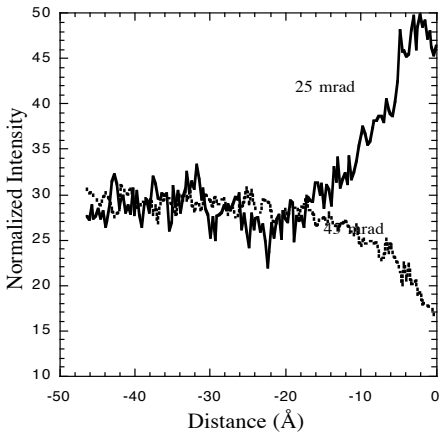

Figure 24. Intensity profiles across the *Z*-contrast images of Fig. 23. The high angle profile shows the dechanneling effect near the interface which can be used to normalise the profile obtained with lower detector angle.

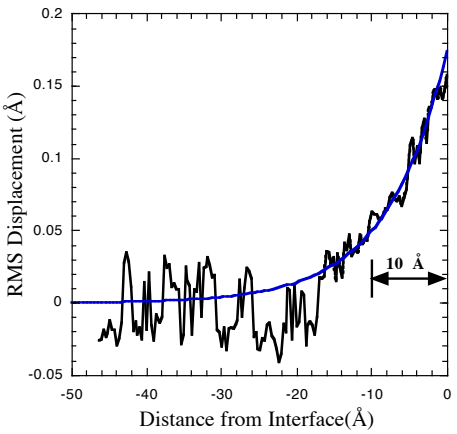

Figure 25. RMS atomic displacement due to static strain induced by the Si/SiO_2 interface and an exponential decay fitted to the data.

4. Retrieval of the Object Function

The key advantage of an incoherent image is avoidance of the phase problem associated with coherent imaging. As there are no phases in an incoherent image, no phase information can be lost, and we have the possibility of direct inversion from the image back to the object. This of course is the reason that an incoherent image can be interpreted intuitively. All that is lost in recording the incoherent image is information on the high spatial frequencies in the object, lost because of the convolution with a probe of finite width. It might naturally be assumed that to retrieve the object, all that would be required would be to deconvolute the probe from the image. But unfortunately this does not work, because it does not retrieve the lost high frequency information. All deconvolution can do is to correct for the decaying transfer function $t(\boldsymbol{\rho})$ by dividing in Fourier space,

$$O(\boldsymbol{\rho}) = \frac{I(\boldsymbol{\rho})}{t(\boldsymbol{\rho})}. \tag{48}$$

This fails as the transfer function approaches zero, and in the Wiener filter method the object frequency is decayed to zero at the cutoff of the transfer function by the addition of a small constant ε

$$O(\boldsymbol{\rho}) = \frac{t^*(\boldsymbol{\rho})I(\boldsymbol{\rho})}{\left|t(\boldsymbol{\rho})\right|^2 + \varepsilon}. \tag{49}$$

The results of such a Wiener filter deconvolution are shown in Fig. 26. A raw image of Si⟨110⟩ is shown in (a) (Nellist and Pennycook, 1998b). The result of a simple Fourier filter to remove the noise is seen in (b), and the effect of the Wiener filter is shown in (c). Unfortunately, the effect of the filter is to produce artifacts between the columns. These arise because of the abrupt cutoff in transfer imposed near the maximum spatial frequency of the data. To avoid such artifacts, a slower cutoff can be imposed on the image but obviously this would also degrade the resolution. In fact, the natural incoherent transfer function is already rather well optimized, and it is not useful to attempt to improve upon it in this way.

An alternative method is required to reconstruct the missing high frequency information. For atoms wide apart, it is reasonable to locate the maximum of each image feature, which is using our *a priori* information that the sample comprises discrete atoms. But this procedure does not work near the limit of resolution because it does not take any account of the probe profile. In Si⟨110⟩, pairs of columns are spaced by distances comparable to the probe size, and the peak image intensity is displaced outwards by a few tenths of an angstrom depending on defocus. A method to accurately account for the effects of the convolution is maximum entropy (Gull and Skilling, 1984). This assumes no prior knowledge concerning the nature of the image, except that it is incoherent.

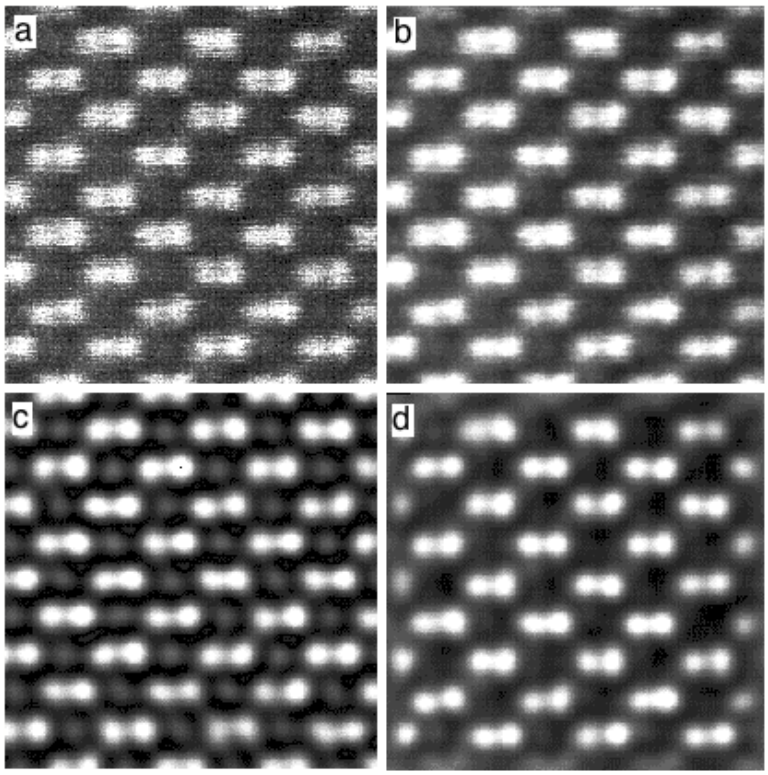

Figure 26. (a) Raw *Z*-contrast image of Si ⟨110⟩; (b) low pass filtered image to reduce noise; (c) deconvolution of the probe function leads to artifacts between the columns; (d) maximum entropy retrieves the correct object giving a reconstructed image free of artifacts.

The maximum entropy method is based on Bayes theorem, which states that

$$\mathrm{p}(a \mid b)\ \mathrm{p}(b) = \mathrm{p}(b \mid a)\ \mathrm{p}(a) \tag{50}$$

that is p($a \mid b$), the probability of a given b, multiplied by the probability of b is equal to the probability of b given a, multiplied by the probability of a. For our case we know the image data and suppose we also know the probe function, or we pick a trial probe function. We then need to be able to assess the probability of different trial object functions. So we write

$$\mathrm{p}(\textit{object function} \mid \textit{image data})\ \mathrm{p}(\textit{image data}) = \mathrm{p}(\textit{image data} \mid \textit{object function})\ \mathrm{p}(\textit{object function}). \tag{51}$$

Now the probability of the image data is a constant (we are tying to assess the most likely object from a given image), and so we have

$$p(\textit{object function} \mid \textit{image data}) = k\, p(\textit{image data} \mid \textit{object function})\, p(\textit{object function}) \tag{52}$$

where k is a constant. As we know the probe profile we can convolute it with our trial object function and compute the probability of the image data given our simulated image using a χ^2 fit for example. All we need is an expression for p(*object function*), a means of assessing the probability of different object functions. Because any distribution of high spatial frequencies beyond our image cutoff will give equally good fits to the data, there are an infinite number of possible object functions that will give the same simulated image, and therefore the same value of p(*image data* | *object function*). It is the function of the prior distribution function, p(*object function*), to assess the one which is the most likely. Maximum entropy weights different object functions according to

$$p(\textit{object function}) = e^{\alpha S} \tag{53}$$

where α is a constant and S is the entropy, given by

$$S = -\sum_i n_i \log n_i \tag{54}$$

where n_i is the value of pixel i in the object function. The maximum entropy prior weights in the direction of least structure (maximum disorder). It will not give three columns where two would do, and is therefore a convenient prior for atomic resolution imaging. Note that nowhere do we assume that the object is comprised of discrete atoms.

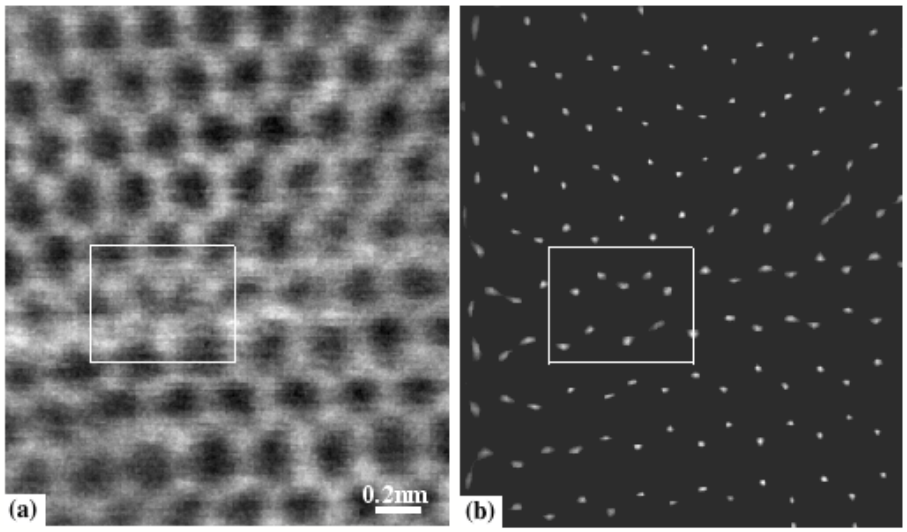

Figure 27. (a) *Z*-contrast image of an end-on threading dislocation in a GaN thin film grown on sapphire, viewed along the ⟨0001⟩ direction. The eight-fold structure of the core is clear from the maximum entropy reconstruction (b).

Figure 26d shows the reconstructed Si⟨110⟩ image, obtained by convoluting the maximum entropy object with the probe profile, where the artifacts between the columns are no longer present. Figure 27 shows a raw image of an end on dislocation core in GaN, with the atomic structure of the core determined by maximum entropy (Xin et al., 1998). In a good reconstruction, the maximum entropy object function consists of fine points, which can be taken directly as the column coordinates. The accuracy of the method is easily checked by measuring spacings far from the dislocation core, and here, as found typically, is approximately ± 0.1 Å for individual columns. More detailed descriptions of the maximum entropy approach applied to *Z*-contrast images, its accuracy and comparison with alternative schemes, are given in McGibbon, Pennycook and Jesson, 1998, Nellist and Pennycook, 1998b.

5. Atomic Resolution Spectroscopy

The *Z*-contrast image uses only the high angle scattering, this leaves the lower angles available for electron energy loss spectroscopy (EELS). Indeed, it is the availability of the incoherent *Z*-contrast image that allows the probe to be located over a selected individual atomic column or plane with very high precision. The probe can be centered over a chosen column just by maximising the annular detector intensity. Provided that incoherent imaging conditions also apply to the EELS signal, the inelastic scattering from that column will also be maximised at the same probe position, giving an atom column resolved analysis (Browning, Chisholm and Pennycook, 1993, Batson, 1993, Browning and Pennycook, 1995).

With the phase contrast image of a CTEM it is not so simple to accurately illuminate an individual column. Neither is it practical to form an energy filtered image at atomic resolution except from low loss electrons. In CTEM the objective lens is behind the specimen (Fig. 3a), so that the energy-loss electrons will suffer chromatic aberration on passing through the lens to form an image. For the energy filtered image to show atomic resolution, the energy window selected by the filter must be kept below 1 eV, which for core edges will lead to a very noisy image. The information is more efficiently gathered in STEM mode by illuminating the chosen column, and collecting the transmission EELS spectrum. Chromatic aberration in the objective lens no longer degrades the spatial resolution. The simultaneous use of atomic resolution imaging and analysis in the STEM has proved to be a powerful means for determining atomic structure, atomic sites of impurity atoms and their valence, and the local band structure seen by the selected column. In principle, atomic resolution is also possible using X-ray fluorescence, but to date it has been demonstrated only for EELS, because of the much lower detection efficiency for X-rays.

Much of the discussion on incoherent imaging with elastically scattered electrons can be carried over to the case of inelastically scattered electrons (Pennycook et al. 1997). For example, it has been realized for some time that collecting all the inelastically scattered electrons will give an incoherent image (Rose, 1976, Ritchie and Howie, 1988). The inelastic object function is given by

$$O'(\mathbf{R}) = \left(\frac{e^2}{\pi\hbar v}\right)^2 \left| \int \frac{\rho_{no}}{q^2}(\mathbf{q}) e^{-i\mathbf{K}\cdot\mathbf{R}} d\mathbf{K} \right|^2 \tag{55}$$

where the momentum transfer is now three-dimensional, $\mathbf{q} = (\mathbf{K},q_z)$, because of the minimum momentum transfer q_z at zero scattering angle. ρ_{no} is the transition matrix element from an initial state $|0\rangle$ to a final state $|n\rangle$ and v is the electron velocity. These are often evaluated in the hydrogenic approximation (Maslen and Rossouw, 1984, Rossouw and Maslen, 1984, Allen, 1993, Holbrook and Bird, 1995, Pennycook et al.,1997, Rafferty and Pennycook, 1998).

Inelastic object functions are significantly broader than elastic object functions, due not so much to the size of the inner shell itself, but to the fact that less transverse momentum is available from the light electron compared to the heavy nucleus. The full-width-half-maxima for several elements are shown in Fig. 28, and are below 1 Å even for the light elements, suggesting atomic resolution will still be possible with a suitably small probe. For considerations of contrast, one can think of the image as given by an effective probe convoluted with delta functions, as in the elastic case. The effective probe is now the real probe convoluted with the inelastic object function, $P^2_{eff}(\mathbf{R}) = P^2(\mathbf{R})*O'(\mathbf{R})$ which is equivalent to a transfer function $t'(\boldsymbol{\rho}) = t(\boldsymbol{\rho})O'(\boldsymbol{\rho})$.

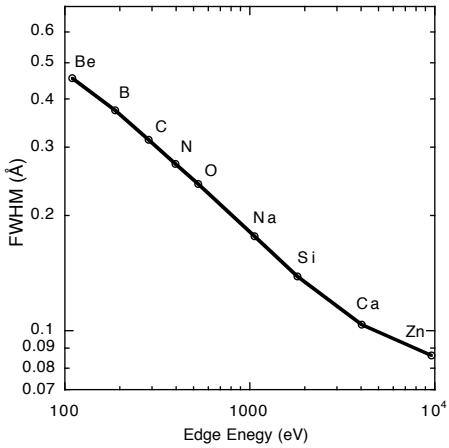

Figure 28. Full-width-half-maxima of EELS object functions for K-shell excitations by 300 kV incident electrons, calculated in the hydrogenic model.

To show that spectroscopy can be performed on individual atomic columns selected from the *Z*-contrast image, Fig. 29 shows an example of a Mn-doped $SrTiO_3$ grain boundary. Significant changes in both concentration and band structure are observed from column to column (Duscher, Browning and Pennycook, 1998). The Mn is seen to prefer the Ti sites in the grain boundary, and from the Ti $L_{II/III}$ ratio, appears to change its valence state from 4+ in the bulk to 3+ at the boundary. Data of this nature is particularly valuable for linking the structure of the grain boundary to its electrical activity. However, the images shown here were taken on the 300 kV STEM whereas the EELS spectra were taken on the 100 kV STEM, with only a 2.2 Å probe size. This is apparent in the fact that even with the probe located over the Sr columns, which do not themselves contain O, a strong O edge is observed from the four O columns only 1.95 Å away.

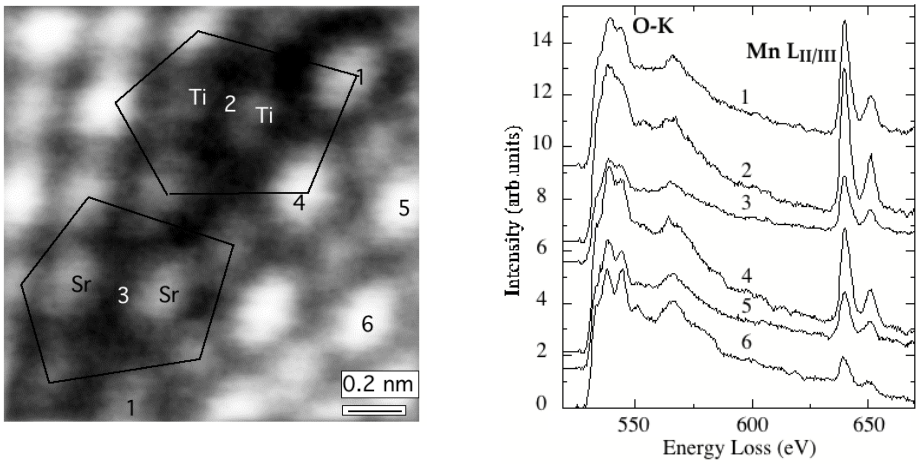

Figure 29. *Z*-contrast image of a Mn-doped $SrTiO_3$ grain boundary with EELS spectra obtained from selected atomic columns.

A more detailed discussion of atomic resolution EELS and comparison to the spatial difference method is given in Duscher, Browning and Pennycook, 1998. EELS is particularly useful for analyzing elements such as oxygen which are too light to be seen directly in the *Z*-contrast image (McGibbon et al., 1994, Dickey et al., 1998). It has been used for example to map hole concentrations in the high temperature superconductors at a spatial resolution below the superconducting coherence length. Significant differences were found at grain boundaries depending on boundary geometry (Browning et al., 1993, 1998).

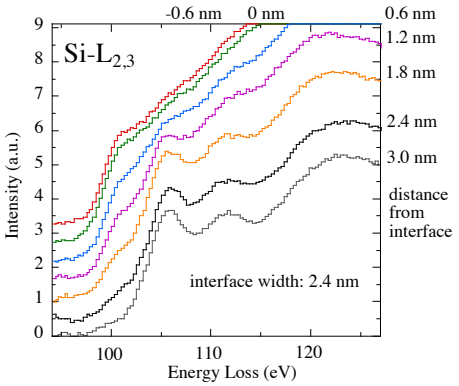

Figure 30. Si-$L_{2,3}$ spectra across a Si/SiO_2 interface showing evolution of the SiO_2 band gap. The full gap is not established until 2.4 nm into the oxide.

EELS is also useful at an amorphous crystal interface, where it can provide valuable information on the composition of the amorphous phase near the interface. Figure 30 shows a series of Si-$L_{2,3}$ ionization edges from a Si/SiO_2 interface produced by oxidation with a gaseous oxygen/nitrogen source (Duscher et al., 1998). Each spectrum is obtained with the probe located at a different distance from the interface, as determined from the *Z*-contrast image. Due to the large band gap of the SiO_2, the onset of the ionization edge is 104 eV in the SiO_2 compared to 99 eV in the Si. It is clear that the edge profile evolves from Si to SiO_2 over a surprisingly extended region, greater than 2 nm. This is almost an order of magnitude greater than the spatial resolution of the measurement. Similar data was obtained from the O-K edge, providing an independent verification of an extended sub-stoichiometric zone.

6. Applications in Materials Science

Theoretical modeling is a valuable complement to these techniques of atomic scale characterization. It can choose between different models that agree with the image, models that differ in their arrangement in the z-direction for example. In particular, first-principles total energy calculations can take the coordinates provided by the image and relax them in a fully self consistent manner, thereby testing if the proposed structure is stable, and avoiding the need for increasing the accuracy of the image inversion. The final total energy can be used to determine grain boundary energies, segregation energies and transformation energies for example. Following relaxation, the electron distribution within the structure is known, from which all properties can in principle be derived and compared to experiment. This combination of *Z*-contrast imaging, EELS and theory is highly synergistic; by providing atomic coordinates from experiment we avoid the need for time-consuming searches of trial model structures with the computer. Theory, in turn, often suggests new directions for experiment. Below are two recent examples of combined experimental and theoretical studies, and some additional recent experimental studies.

6.1. ARSENIC SEGREGATION SITES AT A SILICON GRAIN BOUNDARY

Z-contrast imaging enables low concentrations of high-*Z* impurities to be observed directly, as shown in Fig. 31, a *Z*-contrast image of a grain boundary in Si, after doping with As (Chisholm et al., 1998). The atomic structure of the boundary is directly determined from the positions of the bright features in the image, and is different from all structures proposed previously. It comprises a continuous sequence of dislocation cores, a perfect edge dislocation (1) and two perfect mixed dislocations (2,3) arranged as a dipole, followed by the same sequence (1',2',3') mirrored across the boundary plane. In the ⟨001⟩ projection, these dislocations appear as a connected array of pentagonal and triangular arrangements of atomic columns. One of the dislocation cores contains columns that are 20% brighter on average than other similar columns, indicating the presence of As atoms. The increased intensity corresponds to an average of only 5% As concentration, approximately two As atoms in each atomic column.

Ab-initio theoretical studies added significant further insight. It was found that isolated As atoms have only a small segregation energy of ~ 0.1 eV in all the dislocation cores, not showing preference for any site in the boundary. Arsenic dimers, on the other hand, were found to have a substantial segregation energy (Maiti et al., 1996). Most significantly, with a dilute concentration of As dimers, preference was found for those sites seen bright in the image, with the segregation energy being consistent with the As solubility limit in the bulk at the 700°C annealing temperature. A remarkably detailed and consistent atomic-scale picture of impurity segregation has been achieved for this grain boundary.

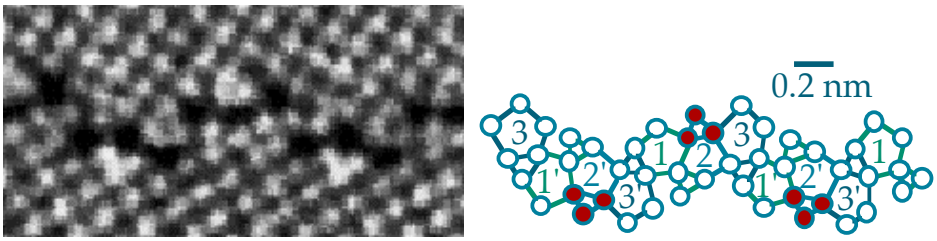

Figure 31. Z-contrast image of a 23° [001] tilt grain boundary in Si showing its unexpectedly complex structure. The five-fold rings (with black centers in the image) are dislocation cores arranged in a repeating sequence along the boundary. Columns shown black in the schematic are those seen brighter in the image due to segregated As.

6.2. IMPURITY-INDUCED GRAIN BOUNDARY TRANSFORMATION IN MgO

In Fig. 32, STEM imaging of an MgO grain boundary (Yan et al., 1998a) reveals a structure that is inconsistent with the widely accepted structure of the boundary proposed by Harris et al. (1996) based on theoretical modelling using classical potentials. The observed structure is similar to that proposed much earlier by Kingery (1974). These two structures are compared in Fig. 32, where the large empty core of the Harris structure is obviously very different from the more dense core of the Kingery model. On careful examination of the intensity in the experimental image, it can be seen that certain specific atomic columns at the grain boundary are significantly brighter than neighboring columns, as arrowed in the figure. This suggests that impurities, with $Z > 12$, may be segregated at these sites. EELS measurements indeed established that significant concentrations of Ca were present in the grain boundary, 0.3 monolayers, consistent with the bright intensity in the image.

To reconcile these observations with the prior experimental and theoretical work, first-principles theoretical calculations were performed. These calculations in fact reproduced the results of the classical potential calculations for the clean grain boundary, indicating the open structure to be 0.5 eV lower in energy per periodic repeat unit. Theory further determined that Ca has a large segregation energy in both boundary structures, but *significantly higher* in the dense structure, sufficient to make the dense structure the lower energy boundary. These calculations therefore established that the dense structure is in fact *stabilized* by the Ca segregation, an example of a segregation induced structural transformation. Examination of the electronic charge distribution revealed just a small perturbation to the oxygen ions next to the Ca atom, indicating the transformation is structural not electronic in origin, i.e. it is driven by the size difference between Ca and Mg ions.

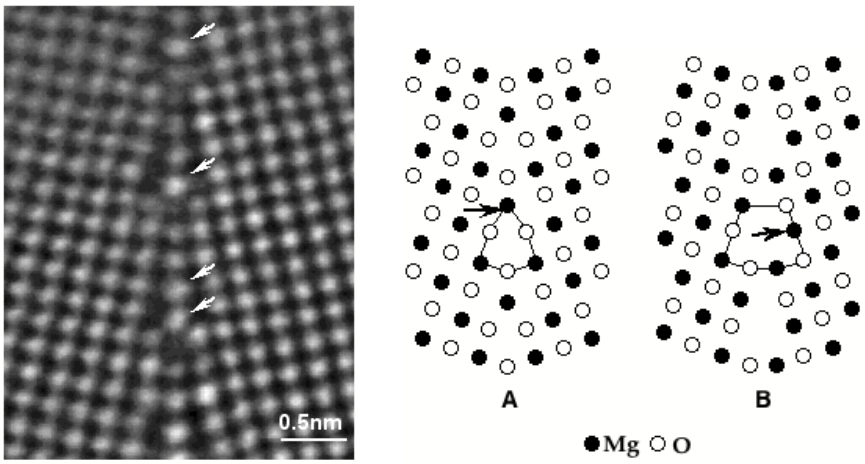

Figure 32. *Z*-contrast image from a 24° ⟨001⟩ tilt grain boundary in MgO showing occasional bright atomic columns at the grain boundary (arrowed), compared to two structures for the 36° ⟨001⟩ tilt grain boundary proposed by (a) Kingery (1974) and (b) Harris et al. (1996). Sites of Ca segregation are arrowed.

6.3. ORDERING IN FERROELECTRIC PEROVSKITES

Z-contrast imaging (Yan et al., 1998b) has resolved a controversy over the ordered structure of the lead-based relaxor ferroelectric $Pb(Mg_{1/3}Nb_{2/3})O_3$ (PMN). Two models have been proposed, the space-charge model and the charge-balanced random-layer model (Chen, Chan and Harmer, 1989) which differ in the distribution of the B-site cations in the doubled unit cell. In the space-charge model, the B^{I} and B^{II} sites are occupied exclusively by the Mg^{2+} and Nb^{5+} cations, respectively, in the form $Pb(Mg_{1/2}Nb_{1/2})O_3$. The resulting net negative charge is assumed compensated by a disordered, Nb^{5+} rich matrix. In the charge-balanced random-layer model, microscopic charge balance is achieved by occupying the B^{II} columns exclusively by Nb^{5+} and the B^{I} columns with a random distribution of Mg^{2+} and Nb^{5+} in a 2:1 ratio. *Z*-contrast imaging along the [110] zone axis can easily distinguish these two cases. In the space charge structure, the ratio of B^{I} column (Mg) to B^{II} column (Nb) intensities is given by approximately 1/17, whereas, in the charge-balanced random-layer structure it is close to 1/4.

Figure 33a shows a *Z*-contrast image of 25% La-doped PMN. The La doping increases the grain size significantly, ensuring that a single domain exists throughout the thickness of the region imaged. The intensity trace taken through the B sublattice clearly shows the intensity ratio is consistent with the 1/4 value expected for the charge-balanced random-layer model. For comparison, Fig 33b shows an image and line trace from $Ba(Mg_{1/3}Nb_{2/3})O_3$ in which the B sites are fully occupied by either Mg or Nb in a 2:1 ratio. The line trace shows the expected very weak intensity from the Mg column. A somewhat higher intensity is observed from the Mg site on the left hand side, indicating that the ordering is not entirely complete.

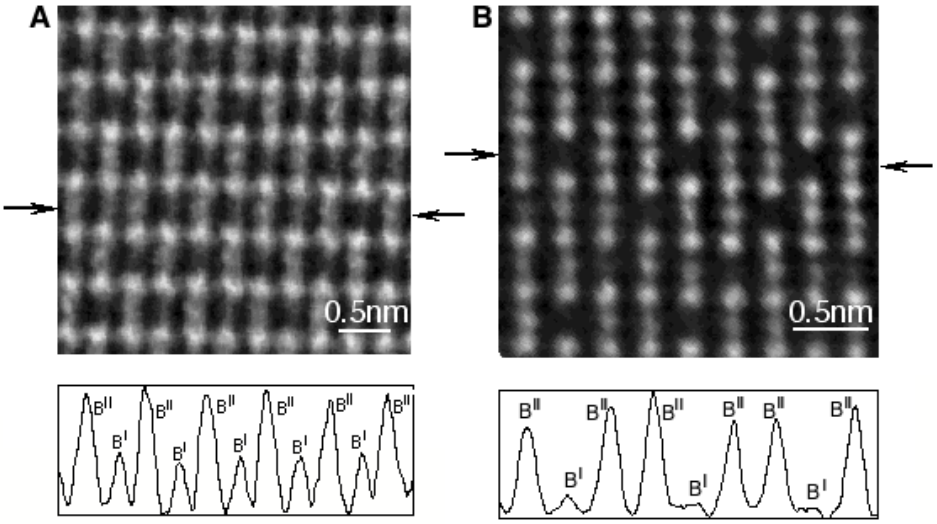

Figure 33. *Z*-contrast images of (a) 25% La-doped $Pb(Mg_{1/3}Nb_{2/3})O_3$, (b) $Ba(Mg_{1/3}Nb_{2/3})O_3$ with intensity profiles across the B sublattice showing the PMN to have the charge-balanced random layer structure.

Images were also taken from thin, undoped PMN. Although the contrast was often reduced due to the overlap of small domains through the sample thickness, the maximum value seen was still 1/4. This indicates that the ordered structure of undoped PMN also follows the charge-balanced random-layer model.

6.4. COMPLEX ATOMIC STRUCTURE OF In_2O_3-ZnO CONDUCTING FILMS

Z-contrast imaging of the transparent conducting oxide $In_2O_3(ZnO)_k$ has revealed a surprisingly complex atomic structure, containing inversion and mirror domain boundaries as an integral part of the structure (Yan et al., 1998c). In and Zn columns are directly distinguishable in a *Z*-contrast image taken along the $[11\bar{2}0]$ zone axis. Figure 34 shows an image of a layered structure observed in a film with nominal composition $In_2O_3(ZnO)_2$, after annealing at 900 °C for six hours in 20% argon 80% oxygen at 3.5 Torr. About 50 % of all grains were found to have this layered structure. It is immediately clear that each In intergrowth layer (seen bright) consists of only a single layer of In atoms, inconsistent with the model proposed by Cannard and Tilley (1988) in which two {111} planes of cubic In_2O_3 were suggested. Additionally, the distance between two adjacent In-O layers varies between 4, 5 or 6 layers of (0002) ZnO. This is evidence of a polytypoid structure, with an average composition obtained from the image of $In_2O_3(ZnO)_{10}$.

Furthermore, the spacing between the In plane and its nearest Zn planes, denoted d1, is measured to be approximately 0.31 nm, which is significantly larger than the Zn (0002) interplanar spacing of 0.26 nm, denoted by d2. This unexpected spacing strongly suggests that the single In layer is in fact an In-O octahedral layer, a result

confirmed by EELS. The oxygen octahedra are oriented in such a way that the O and Zn layers form the regular wurtzite ZnO structure. The polarity of the ZnO slabs either side of the In-O octahedral layer must therefore be opposite, so that the In-O layer is acting as an inversion domain boundary (IDB). In between the In-O layers is another polarity reversal, a mirror domain boundary (MDB) where Zn atoms form a rectangular pattern. It is clear that the microstructure of annealed transparent In_2O_3-ZnO contains closely spaced inversion and mirror domain boundaries as an integral part of its structure.

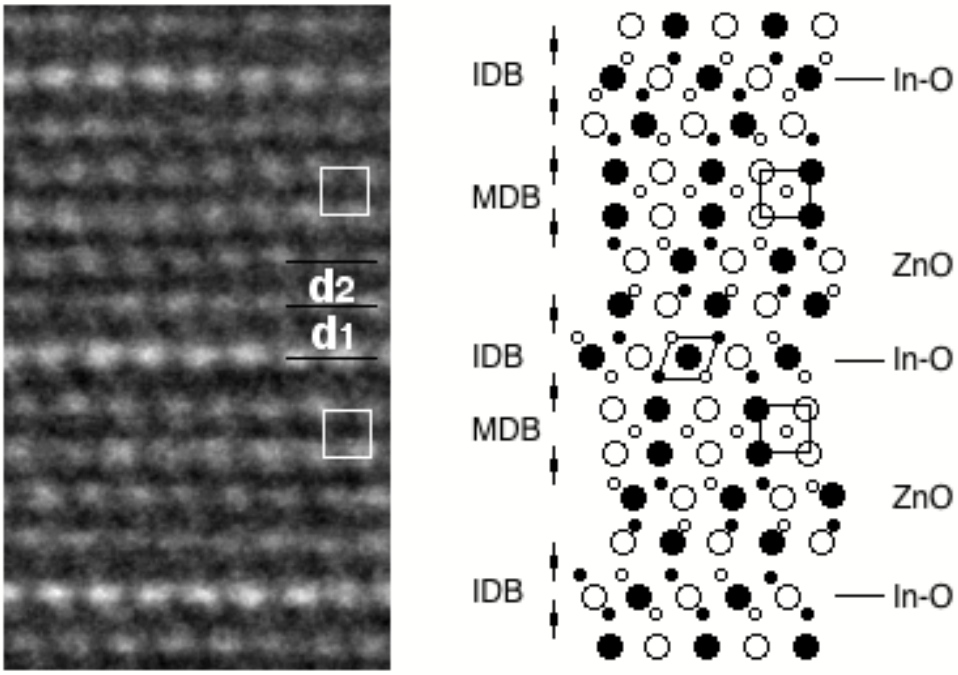

Figure 34. (a) Z-contrast image of an In_2O_3-ZnO film taken along the $[1\bar{1}20]$ zone axis showing a polytypoid structure with an average composition of $In_2O_3(ZnO)_{10}$. (b) structure deduced from the measured cation positions. The single In layer is an In-O octahedral layer inducing an inversion domain boundary (IDB) in the adjacent ZnO layers, which is reversed by the mirror domain boundary (MDB). Large circles denote cations, small circles O, white and black denote different heights.

7. Future Directions

If you consider that the wavelength of a 300 kV electron is less than 0.02 Å, then our achievement of resolutions of 1.3 Å would seem rather poor. Of course, this is due to the enormously large aberrations of the electron lenses which limit the usable apertures to only 10 mrad or so, about half a degree. It has been realized for decades that correction of the spherical aberration would bring great benefits in resolution, and there have been numerous proposals and attempts over the years (Hawkes, 1997,

Dellby, Krivanek and Lupini, 1998). Correctors must resort to elements with non-cylindrical symmetry, and are optically rather complex. Only recently therefore, due in large part to the advent of the computer, have systems been developed with the capability for precise alignment and the necessary stability. Again the STEM appears to have a significant advantage, in that the objective lens is before the sample and we avoid problems with chromatic aberration due to electrons having lost energy in the specimen. Indeed, as the STEM image contrast comes only from overlaps between neighboring illumination discs (see Fig. 12), there is a line down the center of each overlap region that is achromatic, where the path from each disc passes at equal angles through the objective lens. The STEM annular detector image is therefore much more robust towards chromatic aberration effects such as energy spread and fluctuations in objective lens current than the axial phase contrast image (Nellist and Pennycook, 1998c). Because of the rapidly oscillating phase contrast transfer function, such fluctuations lead to an exponential damping envelope, and the so-called information limit is much reduced compared to incoherent imaging.

An example of the anticipated improvement in probe profile is seen in Fig. 35, Not only is the central peak significantly sharpened to a FWHM of 0.5 Å, but the extended probe tails have largely been eliminated. It is precisely these extended tails that were responsible for the intensity seen between the dumbbells in the raw image of Si ⟨110⟩ (Fig 26). For analysis such tails are equally undesirable; with a probe located over a selected column, the tails may put a significant fraction of the total probe current down neighboring columns.

It is clear that if such devices are realized in practice then imaging and analysis will benefit enormously from the increased sensitivity. It is not so much the additional spacings that will become available, it is that we will be able to image and analyze all materials with single atomic column sensitivity. By increasing the current down our selected column, and simultaneously decreasing the current illuminating surrounding columns, we will

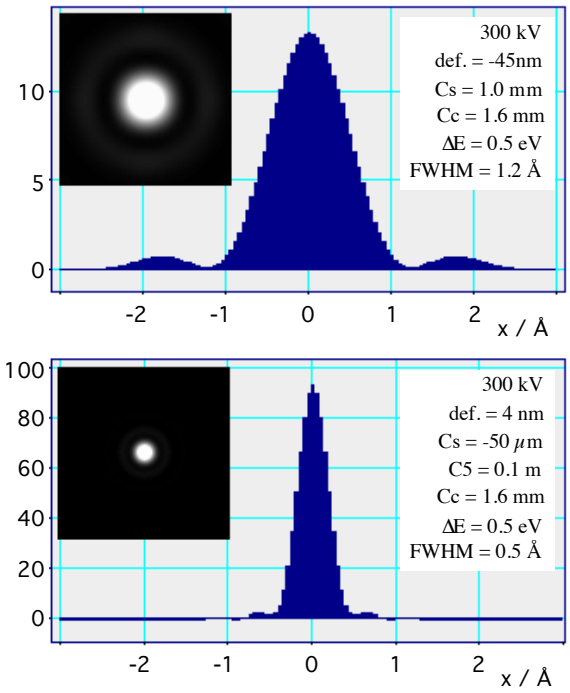

Figure 35. Improvement in probe profile anticipated for the HB603U STEM by correction of spherical aberration. Upper panel shows probe profile and image for our present system, lower panel shows the effect of C_s correction. The integrated intensity in the two probes is equal, but the central peak of the corrected probe is seven times greater. (Courtesy O. Krivanek).

enormously improve the image contrast and the analytical sensitivity. It will be possible to image oxygen columns in the high T_c superconductors, and place the probe over a selected oxygen column to measure local hole concentration. The range of catalyst atoms and clusters visible on the surface of γ-alumina will be greatly extended. It will become possible to image and analyze single impurity atoms in specific columns at a grain boundary or dislocation core, and to see the effect on local electronic structure. Indeed we appear to be on the threshold of finally understanding the atomic origins of materials properties.

8. Acknowledgements

The authors are grateful to their colleagues M. F. Chisholm, Y. Yan, G. Duscher, Y. Xin, H. J. Gao, E. C. Dickey, N. D. Browning, V. P. Dravid, D. E. Jesson, B. Rafferty, N. Tanaka, D. Perovic, A. Maiti and S. T. Pantelides for research collaborations and permission to reproduce results. This research was supported by Lockheed Martin Energy Research Corp. under DOE Contract No. DE-AC05-96OR22464, and by appointments to the ORNL Postdoctoral Research Associates Program administered jointly by ORNL and ORISE.

9. References

Ade, G. (1977) On the incoherent imaging in the scanning transmission electron microsocpe, *Optik* **49**, 113–116.

Allen, L. J. (1993) Electron energy loss spectroscopy in a crystalline environment using inner-shell ionization, *Ultramicroscopy* **48**, 97-106.

Anderson, S. C., Birkeland, C. R., Anstis, G. R., and Cockayne, D. J. H. (1997) An approach to quantitative compositional profiling at near atomic resolution using high-angle annular dark-field imaging, *Ultramicroscopy* **69**, 83–103.

Batson, P. E. (1993) Simultaneous STEM imaging and electron energy-loss spectroscopy with atomic column sensitivity, *Nature* **366**, 727–728.

Browning, N. D., Chisholm, M. F., and Pennycook, S. J. (1993) Atomic-resolution chemical analysis using a scanning transmission electron microscope, *Nature* **366**,143–146.

Browning, N. D., Chisholm, M. F., Nellist, P. D., Pennycook, S. J., Norton, D. P., and Lowndes, D. H. (1993) Correlation between hole depletion and atomic structure at high-angle grain boundaries in $YBa_2Cu_3O_{7-\delta}$., *Physica C* **212**, 185–190.

Browning, N. D. and Pennycook, S. J. (1995) Atomic-resolution electron energy loss spectroscopy in the scanning transmission electron microscope, *J. Microsc.* **180**, 230–237.

Browning, N. D., Buban, J. P., Nellist, P. D., Norton, D. P., Chisholm, M. F., and Pennycook, S. J. (1998) The atomic origins of reduced critical currents at [001] tilt grain boundaries in $YBa_2Cu_3O_{7-\delta}$ thin films, *Physica C* **294**:183–193.

Cannard, P. J. and Tilley, R. J. D. (1988) New intergrowth phases in the ZnO-In_2O_3 system, *J. Solid State Chem.* **73**, 418–426.

Chen, J., Chan, H. M., and Harmer, M. P. (1989) Ordering structure and dielectric properties of undoped and La/Na-doped $Pb(Mg_{1/3}Nb_{2/3})O_3$, *J. Am. Ceram. Soc.* **72**, 593–598.

Chisholm, M. F., Maiti, A., Pennycook, S. J., and Pantelides, S. T. (1998) Atomic configurations and energetics of arsenic impurities in a silicon grain boundary, *Phys. Rev. Lett.* **81**, 132–135.

Cowley, J. M. (1969) Image contrast in a transmission scanning electron microscope, *Appl. Phys. Lett.* **15**, 58–59.

Cowley, J. M. (1986) Principles of image formation in J. J. Hren, J. I. Goldstein, and D. C. Joy (eds.), *Principles of Analytical Electron Microscopy*, Plenum Press, New York, New York, pp. 343–368.

Cowley, J. M. (1976) Scanning transmission electron microscopy of thin specimens, *Ultramicroscopy* **2**, 3–16.

Crewe, A. V., Wall, J., and Langmore, J. (1970) Visibility of single atoms, *Science* **168**, 1338–1340.

Crewe, A. V. and Salzman, D. B. (1982) On the optimum resolution for a corrected STEM, *Ultramicroscopy* **9**, 373–378.

Datye, A. K. and Smith, D. J. (1992) The study of heterogeneous catalysts by high-resolution transmission electron microscopy, *Catal. Rev.-Sci. Eng.* **34**, 129–178.

Dellby, N., Krivanek, O. L. and Lupini, A. R. (1998) Aberration correction in the STEM, *Ultramicroscopy*, (in press).

Dickey, E. C., Dravid, V. P., Nellist, P. D., Wallis, D. J., and Pennycook, S. J. (1998) Three-dimensional atomic structure of NiO-ZrO_2(cubic) interfaces, *Acta Mater.* **46**, 1801–1816.

Duscher, G., Browning, N. D., and Pennycook, S. J. (1998) Atom column resolved electron energy loss spectroscopy, *Phys. Stat. Sol. (a)* **166**, 327–342.

Duscher, G., Pennycook, S. J., Browning, N. D., Rupangudi, R., Takoudis, C., Gao,H-J., and Singh, R. (1998) Structure, composition and strain profiling of Si/SiO_2 interfaces, Proceedings of International Conference on Characterization and Metrology for ULSI Technology, Gaithersburg, Maryland, March 23–27, pp 191-195.

Engel, A., Wiggins, J. W., and Woodruff, D. C. (1974) A comparison of calculated images generated by six modes of transmission electron microscopy, *J. Appl. Phys.* **45**, 2739–2747.

Grigorian, I., Williams, K. A., Fang, S., Sumanasekera, G. U., Loper, A. L., Dickey, E. C., Pennycook, S. J. and Ecklund, P. C. (1998) Reversible intercalation of charged iodine chains into carbon nanotube ropes, *Phys. Rev. Letts***80**, 5560–5563.

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

Hall, C. R. and Hirsch, P. B. (1965) Effect of thermal diffuse scattering on propagation of high energy electrons through crystals, *Proc. Roy. Soc. A* **286**, 158–177.

Hall, C. R., Hirsch, P. B., and Booker, G. R. (1966) Effects of point defects on absorption of high energy electrons passing through crystals, *Phil. Mag.* **14**, 979–989.

Harris, D. J., Watson, G. W., and Parker, S. C. (1996) Atomistic simulation of the effect of temperature and pressure on the [001] symmetric tilt grain boundaries of MgO, *Phil. Mag.* **A74**, 407–418.

Hartel, P., Rose, H., and Dinges, C. (1996) Conditions and reasons for incoherent imaging in STEM, *Ultramicroscopy* **63**, 93–114.

Hawkes, P.W. (1997) Aberrations, in J. Orloff, (ed.), *Handbook of Charged Particles Optics* , CRC Press, Boca Raton and New York.

Holbrook, O. F. and Bird, D. M. (1995) Theoretical modelling of the size and shape of atomic images formed with inelastically scattered electrons, *Proceedings of Microscopy and Analysis 1995* , Jones and Begall, New York, pp. 278-279.

Howie, A. (1979) Image contrast and localized signal selection techniques, *J. Microsc.* **117**, 11-23.

Kingery, W. D. (1974) Plausible concepts necessary and sufficient for interpretation of ceramic grain-boundary phenomena: I, grain-boundary characteristics, structure, and electrostatic potential, *J. Am. Cer. Soc.*, **57**, 1–8.

Isaacson, M. S., Ohtusuki, M., and Utlaut, M. (1979) Electron microscopy of individual atoms, in J. J. Hren, J. I. Goldstein, and D. C. Joy (eds.), *Introduction to Analytical Electron Microscopy*, Plenum Press, New York, New York, pp. 343–368.

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

Jesson, D. E. and Pennycook, S. J. (1995) Incoherent imaging of crystals using thermally scattered electrons, *Proc. Roy. Soc. (London) A* **449**, 273–293.

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

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

Maiti, A., Chisholm, M. F., Pennycook, S. J., and Pantelides, S. T. (1996) Dopant segregation at semiconductor grain boundaries through cooperative chemical rebonding. *Phys. Rev. Lett.* **77**, 1306–1309.

Maslen, V. W. and Rossouw, C. J. (1984) Implications of (e, 2e) scattering for inelastic electron diffraction in crystals: I. theoretical, *Phil. Mag. A* **49**, 735-742.

McGibbon, A. J., Pennycook, S. J., and Jesson, D.E. (1998) Crystal structure retrieval by maximum entropy analysis of atomic resolution incoherent images, *J. Microsc.* (in press).

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

Nakamura, K., Kakibayashi, H., Kanehori, K., and Tanaka, N., (1997) Position dependence of the visibility of a single gold atom in HAADF-STEM image simulation, *J. Electron Microsc.* **1**, 33–43.

Nellist, P. D. and Pennycook, S. J. (1996) Direct imaging of the atomic configuration of ultra-dispersed catalysts, *Science* **274**, 413–415.

Nellist, P. D. and Pennycook, S. J. (1998a) Incoherent imaging using dynamically scattered coherent electrons, *Ultramicroscopy,* (in press).

Nellist, P. D. and Pennycook, S. J. (1998b) Accurate structure determination from image reconstruction in ADF STEM, *J. Microscopy* **190**, 159-170.

Nellist, P. D. and Pennycook, S. J. (1998c) Sub-ångstrom resolution TEM through under-focussed incoherent imaging, *Phys. Rev. Lett.* (in press).

Pennycook, S. J. and Boatner, L. A. (1988) Chemically sensitive structure imaging with a scanning transmission electron microscope, *Nature* **336**, 565–567.

Pennycook, S. J. and Jesson, D. E. (1990) High-resolution incoherent imaging of crystals, *Phys. Rev. Lett.* **64**, 938–941.

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

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

Pennycook, S. J., Jesson, D. E., and Nellist, P. D. (1996) High angle dark field STEM for advanced materials, *J. Electron Microsc.* **45**, 36-43.

Pennycook, S. J., Jesson, D. E., Nellist, P. D., Chisholm, M. F., and Browning, N. D. (1997) Scanning transmission electron microscopy: *Z*-contrast, in S. Amelinckx, D. van Dyck, J. van Landuyt, and G. van Tendeloo (eds), *Handbook of Microscopy*, VCH Publishers, Weinheim, Germany, pp. 595–620.

Perovic, D. D., Rossouw, C. J. and Howie, A., (1993) Imaging inelastic strains in high-angle annular dark field scanning transmission electron microscopy, *Ultramicroscopy* **52**, 353-359.

Perovic, D. D., Howie, A., and Rossouw, C. J. (1993) On the image contrast from dislocations in high-angle annular dark-field scanning transmission electron microscopy, *Phil. Mag. Lett.* **67**, 261–277.

Rafferty, B. and Pennycook, S. J. (1998) Towards atomic column-by-column spectroscopy, *Ultramicroscopy* (in press).

Rayleigh, Lord (1896) On the theory of optical images with special reference to the microscope, *Phil. Mag.* (5) **42**, 167–195.

Ritchie, R. H. and Howie, A. (1988) Inelastic scattering probabilities in scanning transmission electron microscopy, *Phil. Mag. A* **58**, 753–767.

Rose, H. (1976) Image formation by inelastically scattered electrons in electron microscopy, *Optik* **45**, 139–158 and 187–208.

Rossouw. C. J. and Maslen, V. W. (1984) Implications of (e, 2e) scattering for inelastic electron diffraction in crystals II, application of the theory, *Phil. Mag. A* **49**, 743-757.

Scherzer, O. (1949) The theoretical resolution limit of the electron microscope, *J. Appl. Phys.* **20**, 20–29.

Spence, J. C. H. and Cowley, J. M. (1978) Lattice imaging in STEM, *Optik* **50**, 129-142.

Tanaka, N. et al., (1997) High-resolution observation of $Si_{100-x}V_x$ amorphous alloys by HAADF-STEM, *Microscopy and Microanalysis* **3**, Suppl. 2 (Proc. Microsc. And Microan. '97, Springer, New York, 1997), pp 719–720.

Treacy M. M. J., Howie, A. and Wilson, C. J. (1978) Z-contrast of Platinum and Palladium Catalysts, *Phil. Mag.* **38**, 569–585.

Treacy, M. M. J., Howie, A. and Pennycook, S. J. (1980) Z-contrast of supported catalyst particles on the STEM, in *Electron Microscopy and Analysis, 1979*, T. Mulvey (ed.), (Inst. Phys. Conf. Ser. No. 52, Institute of Physics, London and Bristol) pp. 261-264.

Warren, B. E. (1990) *X-Ray Diffraction*, Dover, New York.

Xin, Y., Pennycook, S. J., Browning, N. D., Nellist, P. D., Sivananthan, S., Omnès, F., Beaumont, B., Faurie, J.-P. and Gibart P. (1998) Direct observation of the core structures of threading dislocations in GaN, *Appl. Phys. Lett.* **72**, 2680-2682.

Yan, Y., Chisholm, M. F., Duscher, G., Maiti, A., Pennycook, S. J., and Pantelides, S. T. (1998) Impurity induced structural transformation of a MgO grain boundary, *Phys. Rev. Lett.* **81**, 3675-3678.

Yan, Y., Pennycook, S. J., Xu, Z., and Viehland, D. (1998) Determination of the ordered structures of $Pb(Mg_{1/3}Nb_{2/3})O_3$ and $Ba(Mg_{1/3}Nb_{2/3})O_3$ by atomic-resolution *Z*-contrast imaging *Appl. Phys. Lett* **72**, 3145–3147.

Yan, Y., Pennycook, S. J., Dai, J., Chang, R. P. H., Wang, A., and Marks, T. J. (1998) Polytypoid structures in annealed In_2O_3-ZnO Films, *Appl. Phys. Lett.* 2585-2587.

Yan, Y., Pennycook, S. J. and Tsai, A. P. (1998) Direct imaging of local chemical disorder and columnar vacancies in ideal decagonal Al-Ni-Co quasicrystals, *Phys. Rev. Lett.* (in press).

Zeitler, E. and Thomson, M. G. R. (1970) Scanning transmission electron microscopy. *Optik* **31**, 258–280 and 359–366.