

The Principles and Interpretation of Annular Dark-Field Z-Contrast Imaging

P. D. NELLIST¹ AND S. J. PENNYCOOK²

¹*Nanoscale Physics Research Laboratory, School of Physics and Astronomy,
The University of Birmingham, Birmingham, B15 2TT, UK.*

²*Oak Ridge National Laboratory, Solid State Division, PO Box 2008,
Oak Ridge, TN 37831-6030, USA*

I. Introduction	148
A. Introduction to ADF STEM	148
B. Coherent and Incoherent Imaging	150
C. Outline	153
II. Transverse Incoherence	153
A. STEM Image Formation	153
B. The Conditions for Incoherent Imaging	157
C. The Resolution Limit	160
D. The Thin Specimen Object Function	161
E. Dynamical Scattering Using Bloch Waves	165
III. Longitudinal Coherence	170
A. Kinematical Scattering	171
B. Dynamical Scattering	173
C. HOLZ Effects	174
D. Thermal Diffuse Scattering	175
IV. The Ultimate Resolution and the Information Limit	181
A. Underfocused Microscopy	181
B. Chromatic Aberrations	185
C. Source Size and the Ultimate Resolution	189
V. Quantitative Image Processing and Analysis	190
A. The Absence of a Phase Problem	190
B. Probe Reconstruction	191
C. Deconvolution Methods	192
VI. Conclusions	197
A. Overview	197
B. Future Prospects	198
References	200

I. INTRODUCTION

A. Introduction to ADF STEM

The purpose of this paper is to describe how an annular dark-field (ADF) image is formed in a scanning transmission electron microscope (STEM), and to use that understanding to explain how the image data may be used to provide atomic-resolution information about the specimen. We start with a brief description of a STEM and the ADF detector.

A STEM is in principle very similar to the more commonly known scanning electron microscope (SEM) in that electron optics are used *before* the specimen to focus an electron beam to form an illuminating spot, or *probe* (Figure 1), that is scanned over the specimen in a raster fashion (see Crewe [1980] for a description of the physics of STEM). Various signals produced by the scattering of the electrons can be detected and displayed as

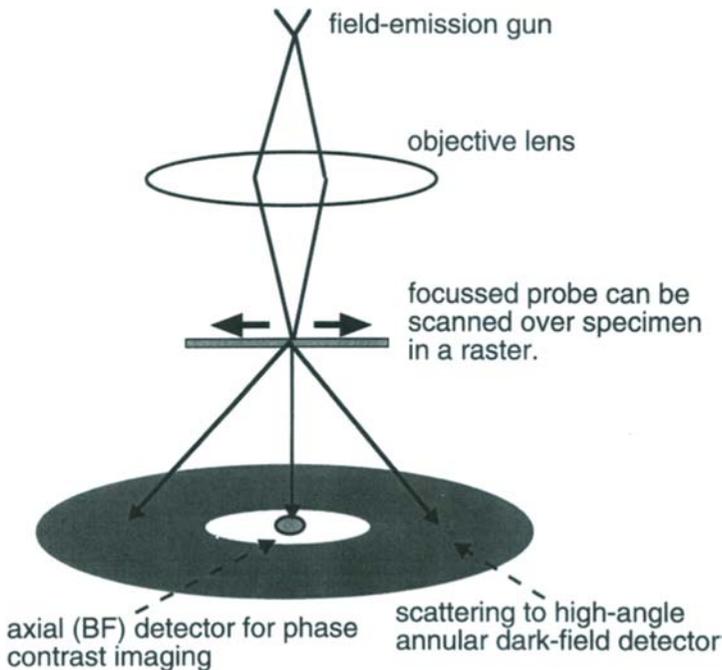


FIGURE 1. Schematic of the scanning transmission electron microscope (STEM) showing the geometry of the annular dark-field (ADF) detector, and the bright-field (BF) detector for phase-contrast imaging.

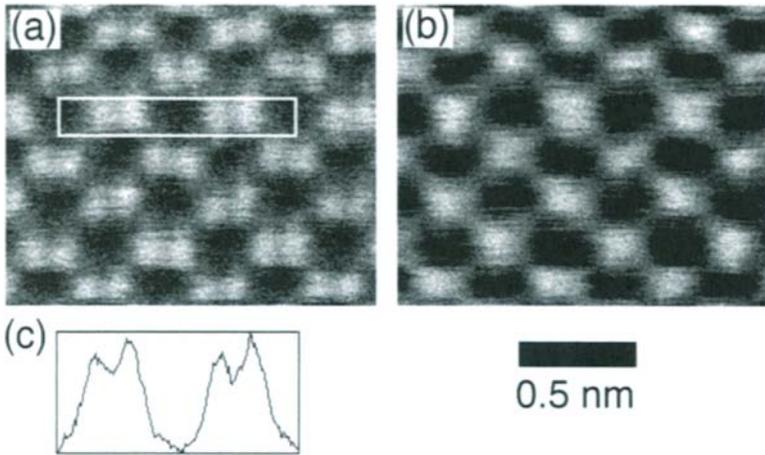


FIGURE 2. Simultaneously recorded (a) ADF and (b) BF images of GaAs $\langle 110 \rangle$. The BF image shows lower resolution than the ADF image, and in this case the atoms are black contrast. The ADF intensity profile (c) shows the polarity of the lattice through the Z-contrast.

a function of the illuminating probe position. The major difference between SEM and STEM is that a thin, electron-transparent specimen is used in STEM, allowing transmitted electrons to be detected. Since there is very little scattering of the electrons in a thin sample, little beam spreading occurs and the spatial resolution is mainly controlled by the illuminating probe size. Typical electron-optical parameters for STEM are comparable to the conventional high-resolution transmission electron microscope, so that typical accelerating voltages are in the range 100–300 kV and the probe forming lens (known as the *objective lens*) has a coefficient of spherical aberration of the order of 1 mm. With such parameters, the illuminating probe can have a dimension similar to that of an atom (a few ångströms) and atomic resolution imaging is possible.

Annular dark-field imaging refers to the use of a particular detector geometry in STEM. A geometrically large annular detector is placed in the optical far field beyond the specimen (Figure 1). The total intensity detected over the whole detector is recorded and displayed as a function of the position of the illuminating probe. Since the detector only receives a signal when the specimen is present, the vacuum appears dark, hence the name, and the heavier the atom, the higher the intensity of the scattering, which leads to atomic number (Z) contrast in the image. The most important feature of ADF imaging is that it can be described as being *incoherent*, which has many advantages at atomic resolution. Figure 2 shows the incoherent

and Z-contrast nature of an ADF image, and also how the phase contrast, bright field (BF) image can be recorded simultaneously for comparison. A major purpose of this paper will be to explain how the incoherence arises, why it is important, and how it may be used. The first step is to understand the difference between coherent and incoherent imaging.

B. Coherent and Incoherent Imaging

1. High-Resolution TEM

Hitherto, the majority of TEM imaging at atomic resolution has been performed using conventional high-resolution TEM (HRTEM) [Spence, 1988], and there is a huge range of applications (see Smith [1997] for a review). Conventional HRTEM is a coherent mode of imaging. The basic principle is that an electron transparent sample is illuminated by coherent plane-wave illumination. The exit surface wavefunction, $\psi(\mathbf{R})$, is then magnified by the objective lens, along with further lenses to provide additional magnification, to form a highly magnified image. Due to the inherent lens aberrations, especially spherical aberration [Scherzer, 1936], there is a blurring of the exit surface wavefunction as it propagates to the magnified image plane. This blurring can be written as a convolution with a point-spread function, $P(\mathbf{R})$, and what is actually measured by a recording medium in the image plane is the *intensity* of this convolution,

$$I_{\text{coh}}(\mathbf{R}) = |P(\mathbf{R}) \otimes \psi(\mathbf{R})|^2 \quad (1)$$

This equation is the mathematical definition of coherent imaging. The convolution of the exit-surface wavefunction with the point-spread function is in *complex amplitude*, which means that the scattering from spatially separated parts of the specimen can interfere in the blurring process. In practice, this coherent convolution means that the image intensity can fluctuate dramatically as the point-spread function is changed by changing the focus, for instance. Contrast reversals can occur depending on whether constructive or destructive interference is occurring, and there is uncertainty over whether atoms should appear as bright or dark contrast in the image [Spence, 1988], which can change depending on the focusing condition. In Figure 2, the atoms appear dark in the HRTEM image, but this is not always the case.

The situation is further complicated by the existence of dynamical diffraction [Bethe, 1928], which can also have a strong effect on the coherent HRTEM image contrast, causing, for instance, contrast reversals as the thickness is changed (for example, Glaisher *et al.* [1989]). As the coherent

electron wavefunction propagates through the crystal, the diffracted beams can be multiply scattered back onto one another leading to strong and complicated reinterference effects. In general, the exit-surface wavefunction cannot be directly interpreted in terms of the specimen structure, especially when defects such as grain boundaries are being studied [Bourret *et al.*, 1988]. Generally, an approach is taken where images are simulated from trial structure models and are systematically compared with the experimental data [Möbus, 1996; Möbus and Dehm, 1996]. For this approach, the microscope imaging parameters need to be accurately known, and there is the ever-present danger that the correct structure will never be tried as a trial object or that an iterative process will find an incorrect local minimum.

It is interesting to speculate as to why coherent imaging has hitherto dominated TEM. The reason may be partly historical and partly because of the instrumentation. Other than high resolution imaging, conventional TEM machines are also required to form diffraction patterns and diffraction contrast images. Both of these applications require coherent illumination to allow interference to occur to form the diffraction pattern. Before the advent of high-resolution imaging, these modes dominated the use of TEM. Additionally, HRTEM is a phase-contrast technique, which requires high coherence. As spatial coherence was lost due to the beam divergence increasing, users will have observed a loss in contrast, and the tendency would therefore have been to keep the spatial coherence as high as possible. The implementation of atomic resolution incoherent imaging has required the development of dedicated STEM machines capable of working at atomic resolution. Such machines have not been nearly as common as conventional TEMs, and thus most users have not had the chance to experience incoherent imaging.

2. Incoherent Imaging

In his classic paper on the resolution limit of the microscope, Lord Rayleigh [1896] described the difference between coherent and incoherent imaging. He pointed out that if a transmission specimen was illuminated with light from a large source giving illumination over a wide range of angles, the specimen could be treated as being self-luminous, and that interference between the radiation emitted from spatially separated parts of the specimen could not interfere, and is thus incoherent. The image then becomes a convolution in *intensity* rather than in complex amplitude,

$$I_{\text{incoh}}(\mathbf{R}) = |P(\mathbf{R})|^2 \otimes |\psi(\mathbf{R})|^2 \quad (2)$$

which is the mathematical definition of incoherent imaging. Compare this with Equation (1). Lord Rayleigh also noted that the resolution limit of the

microscope in the incoherent regime was twice that of a coherently illuminated microscope. The vast majority of images that we observe with our eyes are incoherent images, because most light sources are extended and incoherent. Lord Rayleigh pointed out that incoherent images do not show the sharp interference bands that characterize coherent images, and are therefore much simpler to interpret. It is because of their incoherent nature, that we are able to interpret the images that our eyes see. If we were surrounded by coherent illumination, the strong interference effects that we would observe would be very confusing. We might have to guess at what was near us, and perform an image simulation to compare with the image that our eyes observed!

By remembering that the difficulties in conventional HRTEM, both the strong dependence on the imaging parameters and the problems of dynamical diffraction effects, are mostly born from the coherent nature of the illumination, it is clear that incoherent TEM imaging should hold major benefits. The first steps in this direction were taken in the development of annular dark-field (ADF) imaging in the STEM. An ADF detector had been used in the first STEM developed [Crewe *et al.*, 1968a; Crewe *et al.*, 1970] to collect a signal that was assumed to be largely elastically scattered. Because relatively high-angle scattering was collected by the ADF detector, the signal showed strong atomic number (Z) contrast, although the name Z -contrast was first applied to a signal derived from the ratio of the ADF to inelastic scattering. Later work showed that if any coherently scattered Bragg beams were incident on the ADF detector, the contrast could not purely be related to atomic number [Donald and Craven, 1979]. Various papers then suggested that increasing the angle of the ADF detector such that only incoherent thermal diffuse scattering, rather than coherent Bragg beams, reached the detector could ameliorate the coherent effects in the image [Treacy *et al.*, 1978; Howie, 1979]. The steady improvement in STEM performance eventually resulted in the first ADF images being taken at atomic resolution [Pennycook and Boatner, 1988; Pennycook and Jesson, 1990; Shin *et al.*, 1989]. These results, and Figure 2, illustrate how ADF STEM imaging gives a direct structure image with peaks at the atom sites, without the contrast reversals seen in coherent imaging. Since this initial work there has been a growing interest in the application and interpretation of ADF imaging, with numerous examples now in the literature. Inevitably, the physics behind ADF imaging is more complicated than simply asserting that incoherent imaging requires the collection of incoherent thermally scattered electrons; for example, the detector geometry itself can impose incoherence. The purpose of this article is to present the current understanding of the principles of ADF imaging, and the way the images may be interpreted.

C. Outline

In this article, we have split the discussion of the destruction of coherence in the image into two components parallel and perpendicular to the beam direction, which we refer to as transverse and longitudinal coherence, respectively. In Section II we describe how it is the detector geometry that destroys the transverse coherence, even when purely coherent Bragg scattering is being collected by the ADF detector. This destruction of transverse coherence immediately allows direct interpretation of peaks in the image as being atoms or atomic columns in the specimen, which is extremely important for structure determination.

In Section III it is shown that the detector geometry only weakly destroys coherence parallel to the beam, and that we must rely on phonon scattering to help remove longitudinal coherence effects. This is important in using the Z-contrast nature of the image for compositional determination, and we discuss the current state in this use of ADF images.

Since ADF imaging is very different to conventional HRTEM, we need to examine the effects of other sources of incoherence, such as the source size and the chromatic aberrations. A treatment of such effects is given in Section IV, and we examine what might limit the ultimate ADF resolution. Finally, in Section V we describe how the ADF image might be used quantitatively to accurately determine atomic column positions using the incoherent nature of the image, before drawing some conclusions and putting forward some future prospects in Section VI.

Throughout this article the Fourier transform of a real-space function, $F(\mathbf{R})$, will often be written $\tilde{F}(\mathbf{Q})$, where \mathbf{Q} is a spatial frequency vector in reciprocal space. For consistency, the imaging parameters of a VG Microscopes HB603U STEM will be used in calculations. For this microscope, the accelerating voltage is 300 kV and the coefficient of spherical aberration, C_s , is 1 mm.

II. TRANSVERSE INCOHERENCE

A. STEM Image Formation

We start by developing a general formulation of imaging in a STEM that can be applied to imaging both elastic and inelastic scattering events. For elastic scattering, this formulation also holds for CTEM imaging through the principle of reciprocity [Cowley, 1969; Zeitler and Thomson, 1970], which allows the source and detector planes to be interchanged. The detector in the image plane of a CTEM is equivalent to the source plane in

a STEM, and a certain detector geometry in a STEM corresponds to an illuminating source geometry in a CTEM. Thus a STEM can be thought of as a conventional TEM with the electrons propagating in the reverse direction. It is now clear that a small axial BF detector in a STEM (Figure 1) is equivalent to axial illumination in a TEM, and thus an HRTEM image will be formed with such a STEM detector. Unless explicitly stated, in this article we use the STEM point of view. A partial plane wave in the cone of coherent illumination is focused by the objective lens to form the illuminating probe (Figure 3). This partial plane wave can be labeled by the transverse component, \mathbf{K}_i , of its wavevector; by transverse we mean perpendicular to the optic axis of the microscope. Because of the objective lens aberrations, each partial plane wave will have experienced a phase shift, $\chi(\mathbf{K}_i)$, relative to the $\mathbf{K}_i = \mathbf{0}$ wave, which in the case of spherical aberration

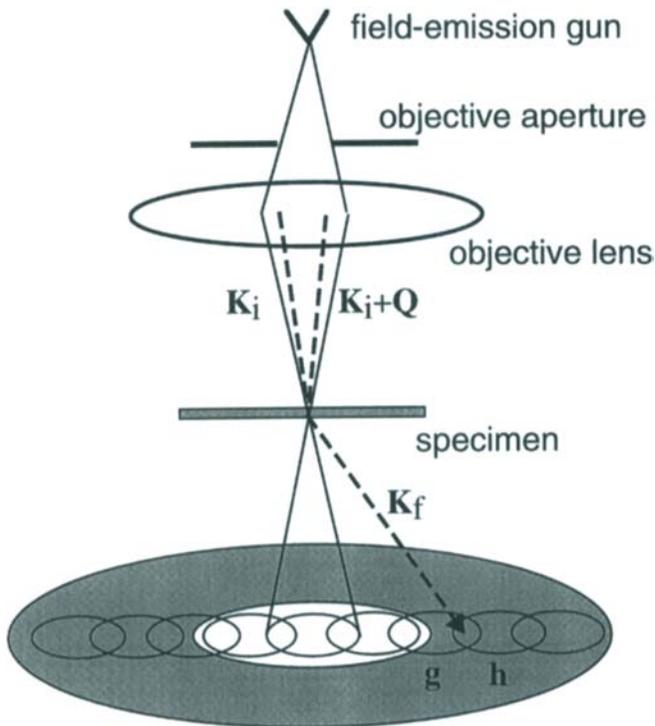


FIGURE 3. Contrast in a STEM image arises from interference between beams in the incident cone that are scattered into the same final wavevector and interfere. For Bragg beams, this interference can only occur in diffraction disc overlap regions.

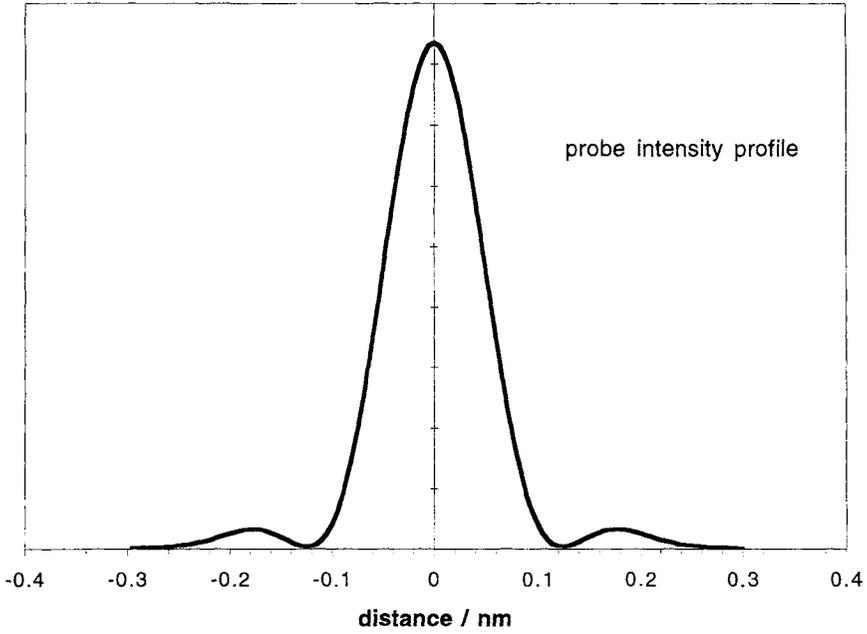


FIGURE 4. The intensity profile of the illuminating probe at Scherzer defocus for the HB603U STEM.

and defocus is

$$\chi(\mathbf{K}_i) = \pi\lambda z|\mathbf{K}_i|^2 + \frac{1}{2}\lambda^3 C_s |\mathbf{K}_i|^4 \quad (3)$$

where λ is the electron wavelength, z is the defocus, and C_s is the coefficient of spherical aberration. Other terms, such as astigmatism and coma, could also be included in this phase function. An objective aperture, described by a circular top-hat support function $H(\mathbf{K}_i)$, is normally used to prevent strongly aberrated, high-angle beams contributing to the image forming process. The functions $H(\mathbf{K}_i)$ and $\chi(\mathbf{K}_i)$ can be combined as the magnitude and phase, respectively, of a complex aperture function, $A(\mathbf{K}_i)$. The illuminating STEM probe wavefunction, $P(\mathbf{R})$ (Figure 4), is a sum over all the partial plane waves,

$$P(\mathbf{R}) = \int A(\mathbf{K}_i) \exp[-i2\pi\mathbf{K}_i \cdot \mathbf{R}] d\mathbf{K}_i \quad (4)$$

which we take as the definition of the inverse Fourier transform of $A(\mathbf{K}_i)$.

Scan coils are present in the STEM to move the probe across the specimen, and such a shift can be incorporated by multiplying $A(\mathbf{K}_i)$ by $\exp(i2\pi\mathbf{K}_i \cdot \mathbf{R}_0)$, which gives $P(\mathbf{R} - \mathbf{R}_0)$ when substituted in Equation (4).

Consider now a scattering event that scatters from the initial partial plane wave, \mathbf{K}_i , into a final plane wave, with wavevector transverse component \mathbf{K}_f , changing the magnitude and phase of the wave by a complex multiplier, $\Psi(\mathbf{K}_f, \mathbf{K}_i)$. Assuming that there is no loss of coherence, we can write the intensity measured in the far field as

$$I(\mathbf{K}_f, \mathbf{R}_0) = \left| \int A(\mathbf{K}_i) \exp[-i2\pi\mathbf{K}_i \cdot \mathbf{R}_0] \Psi(\mathbf{K}_f, \mathbf{K}_i) d\mathbf{K}_i \right|^2 \quad (5)$$

because \mathbf{K}_f defines a position in the far field. Expanding the square gives a double integral

$$I(\mathbf{K}_f, \mathbf{R}_0) = \iint A(\mathbf{K}_i) A^*(\mathbf{K}_i') \exp[i2\pi(\mathbf{K}_i - \mathbf{K}_i') \cdot \mathbf{R}_0] \Psi(\mathbf{K}_f, \mathbf{K}_i) \Psi^*(\mathbf{K}_f, \mathbf{K}_i') d\mathbf{K}_i d\mathbf{K}_i' \quad (6)$$

However, this expression can be reduced back to a single integral by taking the Fourier transform of Equation (6) [Rodenburg and Bates, 1992] to give a function entirely in reciprocal space

$$\tilde{I}(\mathbf{K}_f, \mathbf{Q}) = \int A(\mathbf{K}_i) A^*(\mathbf{K}_i + \mathbf{Q}) \Psi(\mathbf{K}_f, \mathbf{K}_i) \Psi^*(\mathbf{K}_f, \mathbf{K}_i + \mathbf{Q}) d\mathbf{K}_i \quad (7)$$

which is the Fourier transform of the image intensity that would be recorded for a point detector at position \mathbf{K}_f in the far field. From Equation (7) it is clear that the contributions to an image spatial frequency, \mathbf{Q} , come from pairs of incident partial plane-waves separated by the reciprocal-space vector \mathbf{Q} (Figure 3). These two partial plane waves are scattered by the specimen into the same final wavevector, \mathbf{K}_f , where they interfere. If the scattering is purely Bragg diffraction from a crystalline specimen, a series of Bragg discs will be seen in the form of a coherent convergent beam electron diffraction (CBED) pattern. In the disc overlap regions, two incident partial plane waves are being scattered into a single, final wavevector, and interference can occur. Since the probe position, \mathbf{R}_0 , is defined in reciprocal space by a linear phase variation over the incident wave, the phase difference between the two incident waves will be $2\pi\mathbf{Q} \cdot \mathbf{R}_0$. Scanning the probe, therefore, means that the interference between the two incident partial plane waves will cause the intensity to oscillate at a rate given by \mathbf{Q} , leading to image contrast. Thus STEM lattice imaging depends on the detection of interference in the overlap regions of diffracted discs, as noted by Spence and Cowley [1978].

B. The Conditions for Incoherent Imaging

In his discussion of incoherent imaging, Lord Rayleigh [1896] suggested that illuminating an object with incoherent illumination over a large range of angles rendered the object effectively self-luminous, and destroyed any interference between scattering from spatially separated parts of the object. Remembering that a large source in a conventional microscope is equivalent to a large detector in a STEM, we now need to include the effects of the detector. At this point it is easiest to consider a very thin specimen that can be treated as simply multiplying the illuminating wave by a complex function, $\psi(\mathbf{R})$, that describes the magnitude and phase change of the transmitted wave. The Fourier transform of $\psi(\mathbf{R})$ is written, $\Psi(\mathbf{Q})$. The Fourier transform of the image that would be recorded using an annular dark-field detector can now be formed simply by integrating Equation (7) over some detector function, $D(\mathbf{K}_f)$, so that

$$\begin{aligned}\tilde{I}(\mathbf{Q}) &= \int D(\mathbf{K}_f) \int A(\mathbf{K}_i) A^*(\mathbf{K}_i + \mathbf{Q}) \Psi(\mathbf{K}_f - \mathbf{K}_i) \Psi^*(\mathbf{K}_f - \mathbf{K}_i - \mathbf{Q}) d\mathbf{K}_i d\mathbf{K}_f \\ &= \int A(\mathbf{K}_i - \mathbf{Q}/2) A^*(\mathbf{K}_i + \mathbf{Q}/2) \\ &\quad \times \int D(\mathbf{K}_f) \Psi(\mathbf{K}_f - \mathbf{K}_i + \mathbf{Q}/2) \Psi^*(\mathbf{K}_f - \mathbf{K}_i - \mathbf{Q}/2) d\mathbf{K}_f d\mathbf{K}_i\end{aligned}\quad (8)$$

with the shift of \mathbf{K}_i by $-\mathbf{Q}/2$ allowed because the integral has an infinite limit. This is the expression derived by Loane *et al.* [1992]. The domain of integration over \mathbf{K}_i is limited by $A(\mathbf{K}_i - \mathbf{Q}/2) A^*(\mathbf{K}_i + \mathbf{Q}/2)$ to be just the region of overlap between two objective apertures separated by \mathbf{Q} . If $D(\mathbf{K}_f)$ has a geometry that is much larger than the objective aperture, the dependence of the \mathbf{K}_f integral on \mathbf{K}_i becomes very small allowing the integrals to be separated, thus,

$$\begin{aligned}\tilde{I}_{\text{ADF}}(\mathbf{Q}) &= \int A(\mathbf{K}_i - \mathbf{Q}/2) A^*(\mathbf{K}_i + \mathbf{Q}/2) d\mathbf{K}_i \\ &\quad \times \int D(\mathbf{K}_f) \Psi(\mathbf{K}_f - \mathbf{K}_i + \mathbf{Q}/2) \Psi^*(\mathbf{K}_f - \mathbf{K}_i - \mathbf{Q}/2) d\mathbf{K}_f \\ &= T(\mathbf{Q}) \tilde{O}(\mathbf{Q})\end{aligned}\quad (9)$$

where $T(\mathbf{Q})$ is the transfer function for incoherent imaging. By analogy with light optics [Born and Wolf, 1980], this transfer function is referred to as the optical transfer function (OTF), and it contains information about the objective lens defocus, aberrations, and the aperture. Examples of OTFs for

various degrees of spherical aberration have been calculated by Black and Linfoot [1957]. The function, $\tilde{O}(\mathbf{Q})$ is the Fourier transform of the object function, containing information about the specimen scattering and the detector. We will examine its form for various specimen approximations later.

Since Equation (9) is a product in reciprocal-space, taking its Fourier transform results in a convolution in real-space between the Fourier transforms of $T(\mathbf{Q})$ and the object function spectrum, $\tilde{O}(\mathbf{Q})$. The Fourier transform of $T(\mathbf{Q})$, however, is simple to interpret. In Equation (9) it can be seen that $T(\mathbf{Q})$ is the autocorrelation of $A(\mathbf{K}_i)$. The Fourier transform of the autocorrelation of a function is equal to the modulus squared of the Fourier transform of that function. Since the Fourier transform of A is, from Equation (4), the probe function, $P(\mathbf{R})$, we can now write the ADF image intensity as

$$I_{\text{ADF}}(\mathbf{R}_0) = |P(\mathbf{R}_0)|^2 \otimes O(\mathbf{R}_0) \quad (10)$$

which is the incoherent imaging model as also given in Equation (2). So the image intensity can be straightforwardly interpreted as being the convolution between an object function and a positive-real point-spread function (PSF) that is simply the intensity of the illuminating probe. The convolution integral is, therefore, summing over the intensity of the probe, because interference effects between spatially separated parts of the probe are no longer observed, just as if the specimen were self-luminous. Hence the name *incoherent imaging*. The OTF, $T(\mathbf{Q})$, is therefore simply the Fourier transform of the probe intensity function, $|P(\mathbf{R})|^2$. We have plotted the OTF at optimum defocus (-40 nm) (Figure 5) compared to the conventional weak-phase object approximation (WPOA) contrast transfer function (CTF) for the same parameters.

Using this thin specimen approximation, Equation (7) shows that the object function is

$$\tilde{O}(\mathbf{Q}) = \int D(\mathbf{K}_r) \Psi(\mathbf{K}_r - \mathbf{K}_i + \mathbf{Q}/2) \Psi^*(\mathbf{K}_r - \mathbf{K}_i - \mathbf{Q}/2) d\mathbf{K}_r \quad (11)$$

where we require that the integral has no dependence on \mathbf{K}_i to allow the separation of the integrals in Equation (9). The domain of \mathbf{K}_i is a region defined by the overlap of two apertures separated by \mathbf{Q} (Figure 3). If $\Psi(\mathbf{K})$ is varying negligibly slowly over the scale of this domain, then the independence of Equation (11) on \mathbf{K}_i will be fulfilled no matter what the geometry of the detector. A slowly varying $\Psi(\mathbf{K})$ corresponds to a transmittance function, $\psi(\mathbf{R})$, that scatters from a highly localized source. Since the scattering is from a source much smaller than the probe geometry, interfer-

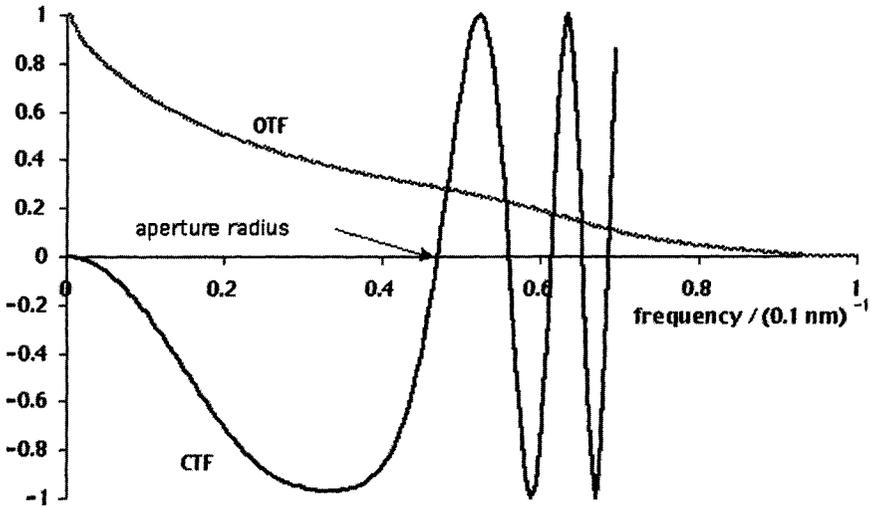


FIGURE 5. The optical transfer function (OTF) for incoherent imaging compared to the phase contrast transfer function (CTF) for the same defocus and spherical aberration. The radius of the objective aperture used for the OTF is marked.

ence between the scattering from spatially separated regions of the probe cannot occur and Equation (9) holds no matter what the detector geometry is. Treacy and Gibson [1995] have noted how a small source will give rise to an incoherent imaging model, but suggested that for more delocalized sources the model failed. However, for delocalized sources it is the detector geometry that can impose incoherence. If $D(\mathbf{K}_f)$ allows the domain of integration of \mathbf{K}_f to be much greater than the domain of \mathbf{K}_i , then any variation of Equation (11) as a function of \mathbf{K}_i will become negligibly small, vanishing completely if $D(\mathbf{K}_f)$ is unity everywhere [Ade, 1977], which means that all transmitted electrons are detected. For a finite detector we are assuming that the detector is summing over complete overlap regions, an exact approximation in Figure 3, and that the contribution of any overlap region intersected by the edge of the detector is negligibly small compared to the rest of the detected signal.

This, then, is the key to forming the incoherent model. For a given spatial frequency, \mathbf{Q} , there are a range of pairs of incident partial plane waves that can contribute. For the incoherent model to apply, all of the available pairs must contribute to the image in a similar way so that a simple integral can be performed over these pairs as in Equation (9). To achieve this, a detector with a geometry much larger than a diffracted disc must be used. This

criterion applies to both the detector and its inner hole. Hartel *et al.* [1996] have suggested that an inner radius at least three times the beam convergence angle should be used. Just as Lord Rayleigh proposed that a large source led to incoherent imaging in a conventional microscope, we have shown how a large detector leads to incoherent imaging in a STEM.

C. The Resolution Limit

One striking feature of Figure 5 is that the resolution limit for incoherent imaging appears to be twice that for the point-resolution in CTEM imaging. To understand the origin of this doubling of resolution, it is worth comparing incoherent imaging, with a large ADF detector, with conventional bright-field imaging that would be performed in HRTEM. Using plane-wave illumination in HRTEM is equivalent by reciprocity to only detecting the intensity at $\mathbf{K}_f = \mathbf{0}$ in a STEM. Under these conditions, Equation (8) is no longer separable into two integrals, so the image intensity cannot be described by a convolution between an object function and a PSF as in the incoherent case. Substituting $\mathbf{K}_f = \mathbf{0}$ into Equation (8) and inverse Fourier transforming gives the coherent imaging model in Equation (1).

There are two important differences between Equations (1) and (10). First, Equation (1) shows that for coherent imaging the phase of the convolution between the probe function and the specimen transmittance, which are both complex quantities, is lost. If one wished to deconvolve the effect of the lens from the image, the phase problem would have to be solved first using holography [Orchowski *et al.*, 1995; Lichte, 1991] or focal-series reconstruction [Coene *et al.*, 1992; Van Dyck and Coene, 1987]. Incoherent imaging does not suffer from a phase problem, and in principle the PSF can be deconvolved directly from the incoherent image described by Equation (10). Second, as seen in Figure 5, the resolution of the ADF is double that for the HRTEM image for the same objective lens and aperture. The origins of this effect can be seen by considering the coherent CBED pattern formed for two crystalline specimens with different lattice spacings (Figure 6). Interference can occur in the overlap regions between the discs, which depends on the phase difference between the diffracted beams, the lens aberrations, and the probe position [Nellist *et al.*, 1995]. All STEM image contrast comes from such overlap regions. Conventional HRTEM imaging requires interference between the zero-order disc and two diffracted discs, and therefore the spacing in Figure 6(b) would not be resolved. However, an ADF detector will detect the interference regions for this crystal, and the spacing will be resolved in the ADF image. Here we

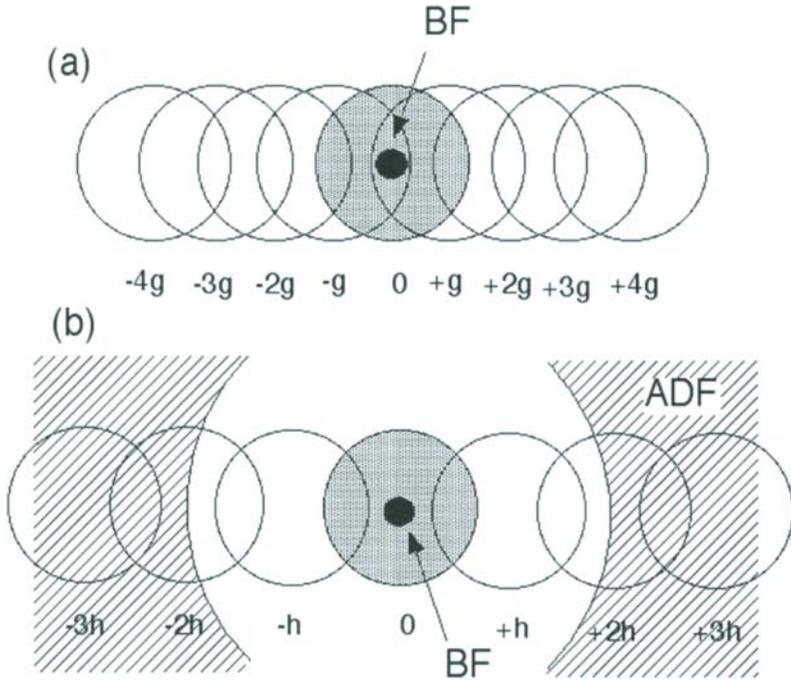


FIGURE 6. A schematic geometry of the STEM detectors relative to the diffracted discs. For crystal (a) the BF detector senses interference between the 0 , g , and $-g$ discs. A crystal with a smaller lattice spacing (b) gives discs more widely space, such that $h > g$. The BF detector senses no interference, whereas the ADF detector senses single-overlap interference.

have essentially restated, via reciprocity, the conclusion of Lord Rayleigh [1896] that the resolution is doubled for incoherent imaging. In real-space we can understand this effect in terms of the probe intensity profile (Figure 4) having approximately half the full width at half maximum of the probe amplitude profile because of the modulus squared being taken, thus improving the resolving power.

D. The Thin Specimen Object Function

Using the thin-object approximation, we now consider the form of the object function, $O(\mathbf{R})$, that applies for incoherent imaging. For illustration we first discuss an amplitude object, and then go on to the slightly more complicated case of a phase object.

1. An Amplitude Object

The theory we are developing is also highly applicable to energy-filtered imaging in a STEM, and to examine the effect of the collector aperture. For a general review of energy filtered imaging see Kohl and Rose [1985]. Let us consider a core excitation of a single atom. Our specimen function, $\psi(\mathbf{R})$, will now have a magnitude that is localized in space. In principle we can collect all the electrons scattered by this excitation by using an infinite collector aperture. In this case $D(\mathbf{K}_f)$ is unity everywhere, and Equation (11) becomes the autocorrelation of $\Psi(\mathbf{K})$,

$$\tilde{O}(\mathbf{Q}) = \int \Psi(\mathbf{K}_f + \mathbf{Q}/2)\Psi^*(\mathbf{K}_f - \mathbf{Q}/2) d\mathbf{K}_f \quad (12)$$

which transformed into real-space gives an object function,

$$O(\mathbf{R}) = |\psi(\mathbf{R})|^2 \quad (13)$$

such that

$$I(\mathbf{R}_o) = |P(\mathbf{R}_o)|^2 \otimes |\psi(\mathbf{R}_o)|^2 \quad (14)$$

identical to Equation (2).

We can see that a large collector aperture gives an image that approaches the perfect incoherent limit, as noted by Kohl and Rose [1985] and Browning and Pennycook [1995]. As the specimen function becomes more localized, as for higher energy excitations, we have seen that the detector geometry becomes irrelevant if the specimen function is much smaller than the probe dimensions. The effect of the collector aperture is not a simple convolution of the object function with the Fourier transform of the collector aperture. The transition from coherent to incoherent imaging as a function of collector angle, and the relative independence of localized excitations to the detector geometry, is all illustrated in Kohl and Rose [1985].

2. A Phase Object

Let us now turn to the question of elastic scattering, which is the important process for Z-contrast imaging. Ignoring the effects of the electron wavefront propagation within the crystal, that is assuming a thin object, the specimen function can be approximated as a pure phase object (see, for example, Cowley [1992]),

$$\psi(\mathbf{R}) = \exp[-i\sigma V_p(\mathbf{R})] \quad (15)$$

where $V_p(\mathbf{R})$ is the projected potential integrated through the thickness of the crystal. The object function for perfect incoherence with an infinite

detector can be formed as it was for inelastic scattering by using Equation (13) which, since we have a pure phase specimen, gives unity everywhere and no contrast. This result is not surprising since no electrons are lost for a phase object, and therefore all the incident electrons will be detected.

To form some contrast into the image, we need to introduce a hole in the detector, hence the use of an annulus in ADF imaging. The effects of the hole in the detector have been considered in the case of a weak-phase object [Jesson and Pennycook, 1993], but here we will generalize to a phase object which, while ignoring the effects of propagation, does include multiple scattering. Ignoring disc-overlap regions intersected by the inner radius of the detector, the object function can be written in reciprocal space,

$$\tilde{O}_{\text{ADF}}(\mathbf{Q}) = \int D_{\text{ADF}}(\mathbf{K}_f) \Psi(\mathbf{K}_f + \mathbf{Q}/2) \Psi^*(\mathbf{K}_f - \mathbf{Q}/2) d\mathbf{K}_f \quad (16)$$

Taking the Fourier transform of Equation (16) and a little algebra gives the real-space object function

$$O_{\text{ADF}}(\mathbf{R}) = \int \left(\delta(\mathbf{B}) - \frac{J_1(2\pi u_i |\mathbf{B}|)}{2\pi |\mathbf{B}|} \right) \psi(\mathbf{R} + \mathbf{B}/2) \psi^*(\mathbf{R} - \mathbf{B}/2) d\mathbf{B} \quad (17)$$

where the term in parenthesis is the Fourier transform of an infinite detector minus a hole with radius, u_i , J_1 is a Bessel function of the first kind, and \mathbf{B} is a dummy real-space vector. Substituting the phase-object approximation, Equation (15), into Equation (17) and writing the integral over a half plane by symmetry gives

$$O_{\text{ADF}}(\mathbf{R}) = 1 - \int_{\text{half plane}} \frac{J_1(2\pi u_i |\mathbf{B}|)}{2\pi |\mathbf{B}|} \cos\{\sigma V_p(\mathbf{R} + \mathbf{B}/2) - \sigma V_p(\mathbf{R} - \mathbf{B}/2)\} d\mathbf{B} \quad (18)$$

Equation (18) shows that the hole in the detector has given rise to a coherence envelope given by the Airy function that is the Fourier transform of the hole. Within this envelope, the electron wavefunction scattered by spatially separated parts of the specimen can interfere, illustrated by the integral over the cosine function in Equation (18). If the projected potential, $V_p(\mathbf{R})$, varies within the coherence envelope, then the cosine function will deviate from 1, and contrast will be seen. Although a finite-sized coherence envelope is required to form contrast, we are at liberty to make it as small as we desire, at the expense of reduced signal. For atomic resolution zone-axis imaging, we select an inner radius that results in a coherence envelope much narrower than the distance between neighboring atomic columns (Figure 7). Each column will then be acting as an independent scattering center, incoherent with its neighbors.

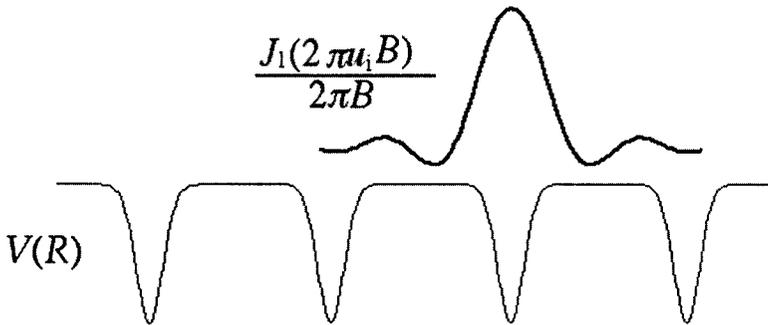


FIGURE 7. A schematic diagram showing the coherence envelope resulting from the hole in the detector and the projected potential. Ideally, the coherence envelope should be smaller than the interatomic distance, but to get contrast the potential must vary within the coherence envelope.

Equation (18) also shows that the object function is sensitive to the rate of change of the projected potential in the vicinity of \mathbf{R} . The potential varies most quickly at the center of the atomic column giving rise to an object function peaked at the atomic column sites. The rate of change is strongly dependent on the nuclear charge, giving a sensitivity to atomic number, hence the name *Z*-contrast.

If we assume that the specimen is thin and the phase shift small, then we can expand Equation (18) to second order in $V_p(\mathbf{R})$,

$$O_{\text{ADF}}(\mathbf{R}) = \int \frac{J_1(2\pi u_i |\mathbf{B}|)}{2\pi |\mathbf{B}|} \{ \sigma V_p(\mathbf{R} + \mathbf{B}/2) - \sigma V_p(\mathbf{R} - \mathbf{B}/2) \}^2 d\mathbf{B} \quad (19)$$

cancelling the constants, which is the result derived by Jesson and Pennycook [1993]. An important difference between weak-phase contrast imaging in a CTEM and *Z*-contrast imaging of weak-phase specimens is that the lowest-order terms in *Z*-contrast imaging go as V_p squared, so it cannot be referred to as linear imaging. Linear imaging requires interference between scattered beams and the unscattered wave, whereas an ADF detector always detects interference between two scattered waves. Of course this effect does not hinder interpretability in any way, but can lead to effects such as an (002) spot present in the Fourier transforms of images of $\text{Si}\langle 110 \rangle$ [Hillyard and Silcox, 1995]. Although the projected potential of $\text{Si}\langle 110 \rangle$ does not contain an (002) Fourier component, the square of the projected potential will unless the potential is exactly a set of infinitely narrow δ -functions.

E. Dynamical Scattering Using Bloch Waves

So far we have considered the destruction of transverse coherence for thin specimens, which do allow the inclusion of multiple scattering effects, but not propagation through the specimen. We must now consider whether we can use an incoherent imaging model in the presence of such propagation. To examine the effects of the detector alone, we will not yet include the effects of coherence loss due to inelastic scattering, for example, phonon scattering. We therefore need to use a theory of dynamical scattering with no absorption effects. We will use the approach of Nellist and Pennycook [1999], and assume a perfectly periodic specimen, so that we have strong Bragg diffracted beams incident on the detector.

Ignoring the effects of higher-order Laue zone (HOLZ) scattering, which is found to have a negligible coherent contribution to the ADF detected intensity [Pennycook and Jesson, 1991; Amali and Rez, 1997], we can expand the solution to the full three-dimensional Schrödinger equation, $\psi(\mathbf{R}, z)$, where z is the thickness, in terms of eigenfunctions of the two-dimensional Hamiltonian (see, for example, Humphreys and Bithell [1992]

$$\psi(\mathbf{R}, z) = \sum_j \Phi_0^{(j)*} \phi(\mathbf{R}) \exp\left(-i\pi \frac{E^{(j)} z}{E_0 \lambda}\right) \quad (20)$$

where

$$\left(\frac{-\hbar^2 \nabla^2}{2m} - eV(\mathbf{R})\right) \phi^{(j)}(\mathbf{R}) = E^{(j)} \phi^{(j)}(\mathbf{R}) \quad (21)$$

and $V(\mathbf{R})$ is the potential averaged along the beam direction. Since $\phi(\mathbf{R})$ must have the periodicity of the crystal lattice, it can be expressed in terms of Fourier components, $\Phi_g^{(j)}$. By assuming plane-wave illumination, the boundary conditions give $\Phi_0^{(j)}$ as the coefficient of excitation of the j th Bloch wave in Equation (20).

Using the Bloch wave formulation of dynamical scattering, we can write the ADF image intensity as

$$I(\mathbf{R}_0, z) = \int D_{\text{ADF}}(\mathbf{K}_f) \left| \int A(\mathbf{K}_i) \exp[i2\pi \mathbf{K}_i \cdot \mathbf{R}_0] \sum_{j\mathbf{g}} \Phi_0^{(j)*}(\mathbf{K}_i) \Phi_g^{(j)}(\mathbf{K}_i) \times \exp\left[-i\pi \frac{zE^{(j)}(\mathbf{K}_i)}{\lambda E_0}\right] \delta(\mathbf{K}_f - \mathbf{K}_i - \mathbf{g}) d\mathbf{K}_i \right|^2 d\mathbf{K}_f \quad (22)$$

where we connected \mathbf{K}_f and \mathbf{K}_i via Bragg scattering through a reciprocal lattice vector \mathbf{g} , the strength of which is given by the usual terms. Expanding the square in Equation (22) gives a double integral similar in form to Equation (6), but can be reduced to a single integral again by taking the

Fourier transform with respect to \mathbf{R}_0 to give an entirely reciprocal space formulation,

$$\begin{aligned} \tilde{I}(\mathbf{Q}, z) = \sum_{\mathbf{g}} D_{\mathbf{g}}^{\text{ADF}} \int A(\mathbf{K}_i) A(\mathbf{K}_i + \mathbf{Q}) \sum_{jk} \Phi_0^{(j)*}(\mathbf{K}_i) \Phi_Q^{(k)}(\mathbf{K}_i) \Phi_{\mathbf{g}}^{(j)}(\mathbf{K}_i) \Phi_{\mathbf{g}}^{(k)*}(\mathbf{K}_i) \\ \times \exp \left[-i\pi \frac{z(E^{(j)}(\mathbf{K}_i) - E^{(k)}(\mathbf{K}_i))}{\lambda E_0} \right] d\mathbf{K}_i \end{aligned} \quad (23)$$

where we have now written the detector function in terms of the discrete Bragg beams that are detected, ignoring partially detected overlap regions.

The advantage of using a completely reciprocal-space formulation is again clear. We can choose to evaluate the summation over the Bragg beams, \mathbf{g} , first to examine the effect of the detector. The \mathbf{g} summation acts only on a product pair of Bloch wave Fourier components resulting in

$$C_{jk}(\mathbf{K}_i) = \sum_{\mathbf{g}} D_{\mathbf{g}}^{\text{ADF}} \Phi_{\mathbf{g}}^{(j)}(\mathbf{K}_i) \Phi_{\mathbf{g}}^{(k)*}(\mathbf{K}_i) \quad (24)$$

so that Equation (23) can now be written

$$\begin{aligned} \tilde{I}(\mathbf{Q}, z) = \int A(\mathbf{K}_i) A(\mathbf{K}_i + \mathbf{Q}) \sum_{jk} C_{jk}(\mathbf{K}_i) \Phi_0^{(j)*}(\mathbf{K}_i) \Phi_Q^{(k)}(\mathbf{K}_i) \\ \times \exp \left[-i\pi \frac{z(E^{(j)}(\mathbf{K}_i) - E^{(k)}(\mathbf{K}_i))}{\lambda E_0} \right] d\mathbf{K}_i \end{aligned} \quad (25)$$

We can choose to evaluate Equation (24) using the real-space representations of the Bloch waves, and assuming the detector to be infinite except for a circular hole gives,

$$C_{jk}(\mathbf{K}_i) = \delta_{jk} - \int \frac{J_1(2\pi u_i |\mathbf{B}|)}{2\pi |\mathbf{B}|} \int \phi^{(j)}(\mathbf{K}_i, \mathbf{C}) \phi^{(k)*}(\mathbf{K}_i, \mathbf{C} + \mathbf{B}) d\mathbf{C} d\mathbf{B} \quad (26)$$

The multiplier containing the Bessel function is the Fourier transform of the hole in the detector, and is acting a kind of coherence envelope, controlling the degree to which different Bloch waves can interfere with each other. By introducing the detector function at this point we can determine which pairs of Bloch waves can interfere in the image forming process, simplifying any calculation by reducing the number of cross-terms that need to be considered rather than including them all and summing over the intensity incident on the detector, as has been used in multislice approaches [Kirkland *et al.*, 1987; Hartel *et al.*, 1996].

In examining the contributions of the various j, k terms to the imaging process, we will refer to the $j = k$ terms as the “diagonal” terms, and the $j \neq k$ terms as “off-diagonal.” At zero thickness, when the exponential

function in Equation (25) can be ignored, the orthogonality of the Bloch states means that the off-diagonal terms give an equal and opposite image to the diagonal terms, resulting from the fact that zero intensity will be scattered to the detector for a zero thickness specimen. As the thickness increases, the integration over \mathbf{K}_i in Equation (25) starts to weaken the off-diagonal terms due to the exponential function and the dispersion of the eigenenergy of the states, and so an image starts to appear. At the limit of infinite thickness, any dispersion of the Bloch states will force the off-diagonal terms to zero, resulting in an image formed purely from the thickness independent diagonal terms, which we will refer to as the residual object function ROF. In this purely elastic and perfectly coherent description of scattering, the ROF is given by the sum of all Bloch waves weighted by their excitation coefficient and C_{jj} . In practice we arrange to use an inner radius such that the coherence envelope in Equation (26) is several times smaller than the smallest intercolumn distance that we wish to image. If the real-space form of the Bloch wave is slowly varying with respect to this coherence envelope, then the integral over \mathbf{C} will approximate closely to being the inner product of $\phi^{(j)}$ with itself, which subtracted from the Kronecker δ -function will give a small value. Therefore, only Bloch waves that are peaked on a scale much smaller than the intercolumn distance will contribute significantly to the ADF images. A calculation of the imaging of InAs<110> at 300 kV using 311 Bloch waves (Figure 8) confirms this, showing that the ROF is dominated by the 1s-type states, localized on the In and As columns, even though for the In columns the 2s state is the most excited.

We have seen that a narrow coherence envelope exists controlling the Bloch wave interference, and this might suggest that an incoherent imaging model can apply here. The narrow coherence envelope ensures that only highly localized Bloch states can contribute to the ROF, and localized states are by their nature relatively nondispersive. Figure 9 shows how the ROF is largely independent of \mathbf{K}_i , which again allows the integral over \mathbf{K}_i in Equation (23) to be performed over just the aperture functions, as in Equation (9). Transforming back to real-space then gives the incoherent imaging model of a convolution between the probe intensity and the ROF.

In addition, Figure 9 shows that the ROF displays strong chemical sensitivity, with the higher atomic-number column being the most intense. Both the insensitivity to \mathbf{K}_i and the chemical sensitivity can be explained physically by considering the high-angle limit of the inner radius. Nellist and Pennycook [1999] have shown that in this limit the diagonal C_{jj} terms become proportional to the transverse kinetic energy of the state,

$$C_{jj}(\mathbf{K}_i) \propto \langle \phi^{(j)}(\mathbf{K}_i) | T | \phi^{(j)}(\mathbf{K}_i) \rangle \quad (27)$$

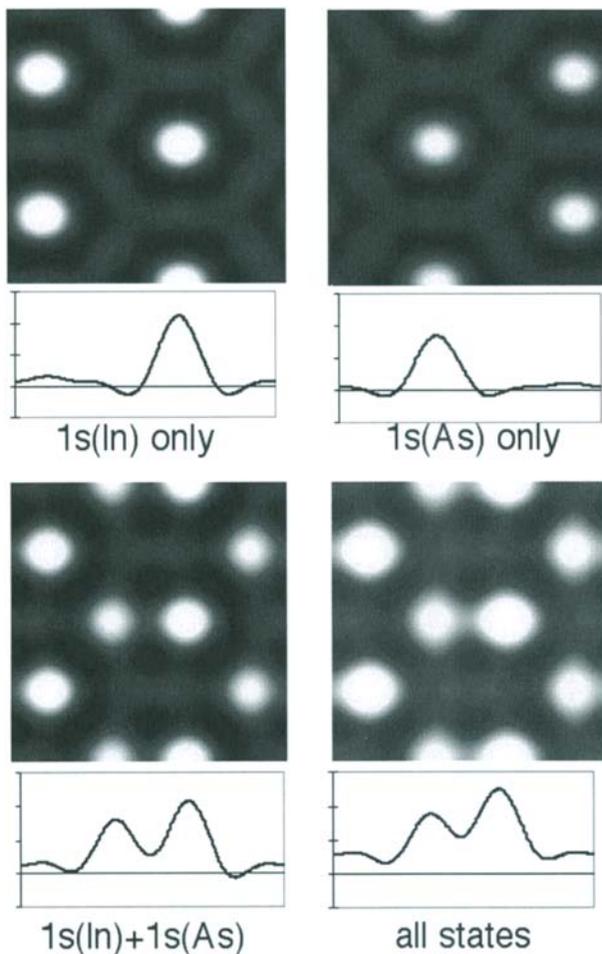


FIGURE 8. The residual object function (ROF) and its profile for InAs at 300 kV calculated using a 311 beam Bloch state calculation. The Bloch states summed over to form the ROF are shown, and illustrate how the 1s states dominate the ROF. The ADF inner radius is 30 mrad.

The C_{jj} multiplier for the diagonal terms is therefore just the expectation value of the transverse kinetic energy for that particular Bloch state. Since we are using a high-angle detector, we find that only Bloch states with sufficient transverse kinetic energy to scatter to the detector contribute, so the detector is acting as a kinetic energy filter with the localized 1s states

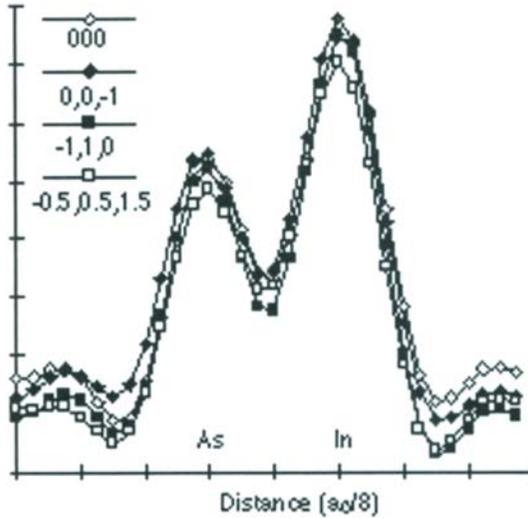


FIGURE 9. The profiles over the dumbbell pairs of residual object function (ROF) for InAs at 300 kV as for Figure 8, but over a range of \mathbf{K}_1 directions as marked. Note how the ROF hardly changes as a function of \mathbf{K}_1 .

dominating. So by forming an object function from the localized $1s$ states weighted by their transverse kinetic energy, we obtain a map of atomic column positions weighted approximately by their atomic number, which we then image incoherently.

An infinite detector with no hole would give perfect incoherence because the integral in Equation (26) would then reduce to δ_{jk} , thus canceling out and destroying any interference between different Bloch waves. However, we mentioned earlier that at zero thickness the scattering by the off-diagonal terms was equal and opposite to the scattering by the diagonal terms, which implies that for an infinite detector, the ROF gives no scattering and, therefore, there is no image contrast. Just as for the phase object in Section II.D above, we need a hole in the detector to introduce some coherence to allow interference to create an observable from the wavefunction phase shifts that occur as the electron passes through the specimen. By increasing the inner radius of the detector we decrease the width of the coherence envelope, and therefore can approach the incoherent limit as closely as we like at the expense of reduced signal. To illustrate the effect of the hole in the detector, consider a situation where both the In and As columns are being illuminated by a large probe. Both the In $1s$ state and the As $1s$ state, denoted here by their subscripts 1 and 2, respectively, will be excited. The

strength of the interference between the two atomic columns can be gauged from the C_{12} term, which we will evaluate here at $\mathbf{K}_i = \mathbf{0}$. Consider first a small axial detector that detects only one Bragg beam, such that only $D_0 = 1$. This is the situation encountered in conventional TEM imaging, since by reciprocity a small STEM detector is equivalent to a low-beam convergence in the TEM illumination. The coherence envelope, being the Fourier transform of the detector, is now relatively large and allows the columns to interfere. The ratio of the magnitudes of $C_{12}(\mathbf{K}_i = \mathbf{0})$ to $C_{11}(\mathbf{K}_i = \mathbf{0})$ is 1.25, and therefore the interference between the In and As columns is significant, varying as a function of thickness, and their contributions must be included in a coherent imaging model. If instead we use an annular dark-field detector with an inner radius of 30 mrad (a typical experimental value), the ratio of the magnitudes of $C_{12}(\mathbf{K}_i = \mathbf{0})$ to $C_{11}(\mathbf{K}_i = \mathbf{0})$ falls to 0.02. Now the coherence envelope is much narrower than the intercolumn spacing, and the interference effects between the two columns is weak. Each column is being imaged as an independent scatterer, and an incoherent imaging model can be used. Thinking of the imaging process as being the detection of interference between overlapping discs in a coherent CBED pattern, intercolumn interference gives rise to features that are smaller than the geometry of the detector, and are therefore averaged out. Only features that are slowly varying over the detector, which come from the contributions from the individual columns, can contribute to the image-forming process.

In this approach we have not considered the question of whether the probe channels down the atomic columns. Propagation of a focused probe through a crystal by a Bloch wave approach has been considered by Fertig and Rose [1981] and Pennycook and Jesson [1990], and more recently by a multislice method by Hillyard *et al.* [1993]. In the approach presented here, we have shown that it is the filtering effect of the *detector* that imposes incoherence. Only the 1s states can contribute to the contrast, and these will only be excited when the probe is located over the corresponding atomic column. Beam spreading and the excitation of other states through the thickness is not important.

III. LONGITUDINAL COHERENCE

We have now discussed in some detail how interference between points that are separated in a direction perpendicular to the electron beam is not detected in ADF imaging, leading to an image that can be described by a convolution between the probe intensity and an object function. It has also

been shown that the object function has a strong dependence on the atomic number, Z , of the species present in a atomic column. We have not yet, however, considered in detail the effects of interference between atoms within the *same* atomic column, referred to as intracolumn interference, which will affect the dependence of the intensity of the peaks in the object function on parameters such as the specimen thickness, or the position of a dopant atom within a column. For example, at the limit of high angle we might conclude that the Rutherford scattering from each atom is proportional to Z^2 . If 10 atoms in a column are being imaged, is the ADF signal then $10 \times Z^2$ (which is the case for perfect incoherence along the column), or $(10 \times Z)^2 = 100 \times Z^2$ (the result if we assume that all the atoms interfere coherently and constructively)? In the latter case we would expect a quadratic dependence on thickness, whereas the former would give a linear dependence. In addition, conservation of energy gives a limit to the total intensity that can be detected, so the intensity cannot rise as a function of thickness forever.

The phase object approximation of Equation (18) shows that at the limit of low thickness (that is low projected potential), the signal will rise quadratically as a function of thickness, as in Equation (19). It is clear that this has perfect longitudinal coherence. As the strength of the projected potential increases, the cosine function gives rise to an oscillatory dependence, which should be observable as thickness fringes. Such fringes are not observed in ADF imaging, so the longitudinal coherence is being broken to some extent. In this section we will first consider the role of the Fresnel propagation within the specimen, using the kinematical approximation. This approach is then extended to include the effect of phonons, and methods of including phonons in image calculations are listed. Finally, we discuss the effects of discontinuities in an atomic column, such as a single dopant atom or a surface adatom.

A. Kinematical Scattering

In Section II we saw how it was an integration in intensity over the transverse component of the scattering vector that destroyed the transverse coherence. Similarly, to destroy coherent interference within a column, known as *intracolumn* interference effects, we need an integration in reciprocal space in a direction parallel to the atomic column. For purely coherent scattering, this integration can only arise from an integration over the longitudinal component of the scattering vector, which comes from the curvature of the Ewald sphere. At high electron energies, however, the Ewald sphere is relatively flat and so the longitudinal components of the

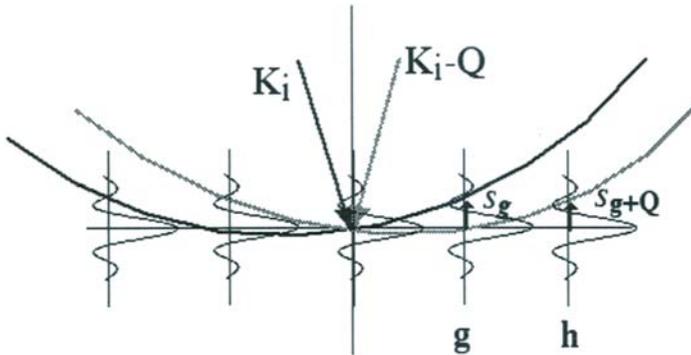


FIGURE 10. A schematic diagram showing the Ewald spheres for two partial plane waves in the incident cone, and the associated scattering amplitude for reciprocal lattice vectors with various excitation errors. The ADF detector sums over the Bragg beams from many such reciprocal lattice vectors.

scattering vector are small (Figure 10), so the detector geometry is much less efficient at breaking the longitudinal coherence than it is at breaking the transverse coherence.

From Figure 10 it can be seen that the ADF detector will sum in intensity over different parts of the shape transform for different Bragg reflections. Kinematical theory (for details see Gevers [1970]) predicts that the amplitude of a reflection follows a sinusoidal dependence on specimen thickness, z ,

$$\Psi_{\mathbf{g}} \propto \frac{V_{\mathbf{g}} \sin(\pi s_{\mathbf{g}} z)}{\pi s_{\mathbf{g}}} \quad (28)$$

and, therefore, the signal detected by the ADF detector as a function of thickness will consist of a sum over many of these sinusoidal oscillations,

$$I_{\text{ADF}}(z) \propto \sum_{\mathbf{g}} D_{\mathbf{g}} |V_{\mathbf{g}}|^2 \frac{\sin^2(\pi s_{\mathbf{g}} z)}{\pi^2 s_{\mathbf{g}}^2} \quad (29)$$

The thickness dependence will therefore depend on the structure and scattering factors of the specimen, and on the detector geometry. In general, the thickness dependence will start by showing an oscillatory behavior that progresses to a constant saturation value as the oscillations in the shape transforms of the reflections are averaged over by the summation over the detector [Treacy and Gibson, 1993]. The intensity oscillations can be

interpreted as being the interference effects between atoms in the same column as the column increases in length, with the decay of these oscillations towards a mean value being the loss of coherence over a longer length column due to the detector geometry. Jesson and Pennycook [1993] point out that the first minimum of such oscillations will be close to the first minimum of the shape transform at a scattering angle given by the inner radius of the detector, which from Equation (29) is given by $z = 2\lambda/\theta_i^2$, assuming that the signal is dominated by scattering to this inner periphery of the detector. Their calculations for 100 kV ADF imaging of Si<110> with an inner radius of 50 mrad show the oscillations decaying over approximately 10 nm.

From Figure 10, if \mathbf{K}_i and $\mathbf{K}_i + \mathbf{Q}$ are diametrically opposite in the cone of incident illumination, then the excitation errors for the interfering scattered Bragg beams, \mathbf{g} and $\mathbf{g} + \mathbf{Q}$, will be equal. As \mathbf{K}_i varies over the incident cone, however, the excitation errors will become different for the two scattered beams, and the intensities of the scattering will vary as a result. Strictly, the integrals in Equation (8) can no longer be separated, and the image intensity can no longer be written as the convolution between a probe intensity and an object function. The explanation for this is that the Fresnel propagation of the probe through the specimen thickness changes the probe profile for scattering from different heights in the specimen, and there is not one single profile that can be applied to all the scattering [Treacy and Gibson, 1995]. We might expect this effect to become more pronounced as the thickness increases, but in that case a full dynamical scattering approach should also be used.

B. Dynamical Scattering

The kinematical approach to the effect of thickness on ADF images has suggested that treating the image intensity as a convolution between the probe intensity and an object function should become less applicable as the thickness increases. Conversely in Section II.E it was observed that at the high-thickness limit, the object function was dominated by the bound 1s states on the columns, and that the low dispersion of these states allowed the integral separation in Equation (25) giving rise to an incoherent imaging model. The multiple scattering gives rise to probe channeling [Fertig and Rose, 1981], which lessens the probe broadening effects.

Having observed that the 1s states dominate the scattering to the ADF detector, it is now instructive to examine the dependence of the dynamical object function given in Section II.E on thickness. Strong oscillations as a

function of thickness are observed [Nellist, to be published], which appear to decay only very slowly and persist over many tens of nanometers. The frequency of these oscillations also depends on the atomic species present, and is controlled mainly by the eigenenergy of the 1s state for that column. The effect is as though standing waves are formed in each column individually, with no crosstalk between the columns. Even though crosstalk may be occurring between adjacent columns, the destruction of intercolumn coherence by the detector prevents there being any effects from crosstalk in the ADF object function. The approach taken by Van Dyck and Op de Beeck [1996] is highly applicable here because of the 1s state domination. The Bloch waves are written as a sum of the 1s states on columns, n , bound to the columns and all the other states are collected as a background,

$$\psi(\mathbf{R}, z) = \sum_n \phi_n^{1s}(\mathbf{R}) \exp\left[\frac{-i\pi z E_n^{1s}}{\lambda E_0}\right] + \sum_{j \neq 1s} \phi_n^{(j)}(\mathbf{R}) \quad (30)$$

When the phase of the 1s state in a given column is close to the other less-bound states, the wavefunction is close to being a plane wave, and there is little scattering to the ADF detector, whereas out of phase, it gives a strong peak in the wavefunction, and therefore in the object function. Whereas the approach of Van Dyck relies on the eigenenergies to separate the 1s states, here it is the ADF detector that is selecting the 1s states, and therefore this approach can be applied over a much greater range of thicknesses.

C. HOLZ Effects

In the kinematical and dynamical analyses given above, the effect of scattering to higher-order Laue zones (HOLZ) has been neglected. In practice, HOLZ scattering does occur at the high scattering angles detected by an ADF detector. Spence *et al.* [1989] pointed out that HOLZ scattering could destroy the incoherent imaging model by the introduction of phase contrast. The reason is that HOLZ scattering gives rise to very fine lines in the CBED pattern rather than discs, and thus only occurs for specific values of \mathbf{K}_\perp . The effect is similar to using a small detector, and the integrals in Equation (8) are most definitely not separable. Measurements by Pennycook and Jesson [1991] and calculations by Amali and Rez [1997] suggest that the contribution to the ADF signal by the HOLZ reflections is small compared to the thermal diffuse scattering (TDS) at those angles. Since TDS is a largely intrinsically incoherent scattering process, it is important to consider how it can affect the coherence in ADF imaging.

D. Thermal Diffuse Scattering

It was suggested by Treacy *et al.* [1978] and Howie [1979] that in order to avoid coherent effects in an ADF image, the inner radius should be taken to an angle where thermal lattice vibrations have attenuated the strength of the coherent scattered Bragg beams, and the scattering is predominantly thermal diffuse scattering (TDS). In practice the situation is more complicated than this. In Section II it was shown that transverse incoherence occurs because of the detector geometry, even when it is detecting Bragg beams, but we might expect TDS to play a role in destroying intracolumn interference. It must be remembered, however, that lattice vibrations can be correlated between nearby atomic sites because the phonon spectrum has larger numbers of long-wavelength phonons, so we cannot assume that simply by detecting TDS, that all coherent effects will be destroyed. The situation is most easily visualized using the kinematical approximation.

1. TDS Using the Kinematical Approximation

The thermal vibrations of the lattice can be described in terms of the normal modes of vibration of the lattice. A quantum of excitation of the lattice is a phonon, so we may have many phonons in each mode. In addition to elastic scattering of the fast electron into the Bragg beams, there may well also be further scattering by a phonon. The phonon scattering can change both the momentum of the electron and its energy (see for further details Cowley [1975]). The latter effect renders the scattered electron incoherent with respect to the elastically scattered electrons. Thus the scattering to a final wavevector, \mathbf{K}_f , can come from a region in reciprocal space given by the distribution of momenta of the scattering phonons (Figure 11). Over this region of scattering vectors, an integration in intensity occurs. Although this makes little difference to the transverse coherence because the detector is much larger than the range of phonon momenta (of the order of one reciprocal lattice), we now have an integration along the beam direction, destroying longitudinal coherence. Jesson and Pennycook [1995] have analyzed in more detail the extent to which the coherence is destroyed, but in general the loss of coherence is not as complete as in the transverse direction.

2. TDS with Multislice Calculations

It is now clear that to fully account for the effects of the partial longitudinal coherence, phonon scattering must be included in the dynamical scattering model. Calculations have been performed using both the multislice method and we will also discuss how a Bloch wave approach may be calculated using matrix diagonalization.

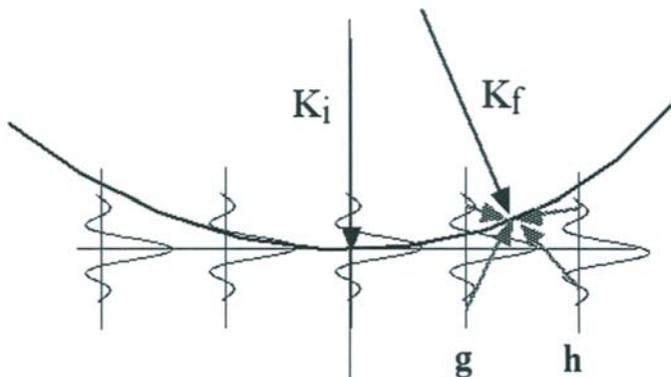


FIGURE 11. The grey arrows depict scattering by phonons in addition to elastic scattering into a final wavevector, K_f . The phonons provide further integration in intensity in both the transverse and longitudinal directions.

The concept of the multislice calculation, proposed initially by Cowley and Moodie [1957] is straightforward. The crystal is divided into a number of thin slices. The projected potential within each finite slice is approximated to lie in an infinitesimal slice, thus acting as a thin-phase object, and the electron wavefunction is propagated between each slice as though it were in vacuum. For STEM, the incident wavefunction at the crystal is the probe's complex amplitude, so the calculation is more complicated than for plane wave illumination because a range of K_i vectors are incident at the specimen. The first use of a multislice calculation for calculating STEM images did not include phonon scattering [Kirkland *et al.*, 1987] and one application was the study of the visibility in ADF of single Au atoms on the surface of thin Si crystals [Loane *et al.*, 1988]. These early simulations suggested that the position of an adatom through the thickness of a crystal could change its resultant effect on the image contrast, complicating the interpretation of intensity.

The multislice approach was then extended to include phonon scattering using the "frozen phonon" approach [Loane *et al.*, 1991]. Since an electron accelerated through 100 kV has a velocity of about half that of light, the time taken for it to pass through a typical TEM sample is of the order of 10^{-16} s, about three orders of magnitude shorter than a typical phonon period of oscillation. The fast electron, therefore, can be regarded as interacting with a frozen "snapshot" of the crystal lattice. The dwell time of the probe at each probe position can be regarded as being long enough that many electrons pass through the sample, each one seeing a lattice with

thermal atomic displacements uncorrelated with the displacements seen by other electrons. The recorded signal is the intensity of the electron scattering summed over many atomic displacement configurations. The accuracy of this approach was first checked by simulating convergent-beam electron diffraction (CBED) patterns that would be observed in the STEM detector plane [Loane *et al.*, 1991], though the patterns simulated did not have overlapping discs so the beam convergence used would not produce a sufficiently small probe for lattice imaging in any STEM imaging modes.

Having shown that the frozen phonon approach does reproduce many of the features observed in the TDS background, ADF STEM images can be simulated by calculating the coherent CBED pattern for each probe position, and summing in intensity over the region detected. Such calculations for atomic-resolution ADF imaging did indeed confirm that the image had transverse incoherence [Loane *et al.*, 1992], and the behavior of the images as a function of thickness could also be studied [Hillyard *et al.*, 1993; Hillyard and Silcox, 1993]. The calculations suggested that the intensity of an atomic column as a function of thickness did not show contrast reversals, such as those observed for coherent TEM imaging, but is not a linear relationship. The strong dependence on the atomic number, Z , was also shown so that heavier elements have more intense columns.

An alternative approach to the inclusion of phonon scattering in a multislice calculation has been developed by Dinges *et al.* [1995]. Instead of displacing the atoms in the calculation, a calculation of the phonon scattering at each slice is made, and this TDS is propagated along with the rest of the electron scattering through the crystal to allow for multiple elastic scattering. However, the TDS is no longer coherent with the elastic scattering because of the inelastic nature of phonon scattering, so no interference effects will occur between the TDS and the elastic scattering. To incorporate this into the calculation, the phonon scattering at each slice is multiplied by a random "statistical phase" shift. The full multislice calculation is then repeated many times with different statistical phases so that any interference effects included by the calculation are eventually averaged over. This method has been applied in a practical way for the atomic resolution analysis of Al concentration in AlGaAs quantum wells [Anderson *et al.*, 1997], where 55 probe positions within a single unit cell were computed for various Al concentrations, and compared with the experiment after substantial noise reduction procedures had been performed. It is worth emphasizing the point made by Hillyard *et al.* [1993] that a resolution limited lattice image consists of just a few complex Fourier components, and so can be completely characterized by just a few numbers. A simulation of a lattice image, therefore, only requires that number of probe positions in judiciously chosen positions to allow the entire image to be simulated.

An approach using simple absorption potentials has been adopted by Nakamura *et al.* [1997] for simulating single Au atoms substituted into a Si lattice. Further elastic scattering of the TDS giving rise to features such as Kikuchi lines is not included in this method, though since this only redistributes the TDS in scattering angle, and a large detector covering many scattering angles is used, it is interesting to consider how important this is. Since most of the absorption is TDS of which much will reach the ADF detector, it is important that this intensity is reintroduced to the final intensity measured in a way consistent with phonon scattering, which would not be peaked at zero scattering angle. Nevertheless, the calculations do suggest that “top-bottom” effects, with the intensity dependent on the Au depth in the crystal, will be observed.

3. Bloch Wave Calculations with TDS

Using the Bloch wave stationary solutions of the Schrödinger equation to compute the electron wavefunction within a crystal has already been discussed and applied in Section II.E, though without any phonon scattering being included. It is possible to include absorption in such a calculation, and to also include an operator for phonon scattering, which operates on the electron wavefunction at all depths in the crystal to compute the TDS [Amali and Rez, 1997]. Such an approach does not include the effects of multiple phonon scattering, but the phonon scattering operator does allow for the creation (or destruction) of multiple phonons in one scattering event, so called “multiphonon” processes. The argument is made that subsequent elastic scattering need not be included if all the Bloch states are included. This argument only holds if all the phonon scattering is detected, which will only be true for a total detector. Physically this is neglecting any redistribution of the TDS from the detector into the hole, which will probably be a small effect for a large detector. Such calculations have been performed for perfect crystals [Amali and Rez, 1997] and stacking faults in crystals [Amali *et al.*, 1997], showing that defects rather than composition changes can also lead to contrast in ADF images.

Pennycook and Jesson [1990] have suggested that if we can assume that the ADF signal has been incoherently generated by phonon scattering at the atom sites, then a model can be used where Bloch waves are used to calculate the electron density at the atom sites, and the intensity at the atom site summed over thickness. It was shown that, for a probe located symmetrically at an atomic column, the 1s states are predominantly excited, which is the origin of channeling. Neglecting the other Bloch states gives rise to a particularly simple thickness dependence.

In Section II.E we saw that the ADF detector acts as a kind of Bloch state

filter by selecting high transverse momentum, spatially localized states, which are the 1s type states for zone axis imaging. Absorption and TDS by phonons also occurs over a highly localized region in the atomic column, and thus is strongest for electron wavefunctions with high transverse momentum, which is the origin of anomalous absorption. Redistribution of intensity by phonons from the Bragg beams to the TDS occurs for high-angle beams, and so selects the 1s states in a similar way to the ADF detector. Note that it is not the phonon momenta that give rise to the high angles of TDS, since phonon momenta are typically of the order of a Brillouin zone, rather that the phonons preferentially scatter high transverse momenta electrons in the crystal. If we make the approximation that all the scattering incident on the ADF detector arises from 1s states, either by coherent scattering or phonon scattering, then we can construct a simple prediction of the thickness behavior of ADF images.

We now make an approximation, used by Van Dyck and Op de Beeck [1996] in the interpretation of conventional HRTEM images, that the eigenenergies of the non-1s states are negligibly small compared to those of the 1s states. The wavefunction in the crystal is approximated as being the 1s states, varying rapidly in phase through the thickness of the crystal, plus the non-1s states all propagating slowly, as described by Equation (30). When the 1s and non-1s states are in phase, as they are at the entrance surface, the wavefunction in the crystal will just be that of the incident wavefunction, which is the STEM probe and does not contain large enough momenta to reach the detector, so no intensity will be recorded. When the 1s state is antiphase to the rest, the wavefunction will be the incident probe minus twice the 1s state, so the electron density will show a large peak at the 1s state and a large ADF signal will be recorded. This oscillatory behavior of the 1s state can be seen in Bloch wave calculations [Fertig and Rose, 1981], and here we give its thickness frequency the symbol, ξ . We now introduce the absorption from the 1s state, with coefficient σ , such that the coherent signal received by the ADF detector varies as

$$I_{\text{coh}}(t) \propto e^{-\sigma t} [1 - \cos(\xi t)] \quad (31)$$

where t is the specimen thickness. The absorption from the 1s state may be integrated over thickness, if we assume it is perfectly incoherent between atoms, to give the total TDS

$$I_{\text{TDS}}(t) \propto (1 - e^{-\sigma t}) - \left(1 + \frac{\sigma^2}{\xi^2}\right)^{-1} \left[\frac{\sigma}{\xi} e^{-\sigma t} \sin(\xi t) + \frac{\sigma^2}{\xi^2} (1 - e^{-\sigma t} \cos(\xi t)) \right] \quad (32)$$

The component of the wavefunction in the 1s state can thus couple to

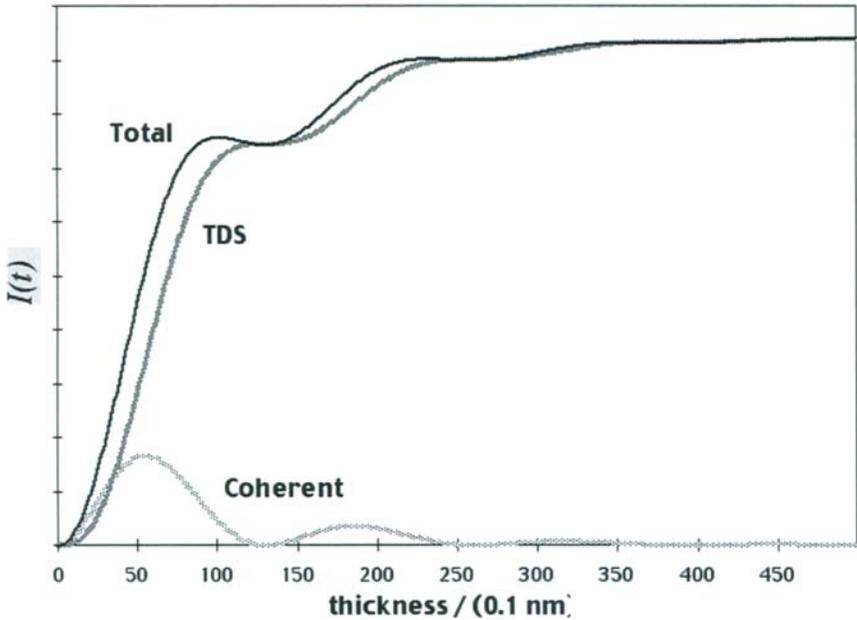


FIGURE 12. A plot of the coherent, TDS and total intensities from Equation (33) using the parameters: $\sigma = 0.12 \text{ nm}^{-1}$, $\xi = 0.48 \text{ nm}^{-1}$, and $\alpha = 0.17$.

high-angle plane-wave states in the vacuum beyond the crystal either at the exit surface, where it emits as coherent Bragg scattering, or through phonon scattering. Let us assume that most of the TDS is detected, as it is biased to high angles, but much of the coherent scattering passes through the hole; the exact fraction depending on the size of the hole. We can define the fraction of coherent to incoherent scatter as α and write the total ADF intensity as

$$I_{\text{ADF}}(t) = I_{\text{TDS}}(t) + \alpha I_{\text{coh}}(t) \quad (33)$$

Equation (33) is therefore a three-parameter model with an analytical expression for the dependence of ADF intensity on thickness. The thickness dependence using Equation (33) with empirical values (Figure 12) compares very closely with the frozen phonon calculation of Hillyard *et al.* [1993] (see Figure 8 of that reference) for a column of In atoms.

Of course, a full simulation is required initially to find the parameters and their dependence on column composition and the experimental conditions. Having parameterized ADF imaging in this way, however, may allow much

faster inversion of intensity data to determine thickness and composition. Probably the most immediate benefit of this type of analysis is to confirm the dominating 1s model, and to show that the image formation mechanism for ADF imaging is becoming understood, and is simple compared to conventional HRTEM imaging. Simplifying the image formation process opens opportunities for quantitative image analysis.

Defects can cause further complications by causing interband transitions within the Bloch states and repopulating 1s states that have been exhausted by absorption. An example of this effect is bright contrast arising from strain around B dopant atoms in Si [Perovic *et al.*, 1993]. It is thus clear that interpreting ADF intensities beyond a qualitative approach still holds many challenges, but nevertheless, the strong Z contrast nature of the imaging is extremely useful.

IV. THE ULTIMATE RESOLUTION AND THE INFORMATION LIMIT

A. Underfocused Microscopy

In Section II.C we have already discussed the resolution limit in terms of the intensity distribution of the illuminating probe, and in reciprocal space the transfer function decreases to zero at a spatial frequency given by twice the objective aperture radius. In conventional HRTEM, however, the form of the contrast transfer function can be modified by changing the degree of defocus, and pass bands can be pushed out in reciprocal space allowing higher spatial frequencies to contribute to the image (see for details Spence [1988]). The ultimate resolution limit of information that can be transferred by the microscope, or *information limit*, is then not defined by the spherical aberration of the objective lens, rather by other sources of incoherence, such as illuminating beam divergence, chromatic aberrations and the overall stability of the microscope. It is now interesting to consider how much of the principles of HRTEM imaging can be applied to ADF imaging. We start by showing that underfocusing the objective lens can counter the effects of spherical aberration leading to a resolution improvement, which then creates the question of what then defines the *information limit* in ADF STEM. Clearly we do not need to consider beam divergence since we have shown by reciprocity that ADF imaging is equivalent to using an extremely high beam divergence to destroy coherence. We must, therefore, go on to look at the effects of chromatic aberrations and the STEM electron source size, which can also destroy coherence. Since we are already describing the imaging as being incoherent, will these further sources of incoherence make any difference?

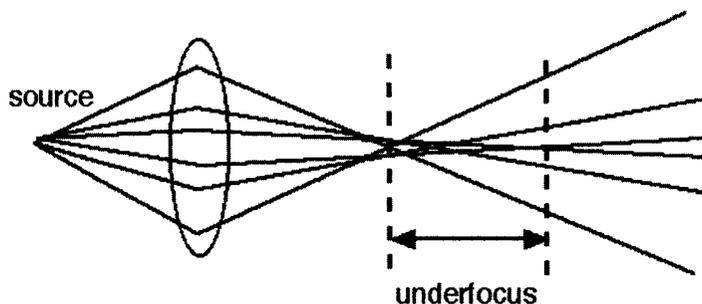


FIGURE 13. Spherical aberration causes overfocusing of the higher angle beams so that they cross over before the gaussian image plane. By underfocusing the lens, the higher-angle beams cross over at the specimen. It is these high angle beams that can carry high-resolution information.

Scherzer [1949] originally showed how, in the presence of spherical aberration, the image resolution could be optimized by underfocusing the objective lens. Geometric ray optics provides a simple way to picture this process (Figure 13). Spherical aberration can be described as being an overfocusing of the beams converging at higher angles, which is countered by an overall underfocusing. In terms of the incoherent OTF (Equation 9) it can be seen that underfocus is used to minimize the variation of the phase of $A(\mathbf{K}_i)$, which is χ , over the integral over \mathbf{K}_i for all values of \mathbf{Q} . Increasing the amount of underfocus will, in general, increase the phase variation and decrease the transfer. Increasing the underfocus, however, has pushed the turning point of χ out to higher values of \mathbf{K}_i , which may be allowed to contribute if the radius of the objective aperture is increased. At certain values of \mathbf{Q} that are beyond the usual resolution limit, the phase variation of A over the disc overlap region in (Equation 9) will become small (Figure 14) and the OTF will show a peak (Figure 15). Although the OTF does not now have the simple form of the Scherzer OTF, it does now extend to much higher spatial frequencies. The corresponding probe intensity distribution does show a smaller central maximum allowing higher resolution information to be passed, but at the expense of lower intensity and long, oscillatory tails (Figure 16). The lack of a phase problem in ADF imaging means that the probe can in principle be readily deconvolved, unlike the case in HRTEM where the phase problem must first be solved using, for instance, holography [Orchowski *et al.*, 1995; Lichte, 1991]. The enhancement in resolution described also can be pictured by geometric ray optics (Figure 13): When the objective lens is highly underfocused, only the high angle beams are crossing at the specimen. These high angle beams are thus

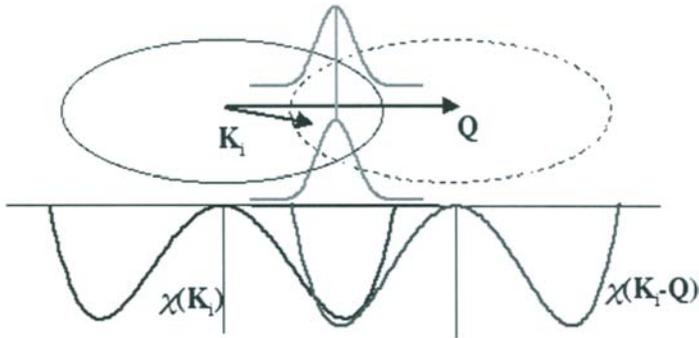


FIGURE 14. For large defocus, the turning point in the quartic form of χ means that there will be some large Q where the integral across the overlapping discs in Equation (9) gives a peak in the transfer function. The attenuation by chromatic defocus spread is shown by the grey gaussian surface.

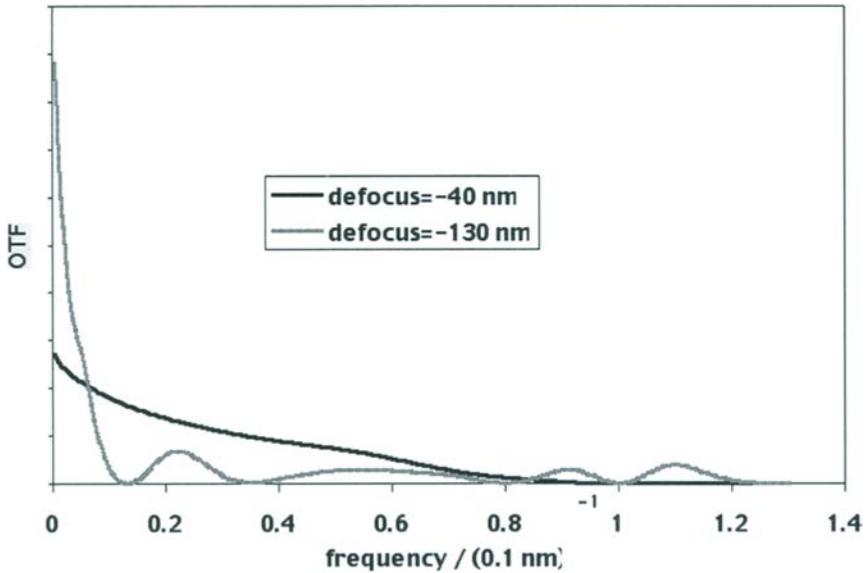


FIGURE 15. The OTF for a large objective aperture (17 mrad) and highly underfocused (-130 nm defocus) compared with that for Scherzer conditions (9.4 mrad; -40 nm). The underfocused OTF shows good transfer at high frequencies. The transfer at the origin is proportional to the area of the objective aperture.

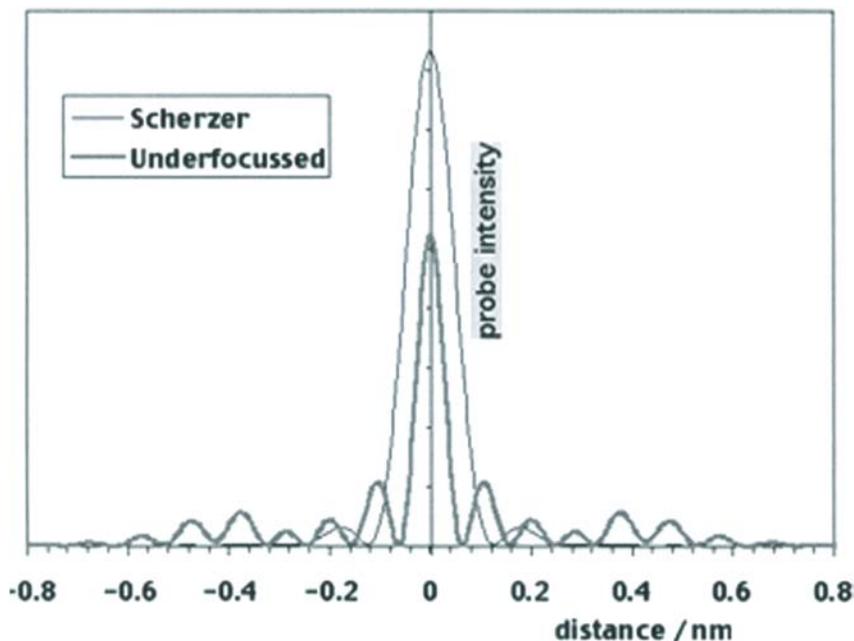


FIGURE 16. Probe intensity profiles for a large objective aperture (17 mrad) and highly underfocussed (-130 nm defocus) compared with that for Scherzer conditions (9.4 mrad; -40 nm). Note how the central maximum has become very much narrower and the first minimum is less than 0.1 nm from the origin.

responsible for the sharp central maximum of the probe, and the out-of-focus lower angle beams give the broader probe tails.

Demonstration of resolution improvement as described has been demonstrated [Nellist and Pennycook, 1998a]. In the $\langle 112 \rangle$ orientation Si has a projected structure of pairs of atomic columns, separated by only 0.078 nm, arranged in a rectangular lattice. An ADF image recorded at Scherzer conditions using the VG Microscopes HB603U STEM (300 kV, $C_s = 1$ mm) is unable to resolve the column pairs, but using an objective aperture with a radius of 17 mrad (approximately twice Scherzer) and a defocus of -130 nm (Scherzer defocus is -44 nm) gives an image where the column pairs are just appearing to be resolved (Figure 17) and the Fourier transform of this image shows that spatial frequencies as far as the $\{444\}$ plane spacing are transferred, which is indeed enough to resolve the column pairs. Despite the probe intensity profile being somewhat complicated, the central maximum of the probe is still the strongest, still allowing direct interpretation to an extent, as shown in the simulation in Figure 18.

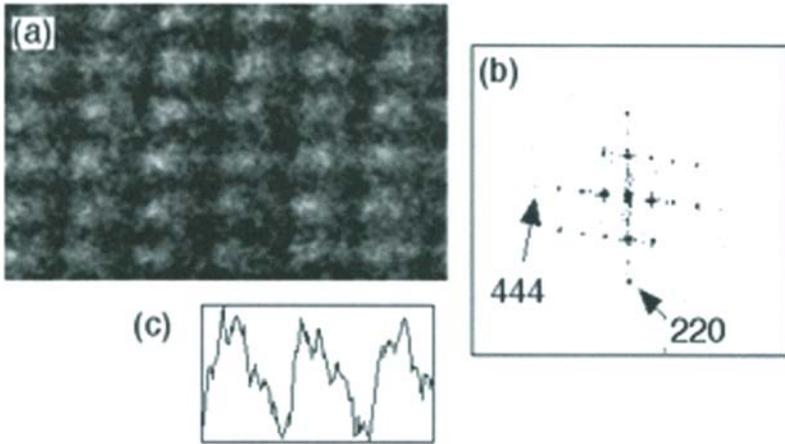


FIGURE 17. (a) An image recorded with an objective aperture of 17 mrad and at approximately -130 nm defocus. The Fourier transform (b) shows transfer to the (444) planes with a spacing of 0.078 nm, resolving the pairs of Si columns, which can just be seen in the image and in profile plot (c) summed perpendicular to the pairs over 150 pixels. The profile plot also shows evidence of the probe sidelobes.

B. Chromatic Aberrations

It is remarkable that resolutions well below an ångström have been achieved in ADF imaging. In HRTEM, information limits below 0.1 nm have been difficult to achieve. The limiting factor is usually the chromatic aberrations of the lens. Electrons of different energy have different focal lengths in the

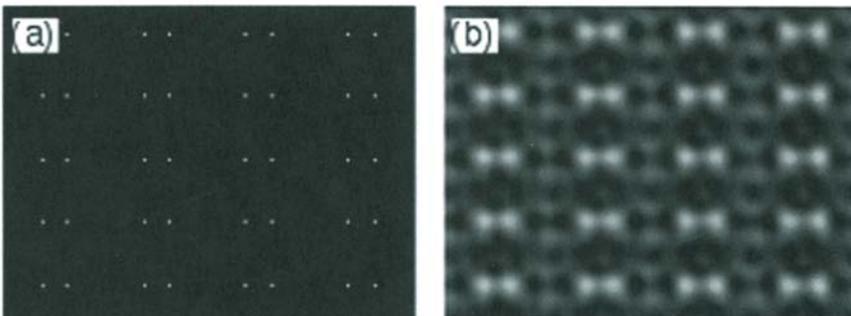


FIGURE 18. (a) The object function for $\text{Si}\langle 112 \rangle$ showing the atomic column positions. The closest spacing between the columns is 0.078 nm. (b) A simulated image by convolving with the probe intensity shown in the underfocused case in Figure 16. Note how the atomic columns are resolved.

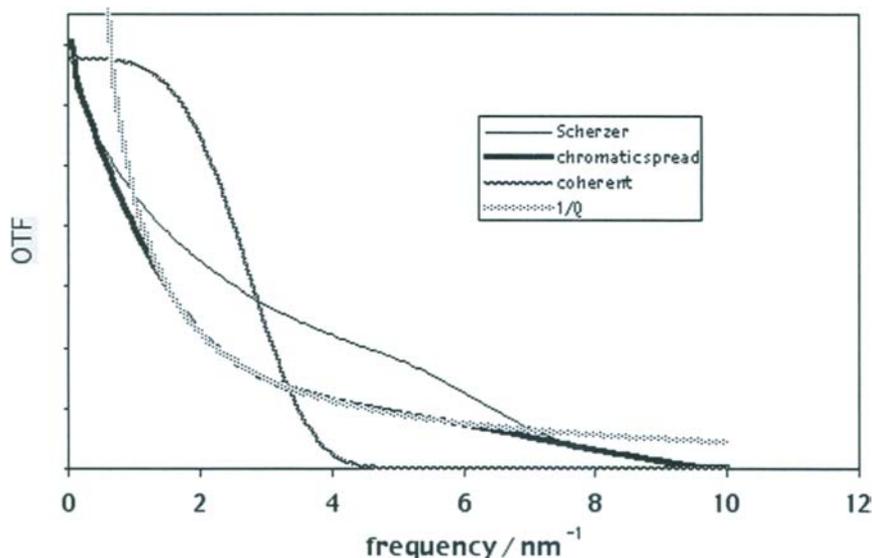


FIGURE 19. The OTF with a defocus spread of 30 nm compared with no defocus spread and the chromatic envelope for HRTEM imaging for the same defocus spread. Note how the effect of chromatic defocus spread is much less severe for incoherent imaging, and just involves being limited in the midrange frequencies by an upper limit proportional to $1/Q$.

objective lens, and the energy spread of the beam leads to an incoherent spread in defocus, Δ , and a coherence envelope in reciprocal space of the form [Wade and Frank, 1977]

$$E_{\text{chr}}(\mathbf{Q}) = \exp\left[-\frac{1}{2}\pi^2\lambda^2\Delta^2|\mathbf{Q}|^4\right] \quad (34)$$

This envelope sharply truncates the transfer of information in HRTEM (Figure 19), and is usually the major factor in controlling the information limit. Using tilted illumination can alleviate the effect giving rise to an “achromatic circle” in the transfer function [Wade, 1976] allowing certain spatial frequencies to be transferred at almost full strength. In this approach, however, only certain spatial frequencies are achromatic, and most are strongly attenuated by the chromatic defocus spread.

Remembering that, from reciprocity, ADF STEM imaging is equivalent to HRTEM imaging with a large incoherent source providing illumination over many tilt angles, we may start to suspect that ADF imaging is robust to chromatic defocus spread since all spatial frequencies will have achromatic contributions. Indeed, it has been observed in calculations by Shao and Crewe [1987] that the intensity of focused electron probes is relatively

insensitive to chromatic defocus spread. A simple quantitative approach to the probe broadening due to chromatic aberrations would be to add the diffraction broadening and chromatic broadening effects in quadrature to calculate the overall probe width. Here we can use the transfer function formulation of Section II.B to use a wave-optical approach to calculating the ADF chromatic envelope.

By considering the transfer function strength calculated in Equation (9) as being an integral over overlapping microdiffracted discs, we can use the approach of Nellist and Rodenburg [1994]. The transfer function must be integrated over a defocus spread that is assumed to be a gaussian distribution

$$T_{\text{chr}}(\mathbf{Q}) = \int \exp[-z'^2/2\Delta^2] \int H(\mathbf{K}_i)H(\mathbf{K}_i - \mathbf{Q}) \\ \times \exp[i\chi(\mathbf{K}_i, z + z') - i\chi(\mathbf{K}_i - \mathbf{Q}, z + z')] d\mathbf{K}_i dz' \quad (35)$$

where χ has been written as a function of defocus, z , explicitly. By substituting Equation (3), and noticing that χ is linear in z , some rearrangement gives

$$T_{\text{chr}}(\mathbf{Q}) = \iint \exp[-z'^2/2\Delta^2] \exp[i\pi z' \lambda (|\mathbf{K}_i|^2 - |\mathbf{K}_i - \mathbf{Q}|^2)] \\ \times H(\mathbf{K}_i)H(\mathbf{K}_i - \mathbf{Q}) \exp[i\chi(\mathbf{K}_i, z) - i\chi(\mathbf{K}_i - \mathbf{Q}, z)] dz' d\mathbf{K}_i \quad (36)$$

The z' integral can be performed first, and is actually the Fourier transform of $\exp[-z'^2/2\Delta^2]$, which gives another gaussian,

$$T_{\text{chr}}(\mathbf{Q}) = \int \exp[-\frac{1}{2}\pi^2 \lambda^2 \Delta^2 (2\mathbf{K}_i \cdot \mathbf{Q} - |\mathbf{Q}|^2)^2] \\ \times H(\mathbf{K}_i)H(\mathbf{K}_i - \mathbf{Q}) \exp[i\chi(\mathbf{K}_i, z) - i\chi(\mathbf{K}_i - \mathbf{Q}, z)] d\mathbf{K}_i \quad (37)$$

after expansion of the moduli-squared and ignoring the overall scaling, which is of no significance here.

From Equation (37) it can be seen that for incoherent imaging, the chromatic defocus spread does not lead to a simple multiplicative coherence envelope scaling the transfer function, as it does for HRTEM. The effect of the chromatic spread depends on the influence of the first exponential function on the \mathbf{K}_i integral. This influence is shown schematically in Figure 14. Along the line given by $\mathbf{K}_i \cdot \mathbf{Q} = |\mathbf{Q}|^2/2$ there is no attenuation. The line is the perpendicular bisector of \mathbf{Q} , and arises from interference between partial plane waves in the convergent beam that have the same angle with respect to the optic axis, and are therefore achromatic with respect to each other. Parallel to \mathbf{Q} , there is an attenuation in the integrand of Equation

(37) following a gaussian form that has a width that is inversely proportional to the defocus spread, Δ , and to the spatial frequency, Q . If the region of disc overlap that is the domain of the integral is narrower than the chromatic gaussian attenuation, then there will be little effect. It is only when the region of overlap, and therefore the transfer function, is large that the chromatic effects will be observed. Thus the effect of chromatic defocus spread is broadly to provide an upper limit to the transfer function. Since the width of the gaussian attenuation is inversely proportional to the spatial frequency, Q , the overall effect of chromatic aberration on the transfer function for incoherent imaging is to provide an upper limit that varies as $1/|Q|$.

The above analysis is illustrated by some calculations. The approximately linear form of the Scherzer OTF (Figure 19) for the VG Microscopes HB603U is limited by the chromatic attenuation for a defocus spread, $\Delta = 30$ nm. Because the transfer function upper limit imposed by the chromatic defocus spread is proportional to $1/|Q|$, only the midrange spatial frequencies are affected. Counterintuitively, the highest spatial frequencies are hardly affected by the defocus spread. The $(0.136 \text{ nm})^{-1}$ spatial frequency (required to resolve the Si<110> dumbbell pairs) is only reduced by a factor of 0.75 for $\Delta = 30$ nm, whereas the attenuation of HRTEM given by Equation (34) for the same defocus spread is 4×10^{-17} , which in practice means that no signal would be observed. This can be understood by realizing that the highest spatial frequencies in ADF arise from partial plane waves that must be close to being the objective aperture *diameter* apart, and are therefore constrained to be almost achromatic with respect to each other. To achieve an attenuation of 0.75 at $(0.136 \text{ nm})^{-1}$ in coherent imaging requires $\Delta = 2.3$ nm, which is much harder to achieve experimentally. For the underfocused case illustrated in Figure 15, it is found that a defocus spread of $\Delta = 10$ nm still allows subångstrom information to be passed at a reasonable strength [Nellist and Pennycook, 1998a], the $1/|Q|$ upper limit being a much more slowly decaying function of Q than the sharp truncation of Equation (34).

The robustness of ADF imaging to chromatic defocus spread has been explained above in terms of interference between partial plane waves in the convergent beam that are largely achromatic with respect to each other. It is also possible to explain it qualitatively in terms of the illuminating probe, since conventional phase contrast HRTEM images can be formed in a STEM by using a small detector. Coherent phase contrast images depend strongly on the phase variation across the coherent illuminating probe to provide the phase contrast, whereas incoherent images, such as ADF images, are insensitive to the phase and just depend on the probe intensity. The phase of the probe is a very sensitive function of defocus, as evidenced

by the rapid contrast reversals that can be observed as the focus is changed, therefore integrating over a defocus spread is able to destroy much of the contrast in a phase contrast image. The intensity distribution in a probe varies more slowly than the phase, and thus the incoherent ADF image is much more robust to the integral over the defocus spread.

C. Source Size and the Ultimate Resolution

It is clear from Section B above that incoherent imaging using ADF STEM is remarkably robust to chromatic aberrations, so we must consider other limitations to the resolution. We can regard the electron optics in a STEM as basically being present to demagnify the electron source such that it is imaged onto the specimen at atomic dimensions, so this demagnification and the effective source size is also important. The electron source can be regarded as being an ensemble of incoherent emitters distributed in space. Each emission point, after demagnification by the STEM electron optics, will give rise to a diffraction limited illuminating probe. The total illuminating probe intensity, J , must therefore be calculated by integrating over this ensemble, after taking into account the demagnification,

$$J(\mathbf{R}) = |P(\mathbf{R})|^2 \otimes S(\mathbf{R}) \quad (38)$$

where S is the geometric image of the source after taking into account the demagnification. Taking the Fourier transform of Equation (38) gives the OTF with a finite source,

$$T_{\text{sou}}(\mathbf{Q}) = T(\mathbf{Q})\tilde{S}(\mathbf{Q}) \quad (39)$$

where \tilde{S} is the Fourier transform of the geometric image of the source. Now we can see that the finite source gives rise to a multiplicative coherence envelope on the transfer function and limits it. To have strong transfer at the highest spatial frequencies allowed by the transfer function, a geometric image of the electron source that is significantly smaller than the diffraction limit is required.

In addition to the objective lens, a STEM instrument is also fitted with condenser lenses, which allow the demagnification of the source to be continuously varied, so should not we simply use the highest demagnification possible? The answer is that we will have no current in the probe if we use this approach. In an optical system, the quantity known as the brightness is conserved, independent of the demagnification [Born and Wolf, 1980]. The brightness of a source is defined as the total current emitted divided by the emission and divided by the solid angle that the radiation is emitted into. In geometric ray optics, this quantity is conserved

and at the probe we can calculate the total current in the probe as follows: The solid angle in the cone of illumination is held constant because of the objective aperture, whose radius is essentially defined by the spherical aberration. Thus the total current is proportional to the area of the geometric image of the source. As we demagnify to reduce the source size, we lose current. This current is being lost because the condenser lenses are more highly excited leading to the beam being more dispersed in angle, and thus being lost by the fixed objective aperture. It is now clear that an extremely high brightness source is required. Indeed the development of the STEM [Crewe *et al.*, 1968a] required the invention of the cold field-emission gun (FEG) [Crewe *et al.*, 1968b] to provide the required brightness. Given an electron gun of certain brightness, the operator has then complete freedom to trade geometric source size against probe current to optimize their experiment and the spatial resolution they require. It can be seen in Section IV.B above, however, that a 300 kV FEG has sufficient brightness for subångstrom imaging to be achieved. The final limits to resolution that are likely to be encountered are those arising from mechanical and electrical instabilities, which, as in any high-resolution microscope installation, should be minimized as much as possible.

Finally, we turn to the prospects for resolution improvement using spherical aberration corrector systems that are currently being developed [Haider *et al.*, 1998; Krivanek *et al.*, 1997]. If such a corrector corrects spherical aberration but not chromatic aberration, then it is clear from the analysis presented in Section B that it will have greatest application to incoherent imaging. As the spherical aberration is corrected, and the diffraction broadening of the probe is reduced, it will be necessary to demagnify the electron source further. It is important to note, however, that total probe current will not be lost because the corrected spherical aberration will allow the objective aperture radius to be increased, thereby exactly compensating. The current challenge appears to be the electrical stability required.

V. QUANTITATIVE IMAGE PROCESSING AND ANALYSIS

A. *The Absence of a Phase Problem*

The mathematical formulation of incoherent imaging as being the convolution between an object function and the illuminating probe intensity has an important consequence that the probe intensity may be directly deconvolved from an image without the prior requirement to solve the phase problem. The implications of this on the way the image intensities can be

analyzed in order to quantitatively determine atomic positions is discussed in this section.

First it is necessary to understand the implications of the phase problem for deconvolutions in coherent HRTEM. The intensity in an HRTEM image can be written in the form of Equation (1). Both P and ψ are complex quantities, and by taking the final modulus-squared the overall phase of the convolution is lost. This so-called *phase problem* is important in many areas of diffraction and imaging [Burge, 1976] since it prevents direct inversion to the specimen structure. The probe complex amplitude, P , is only dependent on the microscope parameters and is, therefore, in principle, known. If the phase problem could be solved in Equation (1), then P could be deconvolved leaving the specimen function. Deconvolving P is thereby mathematically compensating for the effects of spherical aberration, and solving the phase problem in order to achieve this was the motivation behind the invention of holography [Gabor, 1948]. More recently, techniques such as off-axis holography [Orchowski *et al.*, 1995; Lichte, 1991] and focal-series reconstruction [Coene *et al.*, 1992; Van Dyck and Coene, 1987] have been applied to this problem. Even having determined, ψ , it is still necessary to carry out relatively long computer simulations of the dynamical electron diffraction within the specimen [Möbus, 1996; Möbus and Dehm, 1996].

In Equation (2) we can see that there is no phase problem. The probe intensity is real and positive, and can be directly deconvolved. Our aim is to result in the ADF object function, which should consist of sharp peaks at the atomic column positions, the positions of which could be directly measured to quantitatively determine the atomic positions. The first step to achieve this is to compute or determine the probe intensity, $P(\mathbf{R})$.

B. Probe Reconstruction

The probe function is purely a function of the microscope, and therefore a completely known parameter. In practice the parameters on which the probe function depends, such as defocus and astigmatism, will vary from image to image as, for example, the height of the specimen within the objective lens changes. In many images, however, the feature of interest is a localized defect, such as a dislocation core, located within a large region of perfect crystal of known structure. We can therefore estimate the object function in that perfect crystalline region and use the image data to determine the probe. A more detailed description of such a scheme is given by Nellist and Pennycook [1998b], but the basic ideas are as follows: The approach is to first determine the OTF for that particular image. The Fourier transform of such an image can be taken and the intensities of the

corresponding spots measured. Making the approximation that the object function consists of δ -function-like peaks at the atomic column positions weighted by the square of the atomic number (Z^2) allows the Fourier components of the object function, or the incoherent structure factors, to be determined. These would be the expected spot strengths for a hypothetical perfect microscope with no resolution limit. Dividing the measured spot strengths by these computed Fourier components allows the microscope transfer function at these spatial frequencies to be determined. James and Browning [1999] have performed this procedure for different focus settings, and find that the determined transfer function values are close to those expected for the estimated focus.

The observation can now be made that an image of a crystalline material does not allow the OTF to be uniquely determined, because only certain spatial frequencies are present. In general, perfectly periodic images contain very little information and can be characterized with only a few parameters. We are therefore to impose some form of constraint to proceed further. We assume that the transfer function is a relatively slowly varying function in reciprocal space, which is equivalent to assuming that we have a well-localized probe in real space. A linear interpolation is therefore taken between the derived transfer values, which has been shown to be a good approximation [Nellist and Pennycook, 1998b]. The estimated transfer function can now be transformed back to reciprocal space to calculate the estimated probe intensity distribution. In the above example of Nellist and Pennycook [1998b] using GaAs<110>, the final estimated probe was found to be marginally broader than the optimum diffraction limited probe, which is to be expected taking into account the residual instabilities of a microscope.

C. Deconvolution Methods

1. Multiplicative Deconvolution

In reconstructing the probe above, what we have done is deconvolved the object function from the image resulting in the probe intensity distribution. By dividing the Fourier transform of the image by the Fourier transform of the object function, we have performed a *multiplicative deconvolution*. A variety of different deconvolution methods exists [Bates and McDonnell, 1986], and we will explore the application of some of them to ADF object function reconstruction here, following the approaches of Nellist and Pennycook [1998b].

Having used a multiplicative deconvolution to reconstruct the probe from an image region containing perfect crystal, perhaps now the obvious place

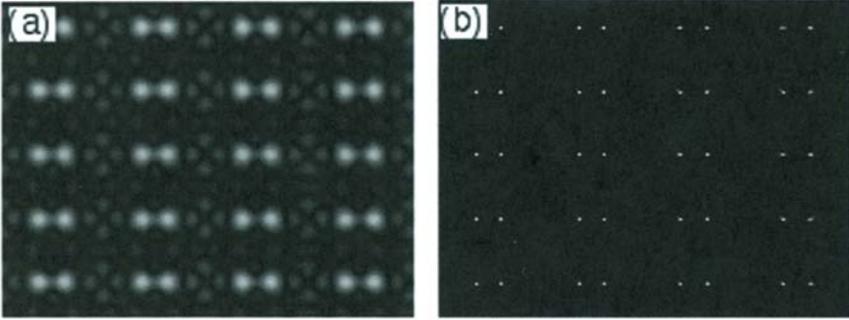


FIGURE 20. The simulated image from Figure 18 with the probe intensity having been deconvolved using: (a) a Wiener filter; (b) the CLEAN algorithm. Both these object functions are completely consistent with the experimental data.

to start is to apply a multiplicative deconvolution to the image area of interest, a defect say. Dividing the Fourier transform of the image by the transfer function will result in the Fourier transform of the object function. The difficulty is that the transfer function goes to zero at the resolution limit, and dividing by close to zero transfer values will dramatically amplify noise. To protect against this, a Wiener filter may be used (for a description see Bates and McDonnell [1986], which in our notation has the form

$$\tilde{o}_{\text{rec}}(\mathbf{Q}) = \frac{\tilde{I}(\mathbf{Q})T^*(\mathbf{Q})}{|T(\mathbf{Q})|^2 + \varepsilon} \quad (40)$$

where \tilde{I} is the Fourier transform of the recorded image, T^* is the complex conjugate of the transfer function, and \tilde{o}_{rec} is the reconstructed object function. A simulated image of Si<112> taken at Scherzer defocus is shown in Figure 18. The ε parameter prevents an attempt to reconstruct the object function where the transfer is very low to minimize the impact of the noise. The result of applying a Wiener filter with ε set to 10^{-3} of the maximum transfer value is shown in Figure 20. It is initially somewhat surprising that deconvolving the probe this way has made little difference to the image, and has not resulted in a sharply peaked object function from which we can read off the atom positions.

The explanation of why the Wiener filter has been ineffectual lies in the form of the transfer functions. The initial OTF is a smoothly decaying function that goes to zero at the resolution limit. The Wiener filter results in a new *effective* transfer function that has a value of 1 until spatial

frequencies where the transfer is weak and ε starts to dominate in Equation (40). At this point the effective transfer function is decayed to zero. The result is an effective transfer function that goes to zero much more sharply than the original OTF. Sharp features in a transfer function lead to a degree of delocalization in the real-space image, and should be avoided. An optimum transfer function is one that decays in a smooth way to zero at the resolution limit, much like the original OTF. It is now clear that it is hard to improve upon the transfer function imposed by the microscope in ADF STEM, which also explains why ADF STEM images are straightforward to process and are often analyzed intuitively from the raw data or after a simple low-pass noise filter.

It now becomes apparent that the obstacle to reconstructing a sharply peaked object function is not the form of the transfer function, rather the fact that information has been lost in the image-forming process. The object function contains information throughout reciprocal space, but above the resolution limit of the transfer function it is all lost. Since this information is thrown away, there can be many object functions that, once convolved with the probe intensity, can give rise to an observed image intensity. Such object functions would differ from each other in the information that exists beyond the resolution limit. To reconstruct to a single object function we need to impose a constraint further to the data we have measured in the microscope, using our prior knowledge and experience.

2. *The CLEAN Algorithm*

At this point it is worth mentioning an alternative approach to deconvolution, known as subtractive deconvolution. In attempting to deconvolve oscillatory point-spread function from astronomical radio telescope images, Högbom [1974] invented an algorithm given the acronym, CLEAN. It has subsequently been applied to Patterson maps calculated from HOLZ electron diffraction patterns [Sleight *et al.*, 1998], and here we examine its applicability to ADF images.

The CLEAN algorithm has the following form:

1. Locate the pixel in the raw data that has the maximum intensity.
2. Transfer a fraction, γ , known as the *loop gain*, of the maximum intensity to the CLEANed reconstructed object function at that pixel location.
3. Subtract a PSF from the raw data, centered at the peak pixel position and with a height of γ times that intensity.
4. Test the raw data to see if the contrast has fallen below some previously specified criterion.

The results of applying the CLEAN algorithm to the image shown in Figure 18 is shown in Figure 20. It is clear that a much sharper object function has been reconstructed than that found using the Wiener filter. A closer inspection of the algorithm, however, shows that there will always be a spatial spread of pixel values reconstructed in the object function, and so single peaks at the atom positions will not be achieved. Furthermore, the sharpening of the object function shows that a constraint has been applied, but it is not clear what the nature of this constraint is. Tan [1986] has analyzed the CLEAN algorithm to determine the constraint, and found the algorithm broadly tends to reconstruct a “sharp” object function, which, of course, is ideal for atomic-like object functions. The mathematical details of the constraint, however, are found to vary strongly as a function of the loop gain, γ , and the stopping parameter. Since these will be varied depending on the noise in the raw data, it cannot be said that the CLEAN algorithm will quantitatively give a consistent result. Having said that, the fast and simple implementation of the CLEAN algorithm, and its reconstruction of peak objects, makes it an attractive proposition, and it may well be worthy of further investigation as a tool for ADF imaging.

3. Bayesian Methods

The attempt in Section 1 above to use a Wiener filter to deconvolve the probe from an image demonstrated how it was not possible to determine a unique object function from the image data alone. The Wiener filter resulted in a blurred object function, which was a consistent object function for the image data, but was not the original one used consisting of an array of delta function. Both object functions are consistent with the data. To select between these valid object functions we need to incorporate some kind of constraint, but how should this be done mathematically?

An approach that has been applied to image processing is to assert that the best we can hope to quantify is a *probability distribution* of the object functions given the experimental data (for example Sivia [1993]), written $p(\text{object}|\text{data})$. The object function would allow us to determine both the most probably object function given the experimental data, but also allow us to quantitatively determine our confidence in that deduction [Skilling, 1998]. Using Bayes theorem, it is possible to relate this distribution to other probability distributions,

$$p(\text{object}|\text{data}) = \frac{p(\text{data}|\text{object})p(\text{object})}{p(\text{data})}$$

For a noisy image, the probability of getting the measured data set for a

given object function, $p(\text{data} | \text{object})$, can be calculated by convolving the object function with the point spread function and performing a chi-squared comparison [Gull, 1978]. The probability of the data, $p(\text{data})$, is a constant, because the measured data is a constant. The probability of the object function, $p(\text{object})$, is the probability of that object function existing, without any other constraints. It is within this probability distribution that we must include our constraints, or prior knowledge about the system we are studying. It is therefore referred to as the *prior* probability distribution. Bayes theorem is making clear to us that there is no such thing as an unambiguous experiment. To interpret any experimental data, whether from an electron microscope or not, always requires prior experiences and knowledge.

For ADF imaging, therefore, we need to find mathematical forms in which to encode our prior knowledge about the specimen. A popular prior in many different fields requiring image processing, for example radio, astronomy [Gull, 1978], has been entropy (for a review of applications see Buck [1991]). The prior has the form,

$$p(\text{object function}) \propto \exp(\alpha S)$$

where

$$S = -\sum_i f_i \log f_i$$

is the entropy, f_i are the pixel values for the object function, α is a constant. The idea is that entropy will always favor a smooth object; the highest entropy object is one that is completely uniform. Any structure in the object will result in a lower prior probability and will be less favored. Use of this prior in a Bayesian reconstruction will result in an object being reconstructed with the minimum structure consistent with the experimental data. An algorithm designed to locate the object function with the highest $p(\text{object} | \text{data})$ has been implemented by Skilling and Bryan [1984] and successfully applied to ADF data (McGibbon *et al.*, 1994; McGibbon *et al.*, 1995).

For precise quantitative applications, however, it has become clear that entropy cannot be used [Skilling, 1998]. A simple explanation is that the result obtained will depend on the number of pixels in the image, that is, on the sampling of the data. Clearly the object function should just depend on the specimen, not on the way that the data has been collected. Quantitatively correct prior probability distributions should fall into a specific class of mathematical functions [Sibisi and Skilling, 1997]. Such priors do exist, and an extremely promising one is known as massive inference [Skilling, 1998], since it emphasizes objects with a small number of highly localized sources, ideal for atomic images.

Clearly the lack of a phase problem in ADF imaging, and freedom from complicating effects such as contrast reversals, gives many opportunities for image processing and analysis. The vast majority of images are still analyzed in a qualitative way, but there is great scope for the development of quantitative methods that will provide investigators quickly and reliably with quantitative measurements of atomic column positions.

VI. CONCLUSIONS

A. Overview

What we have tried to show in this paper is that annular dark-field imaging in a scanning transmission electron microscope is simply the implementation in TEM of incoherent imaging as discussed by Lord Rayleigh [1896]. In general, incoherent imaging is far more common than coherent imaging. The images formed by our eyes are generally incoherent, since we usually have large incoherent light sources, and most optical microscopy is also incoherent. The great advantage to us of using incoherent imaging is that most of the interference effects that can complicate an image are destroyed, leading to much more straightforward interpretation of images. We would struggle to interpret what our eyes were seeing if we used mostly coherent sources, such as lasers, resulting in large fields of interference fringes. It is somewhat surprising that incoherent imaging in TEM has been so long in arriving, and must have much to do with requiring the microscope to form diffraction patterns and diffraction contrast images, both of which require coherent illumination. Such constraints do not apply at high resolution, and we hope we have demonstrated that incoherent imaging holds many advantages.

It must be acknowledged that incoherent image formation has lost information relative to coherent image formation. The use of a large detector geometry has averaged over many of the interference effects observable in the STEM detector plane. The complicated interference fringes seen in coherent imaging do contain information, but it is not easily invertible to a specimen structure. Generally the only way to proceed is to simulate images from trial structures to match against the images, with the danger that the true structure may be missed. The removal of information in the incoherent image formation process simplifies the interpretation of the data, and enough information is retained to determine the object structure. This allows the opportunity to observe unexpected structures, which is of crucial importance since it would be somewhat arrogant to assume that we could predict all the atomic structures that we are ever likely to observe.

The type of information that can be derived from an ADF image falls into two broad classes: the determination of the projected position of atoms and atomic columns, and the determination of the elemental composition of the specimen. The former of these applications relies mainly on the destruction of transverse coherence, which results in the ADF image intensity being a convolution between the probe intensity and an object function with sharp peaks at the atom positions. Here we have shown that transverse coherence is destroyed by the geometry of the detector. Even if we assume purely coherent scattering, a large detector in a STEM is equivalent by reciprocity to a large incoherent source in conventional TEM, resulting in incoherent imaging. The destruction of transverse coherence is almost complete, and opens the way to a whole variety of image processing and analysis methods.

The two-dimensional transverse nature of the ADF detector, while being highly efficient at destroying transverse incoherence, is much less efficient at destroying coherence between atoms along the beam direction. Here we must rely much more on the fact that at high angles we collect a large proportion of electrons that have been scattered by both an elastic scattering event and a phonon. The phase randomization by the phonon scattering can help in destroying the coherence between the scattering from different atoms, but the coherence destruction process is much less complete in this longitudinal direction. This residual coherence can be seen in effects such as strain contrast in the images, and provides to a limit to the degree to which the column intensities in an image can be quantitatively interpreted in terms of composition. The high angle nature of ADF imaging, though, does mean that the scattered intensities are a much stronger function of atomic number than in HRTEM, which relies on much lower angle scattering. This Z-contrast nature of ADF STEM is extremely useful and has found many applications, albeit generally of a qualitative nature.

B. Future Prospects

Although, to date, the number of machines worldwide capable of performing ADF STEM imaging has been rather low, there are already many examples of its application to materials problems ranging from semiconductor interfaces [McGibbon *et al.*, 1995] to supported catalysts [Nellist and Pennycook, 1996]. The main hindrance has been that dedicated STEM instruments have been required, which have been very much research machines requiring dedicated staff. Recently, field-emission gun TEMs (FEGTEMs) have been displaying impressive capabilities as STEM instruments [James and Browning, 1999], thus combining TEM and STEM capabilities within one machine. With the prolific growth in the number of

FEGTEMs being acquired, there will be a major growth in the use of the STEM approach, particularly ADF imaging, especially as the benefits of incoherent imaging are observed first-hand by a growing bunch of users.

Naturally the growth in the number of applications of ADF imaging will promote further developments in the STEM capabilities of machines. Improvements may include better electrical and mechanical stability, improved scanning capabilities, and higher detector efficiencies. Such improvements will be of benefit also in the use of STEM for high spatial resolution microanalysis.

In terms of image interpretation, the future lies in rapid quantitative deductions about the specimen. There is no doubt that qualitative interpretation of the image data will be the first approach used, as it is with most image data, but we have shown here that ADF imaging provides excellent opportunities for quantitative interpretation. Determination of atomic positions from the image data has been achieved, being relatively straightforward information to retrieve, and should quickly become fast and routinely done. Use of the image intensities for quantitative compositional determination is more difficult because of residual coherence, especially in the presence of, for example, strain. To date, attempts to perform quantitative determination have involved long calculations. Such calculations generally retain all the coherent scattering information, which is finally lost at the end when the summation over the detector is done, and therefore there is a lot of wasted time in the calculation. What is required are approaches where the incoherence is built in so that only the relevant terms in the calculation are retained. Further theoretical work is required here. More than one image may also be required for more detailed interpretation. For example, a series of images of the same area could be recorded over a series of different inner radii. This data set should be able to distinguish between the effects of strain and atomic number contrast.

Finally, ADF STEM has already established itself as an important technique for the atomic-resolution investigation of materials, and will act as a catalyst for a much greater interest in all the applications of STEM.

ACKNOWLEDGMENTS

This work was supported by The Royal Society (London) and by the US DOE under Contract No. DE-AC05-96OR22464 with Lockheed Martin Energy Research. Some of the work was performed at Oak Ridge National Laboratory through an appointment to the ORNL Postdoctoral Program administered by ORISE.

REFERENCES

- Ade, G. (1977). *Optik* 49, 113–116. “On the incoherent imaging in the scanning transmission electron microscope.”
- Amali, A. and Rez, P. (1997). *Microsc. Microanal.* 3, 28–46. “Theory of Lattice Resolution in High-angle Annular Dark-field Images.”
- Amali, A., Rez, P. and Cowley, J. M. (1997). *Micron* 28, 89–94. “High Angle Annular Dark Field Imaging of Stacking Faults.”
- Anderson, S. C., Birkeland, C. R., Anstis, G. R. and Cockayne, D. J. H. (1997). *Ultramicroscopy* 69, 83–103. “An approach to quantitative compositional profiling at near-atomic resolution using high-angle annular dark field imaging.”
- Bates, R. H. T. and McDonnell, M. J. (1986). *Image Restoration and Reconstruction*. OUP, Oxford.
- Bethe, H. (1928). *Ann. Phys. Lpz.* 87, 55–129. “Theorie der Beugung von Elektronen an Kristallen.”
- Black, G. and Linfoot, E.H. (1957). *Proc. R. Soc. (Lond.) A* 239, 522–540. “Spherical aberration and the information content of optical images.”
- Born, M. and Wolf, E. (1980). *Principles of Optics*. Pergamon Press, Oxford.
- Bourret, A., Rouviere, J. L. and Penisson, J. M. (1988). *Acta Cryst. A* 44, 838–847. “Structure Determination of Planar Defects in Crystals of Germanium and Molybdenum by HREM.”
- Browning, N. D. and Pennycook, S. J. (1995). *J. Microsc.* 180, 230–237. “Atomic-resolution electron energy-loss spectroscopy in the scanning transmission electron microscope.”
- Buck, B. and Macaulay, V. A. (1991). *Maximum Entropy in Action*. Oxford, OUP.
- Burge, R. E. (1976). *Proc. R. Soc. (London) A* 350, 191–212. “The phase problem.”
- Coene, W., Janssen, G., Op de Beeck, M. and Van Dyck, D. (1992). *Phys. Rev. Lett.* 69, 3743–3746. “Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy.”
- Cowley, J. M. (1969). *Appl. Phys. Lett.* 15, 58–59. “Image contrast in a transmission scanning electron microscope.”
- Cowley, J. M. (1975). *Diffraction Physics*. North Holland, Amsterdam.
- Cowley, J. M. (1992). In *Electron Diffraction Techniques* (J. M. Cowley, ed.). Vol. 1, pp. 1–74. OUP, Oxford. “Electron Diffraction: An Introduction.”
- Cowley, J. M. and Moodie, A. F. (1957). *Acta Cryst.* 10, 609–619. “The scattering of electrons by atoms and crystals I. A new theoretical approach.”
- Crewe, A. V. (1980). *Rep. Prog. Phys.* 43, 621–639. “The physics of the high-resolution STEM.”
- Crewe, A. V., Wall, J. and Welter, L. M. (1968a). *J. Appl. Phys.* 39, 5861–5868. “A high-resolution scanning transmission electron microscope.”
- Crewe, A. V., Eggenberger, D. N., Wall, J. and Welter, L. M. (1968b). *Rev. Sci. Instr.* 39, 576–583. “Electron gun using a field emission source.”
- Crewe, A. V., Wall, J. and Langmore, J. (1970). *Science* 168, 1338–1340. “Visibility of single atoms.”
- Dinges, C., Berger, A. and Rose, H. (1995). *Ultramicroscopy* 60, 49–70. “Simulation of TEM images considering phonon and electronic excitations.”
- Donald, A. M. and Craven, A. J. (1979). *Phil. Mag. A* 39, 1–11. “A study of grain boundary segregation in Cu-Bi alloys using STEM.”
- Fertig, J. and Rose, H. (1981). *Optik* 59, 407–429. “Resolution and contrast of crystalline objects in high-resolution scanning transmission electron microscopy.”
- Gabor, D. (1948). *Nature* 161, 777–778. “A new microscope principle.”
- Gevers, R. (1970). In *Modern Diffraction and Imaging Techniques in Materials Science* (S.

- Amelinckx, G. Gevers, G. Remaut and J. Van Landuyt, eds.), pp. 1–33. North Holland, Amsterdam. “Kinematical Theory of Electron Diffraction.”
- Glaisher, R. W., Spargo, A. E. C. and Smith, D. J. (1989). *Ultramicroscopy* 27, 35–52. “A systematic analysis of HREM imaging of elemental semiconductors.”
- Gull, S. F. and Daniell, G. J. (1978). *Nature* 272, 686–690. “Image reconstruction from incomplete and noisy data.”
- Haider, M., Uhlemann, S., Schwan, E., Rose, H., Kabius, B. and Urban, K. (1998). *Nature* 392, 768–769. “Electron microscopy image enhanced.”
- Hartel, P., Rose, H. and Dinges, C. (1996). *Ultramicroscopy* 63, 93–114. “Conditions and reasons for incoherent imaging in STEM.”
- Hillyard, S. and Silcox, J. (1993). *Ultramicroscopy* 52, 325–334. “Thickness effects in ADF STEM zone axis images.”
- Hillyard, S. and Silcox, J. (1995). *Ultramicroscopy* 58, 6–17. “Detector geometry, thermal diffuse scattering and strain effects in ADF STEM images.”
- Hillyard, S., Loane, R. F. and Silcox, J. (1993). *Ultramicroscopy* 49, 14–25. “Annular dark-field imaging: resolution and thickness effects.”
- Högbom, J. A. (1974). *Astron. Astrophys. Suppl.* 15, 417–426. “Aperture synthesis with a non-regular distribution of interferometer baselines.”
- Howie, A. (1979). *J. Microsc.* 117, 11–23. “Image contrast and localized signal selection techniques.”
- Humphreys, C. J. and Bithell, E. G. (1992). In *Electron Diffraction Techniques* (J. M. Cowley, ed). Vol. 1, pp. 75–151. OUP, Oxford. “Electron Diffraction Theory.”
- James, E. M. and Browning, N. D. (1999). *Ultramicroscopy* 78, 125–139. “Practical aspects of atomic resolution imaging and analysis in STEM.”
- Jesson, D. E. and Pennycook, S. J. (1993). *Proc. R. Soc. (London) A* 441, 261–281. “Incoherent imaging of thin specimens using coherently scattered electrons.”
- Jesson, D. E. and Pennycook, S. J. (1995). *Proc. R. Soc. (London) A* 449, 273–293. “Incoherent imaging of crystals using thermally scattered electrons.”
- Kirkland, E. J., Loane, R. F. and Silcox, J. (1987). *Ultramicroscopy* 23, 77–96. “Simulation of annular dark field STEM images using a modified multislice method.”
- Kohl, H. and Rose, H. (1985). *Adv. Elect. & Electron. Phys.* 65, 173–227. “Theory of image formation by inelastically scattered electrons in the electron microscope.”
- Krivanek, O. L., Dellby, N., Spence, A. J., Camps, R. A. and Brown, L. M. (1997). *Proc. EMAG97, Cambridge, 1997* 35–40. “Aberration correction in the STEM.”
- Lichte, H. (1991). *Adv. in Opt. & Electron Microsc.* 12, 25–91. “Electron image plane off-axis holography of atomic structures.”
- Loane, R. F., Kirkland, E. J. and Silcox, J. (1988). *Acta Cryst. A* 44, 912–927. “Visibility of single heavy atoms on thin crystalline silicon in simulated dark-field STEM images.”
- Loane, R. F., Xu, P. and Silcox, J. (1991). *Acta Cryst. A* 47, 267–278. “Thermal vibrations in convergent-beam electron diffraction.”
- Loane, R. F., Xu, P. and Silcox, J. (1992). *Ultramicroscopy* 40, 121–138. “Incoherent imaging of zone axis crystals with ADF STEM.”
- Lord Rayleigh (1896). *Phil. Mag. (5)* 42, 167–195. “On the theory of optical images with special reference to the microscope.”
- McGibbon, M. M., Browning, N. D., Chisholm, M. F., McGibbon, A. J., Pennycook, S. J., Ravikumar, V. and Dravid, V. P. (1994). *Science* 266, 102–104. “Direct determination of grain boundary atomic structures in SrTiO₃.”
- McGibbon, A. J., Pennycook, S. J. and Angelo, J. E. (1995). *Science* 269, 519–521. “Direct observation of dislocation core structures in CdTe/GaAs(001).”

- Möbus, G. (1996). *Ultramicroscopy* 65, 205–216. “Retrieval of crystal defect structures from HREM images by simulated evolution. I. Basic technique.”
- Möbus, G. and Dehm, G. (1996). *Ultramicroscopy* 65, 217–228. “Retrieval of crystal defect structures from HREM images by simulated evolution. II. Experimental image evaluation.”
- Nakamura, K., Kakibagashi, H., Kanehori, K. and Tanaka, N. (1997). *J. Electron Microsc.* 46, 33–43. “Position dependence of the visibility of a single gold atom in silicon crystals in HAADF-STEM image simulation.”
- Nellist, P. D. and Pennycook, S. J. (1996). *Science* 274, 413–415. “Direct imaging of the atomic configuration of ultradispersed catalysts.”
- Nellist, P. D. and Pennycook, S. J. (1998a). *Phys. Rev. Lett.* 81, 4156–4159. “Subångstrom resolution by underfocussed incoherent transmission electron microscopy.”
- Nellist, P. D. and Pennycook, S. J. (1998b). *J. Microsc.* 190, 159–170. “Accurate structure determination from image reconstruction in ADF STEM.”
- Nellist, P. D. and Pennycook, S. J. (1999). *Ultramicroscopy* 78, 111–124. “Incoherent imaging using dynamically scattered coherent electrons.”
- Nellist, P. D. and Rodenburg, J. M. (1994). *Ultramicroscopy* 54, 61–74. “Beyond the conventional information limit: the relevant coherence function.”
- Nellist, P. D., McCallum, B. C. and Rodenburg, J. M. (1995). *Nature* 374, 630–632. “Resolution beyond the ‘information limit’ in transmission electron microscopy.”
- Orchowski, A., Rau, W. D. and Lichte, H. (1995). *Phys. Rev. Lett.* 74, 399–401. “Electron holography surmounts resolution limit of electron microscopy.”
- Pennycook, S. J. and Boatner, L. A. (1988). *Nature* 336, 565–567. “Chemically sensitive structure-imaging with a scanning transmission electron microscope.”
- Pennycook, S. J. and Jesson, D. E. (1990). *Phys. Rev. Lett.* 64, 938–941. “High-resolution incoherent imaging of crystals.”
- Pennycook, S. J. and Jesson, D. E. (1991). *Ultramicroscopy* 37, 14–38. “High-resolution Z-contrast imaging of crystals.”
- Perovic, D. D., Rossouw, C. J. and Howie, A. (1993). *Ultramicroscopy* 52, 353–359. “Imaging elastic strain in high-angle annular dark-field scanning transmission electron microscopy.”
- Rodenburg, J. M. and Bates, R. H. T. (1992). *Phil. Trans. R. Soc. (Lond.) A* 339, 521–553. “The theory of super-resolution electron microscopy via Wigner-distribution deconvolution.”
- Scherzer, O. (1936). *Z. Phys.* 101, 593–603. “Über einige Fehler von Elektronenlinsen.”
- Scherzer, O. (1949). *J. Appl. Phys.* 20, 20–29. “The theoretical resolution limit of the electron microscope.”
- Shao, Z. and Crewe, A. V. (1987). *Ultramicroscopy* 23, 169–174. “Chromatic aberration effects in small electron probes.”
- Shin, D. H., Kirkland, E. J. and Silcox, J. (1989). *Appl. Phys. Lett.* 55, 2456–2458. “Annular dark field electron microscope images with better than 2Å resolution at 100 kV.”
- Sibisi, S. and Skilling, J. (1997). *J. R. Statist. Soc. B* 59, 217–235. “Prior distributions on measure space.”
- Sivia, D. S., David, W. I. F., Knight, K. S. and Gull, S. F. (1993). *Physica D* 66, 234–242. “An introduction to Bayesian model selection.”
- Skilling, J. and Bryan, R. K. (1984). *Mon. Not. R. Astr. Soc.* 511, 111–124. “Maximum entropy image reconstruction: general algorithm.”
- Skilling, J. (1998). *J. Microsc.* 190, 28–36. “Probabilistic data analysis: An introductory guide.”
- Sleight, M. E., Midgley, P. A. and Vincent, R. (1998). *Proc. EUREM-11, Dublin, 1996* 2, 488–489. “Image improvement and information retrieval using the CLEAN algorithm.”
- Smith, D. J. (1997). *Rep. Prog. Phys.* 60, 1513–1580. “The realisation of atomic resolution with the electron microscope.”
- Spence, J. C. H. (1988). *Experimental High-Resolution Electron Microscopy*. OUP, New York.

- Spence, J. C. H. and Cowley, J. M. (1978). *Optik* 50, 129–142. “Lattice imaging in STEM.”
- Spence, J. C. H., Zuo, J. M. and Lynch, J. (1989). *Ultramicroscopy* 31, 233–240. “On the HOLZ contribution to STEM lattice images formed using high-angle dark-field detectors.”
- Tan, S. M. (1986). *Mon. Not. R. Astr. Soc.* 220, 971–1001. “An analysis of the properties of CLEAN and smoothness stabilised CLEAN — some warnings.”
- Treacy, M. M. J. and Gibson, J. M. (1993). *Ultramicroscopy* 52, 31–53. “Coherence and multiple-scattering in z-contrast images.”
- Treacy, M. M. J. and Gibson, J. M. (1995). *J. Microsc.* 180, 2–11. “Atomic contrast transfer in annular dark-field images.”
- Treacy, M. M. J., Howie, A. and Wilson, C. J. (1978). *Phil. Mag. A* 38, 569–585. “Z contrast imaging of platinum and palladium catalysts.”
- Van Dyck, D. and Coene, W. (1987). *Optik* 77, 125–128. “A new procedure for wave function restoration in high resolution electron microscopy.”
- Van Dyck, D. and Op de Beeck, M. (1996) *Ultramicroscopy* 64, 99–107. “A simple intuitive theory for electron diffraction.”
- Wade, R. H. (1976). *Optik* 45, 87–91. “Concerning tilted beam electron microscope transfer function.”
- Wade, R. H. and Frank, J. (1977). *Optik* 49, 81–92. “Electron microscope transfer functions for partially coherent axial illumination and chromatic defocus spread.”
- Zeitler, E. and Thomson, M. G. R. (1970). *Optik* 31, 258–280, 359–366. “Scanning transmission electron microscopy.”